

**Modeling Childhood Wheezing in Small Areas in Manitoba**

by

Charanpal Singh

A Thesis submitted to the Faculty of Graduate Studies of

The University of Manitoba

in partial fulfilment of the requirements of the degree of

MASTER OF SCIENCE

Department of Community Health Sciences

University of Manitoba

Winnipeg

---

## Abstract

Asthma has a significant impact on the Manitoba healthcare system. Asthma-related health expenditures in Canada are around \$2 billion annually and are the leading causes of emergency treatment for the younger demographic. Asthma is, however, challenging to diagnose at an earlier age and routine checks are not possible within the younger age groups. Wheezing however is one of the symptoms of asthma but is not exclusive to asthma. The link between wheezing and asthma sets the motivation to a model being developed for this study which would predict wheezing in children in the future. The intention is that the model developed could be used to assess whether a child is at risk of developing asthma in the future. The model developed was a small area longitudinal model which would borrow strength from follow-ups and other auxiliary variables within the Canadian Healthy Infant Longitudinal Development (CHILD) study. The model building process was guided using the CHILD cohort with an emphasis on the Manitoba site. The goal is to model wheezing through the years and produce accurate small area statistics within Manitoba by preserving as much of the individual level data, thus going beyond unit-level and time series models. The result was a model that shows how location and other covariates like previous medical history affect number of wheezing episodes and wheezing severity. By breaking down Manitoba into small geographical areas and using an indirect small area model approach, this allows to develop small area estimates which would have not been possible due to sample size limitations. In turn the small area model will allow researchers and policymakers to determine why some areas in Manitoba have higher wheezing rates than the rest of the population. Furthermore, once the CHILD study is completed and more data is available the model in this study could be used to predict asthma in early life which is often difficult to diagnose.

# Contents

<b>List of Figures</b>	<b>v</b>
<b>List of Tables</b>	<b>vii</b>
<b>1 Introduction</b>	<b>1</b>
1.1 Background . . . . .	1
1.2 Purpose and Objectives . . . . .	4
<b>2 Literature Review and Contribution</b>	<b>5</b>
2.1 Asthma and Wheezing: Impacts and Risk Factors . . . . .	5
2.2 Theoretical Background . . . . .	7
2.3 Research Contribution . . . . .	10
2.4 Ethics . . . . .	11
<b>3 Methods</b>	<b>12</b>
3.1 Research Design . . . . .	12
3.2 Study Area . . . . .	12
3.3 Data . . . . .	14
3.3.1 Covariates . . . . .	16
3.3.2 Response Variable . . . . .	21

---

3.4	Small Area Models . . . . .	29
3.4.1	Small Area Parameters . . . . .	33
3.4.2	Model Parameter Estimation . . . . .	34
3.5	Statistical Contribution . . . . .	37
3.5.1	Ordinal Logistic Model . . . . .	39
3.5.2	Zero-Inflated Poisson Model . . . . .	40
3.5.3	Area-Level Proportions and Means . . . . .	42
<b>4</b>	<b>Simulation Study</b>	<b>45</b>
4.1	Ordinal Logistic Model . . . . .	45
4.2	Binary Logistic Model . . . . .	48
4.3	Zero-Inflated Poisson Model . . . . .	50
<b>5</b>	<b>CHILD Study: Data Analysis</b>	<b>53</b>
5.1	Logistic Model . . . . .	53
5.2	Zero-Inflated Poisson Model . . . . .	59
5.3	Area-Level Models . . . . .	63
5.3.1	Proportion of Wheezing Severity . . . . .	67
5.3.2	Average Rate of Wheezing Episodes . . . . .	70
<b>6</b>	<b>Discussion &amp; Conclusions</b>	<b>75</b>
6.1	Wheezing in Small Areas in Manitoba . . . . .	75
6.2	Wheezing Throughout Childhood . . . . .	78

---

---

6.3	Strengths and Limitations . . . . .	80
6.4	Conclusions . . . . .	81
	<b>References</b>	<b>82</b>
	<b>Appendices</b>	<b>86</b>
A	Region Codes	87
B	Ordinal Logistic Simulation Probabilities	91
C	Logistic Model - Univariate Analysis	93
D	Logistic Model Building Tables	98
E	ZIP Model - Univariate Analysis	102
F	ZIP Model Building Tables	107
G	Area-Level Logistic Model- Univariate Analysis	110
H	Area-Level Logistic- Model Building	114
I	Area-Level ZIP- Univariate Analysis	115
J	Area-Level ZIP Model- Covariates for Both Components	119
K	Model Building for Area ZIP- Covariates for Both Components	124

## List of Figures

3.1	RHADs in Manitoba, with Winnipeg specific RHADs shown on right. (Region codes given in Appendix A) . . . . .	14
3.2	Map of all 690 participants in the study by RHAD in Manitoba (left) and Winnipeg (right). . . . .	24
3.3	Number of children that wheezed three months after birth by RHAD in Manitoba (left) and Winnipeg (right). . . . .	24
3.4	Number of children that wheezed six months after birth by RHAD in Manitoba (left) and Winnipeg (right). . . . .	25
3.5	Number of children that wheezed one year after birth by RHAD in Manitoba (left) and Winnipeg (right.) . . . . .	25
3.6	Number of children that wheezed eighteen months after birth by RHAD in Manitoba (left) and Winnipeg (right). . . . .	26
3.7	Number of children that wheezed two years after birth by RHAD in Manitoba (left) and Winnipeg (right). . . . .	26
3.8	Number of children that wheezed two and half years after birth by RHAD in Manitoba (left) and Winnipeg (right). . . . .	27
3.9	Number of children that wheezed three years after birth by RHAD in Manitoba (left) and Winnipeg (right). . . . .	28
3.10	Number of children that wheezed four years after birth by RHAD in Manitoba (left) and Winnipeg (right). . . . .	28

---

3.11 Number of children that wheezed five years after birth by RHAD in Manitoba (left) and Winnipeg (right) . . . . .	29
5.1 Predicted proportion of wheezing severity using the final area-level logistic model in Manitoba. Winnipeg specific are proportions shown on the right panel (Region codes are given in Appendix A) . . . . .	69
5.2 Proportion of wheezing severity obtained from the data in Manitoba (aggregated wheezing severity at 3 months divided by sample size in each area). Winnipeg specific proportions are shown on the right panel (Region codes are given in Appendix A) . . . . .	70
5.3 Predicted average number of wheezing episodes in Manitoba using the final area-level (new) ZIP model. Winnipeg specific proportions are shown on the right panel (Region codes are given in Appendix A) . . . . .	73
5.4 Number of wheezing episodes obtained from the data in Manitoba (aggregated wheezing episodes at 3 months). Winnipeg specific proportions are shown on the right panel (Region codes are given in Appendix A) . . . . .	74
6.1 Number of wheezing episodes for children who started wheezing in every follow-up and if they continued wheezing until the last follow-up of 5 years after birth. . . . .	79
B.1 Plot of average $y_{ijk}$ 's vs corresponding average probabilities from 500 simulation runs for area one. . . . .	91
B.2 Plot of average $y_{ijk}$ 's vs corresponding average probabilities from 500 simulation runs for area four. . . . .	92

## List of Tables

3.1	Frequency table of children diagnosed with asthma at ages three and five. . . . .	16
3.2	Maternal Health: Frequency table of maternal history of wheezing. . . . .	17
3.3	Parental Health: Frequency table of if father wheezed at 18 weeks (prenatal). . . . .	17
3.4	Parental Health: Frequency table of maternal history of smoking, asked at 18 weeks (prenatal) and at 1 year follow-up. . . . .	17
3.5	Parental Health: Frequency table of paternal history of smoking during pregnancy of mother (prenatal). . . . .	18
3.6	Maternal Health: Frequency table of if mother ever had asthma asked at 18 weeks pregnant (prenatal). . . . .	18
3.7	Maternal Health: Frequency table of if mother was ever diagnosed with asthma or ever had asthma and if the mother was diagnosed and treated for asthma or wheezing (all asked at 1-year follow-up). . . . .	19
3.8	Parental Health: Frequency table of if father was ever diagnosed with asthma or ever had asthma and if the father was diagnosed and treated for asthma or wheezing at 18 weeks pregnancy (prenatal). . . . .	19
3.9	Frequency table indicating if the household had furry pets at six time points throughout the study. . . . .	19
3.10	SES: Paternal frequency table with statistics for number of years of education. . . . .	20

---

3.11 SES: Distribution of household income of the entire sub-cohort. Note “prefer not to say” is also imputed when cleaning this variable to calculate SES score. . . . .	20
3.12 Environmental: Frequency table showing distribution of participants within five environmental factors. . . . .	21
3.13 Child Health: Frequency table displaying distribution of wheezing for all nine follow-ups. . . . .	21
3.14 Child Health: Severity of wheezing frequency distribution for six follow-ups. Scale of 1-3, 1 being the least severe. . . . .	22
3.15 Child Health: Frequency table displaying the counts of wheezing episodes for all nine follow-ups. Cleaning by using the prior question, if child did not wheeze then the number of wheezing episodes would be 0. . .	23
4.1 Ordinal Simulation Study: Three chains are used with 60,000 samples from each chain. Furthermore, the Gelman diagnostics indicated convergence of model parameters. These estimates are to be used as the true parameters for the ordinal logit model. . . . .	45
4.2 Bias, relative bias, and MSE of model parameter estimates based on 500 simulation runs . . . . .	47
4.3 Binary Logit Simulation Study: Three chains are used with 60,000 samples from each chain. Furthermore, the Gelman diagnostics indicated convergence of model parameters. These estimates are to be used as the true parameters for the binary logit model. . . . .	49
4.4 Bias, relative bias, and MSE of model parameter estimates based on 500 simulation runs from the proposed logit model . . . . .	49

---

4.5 ZIP Simulation Study: Three chains are used with 60,000 samples from each chain. Furthermore, the Gelman diagnostics indicated convergence of model parameters. These estimates are to be used as the true parameters for the ZIP model. . . . .	51
4.6 Bias, relative bias, and MSE of model parameter estimates based on 500 simulation runs from the proposed ZIP model. . . . .	51
5.1 Binary logistic random effects model. . . . .	54
5.2 Final logistic model based upon model building steps with intercept and random effects. The model has six covariates; odds ratios are calculated by taking the exponent of the covariate parameter estimates. Credible intervals are provided for the parameter estimates. . . . .	55
5.3 Probabilities of wheezing severity for all six follow-ups. . . . .	56
5.4 Logistic model with prenatal covariates, based upon model building steps with intercept and random effects. . . . .	57
5.5 Logistic model with two follow-ups, based upon model building steps with intercept and random effects. . . . .	58
5.6 Unit-level logistic model with using only the 3 month follow-up of wheezing. Model has prenatal covariates with intercept and only area-level random effects. . . . .	59
5.7 ZIP model with six covariates from the model building process. . . .	61
5.8 ZIP model with four covariates. Covariates chosen based upon the models in Appendix F. . . . .	62

---

5.9	Simple unit-level ZIP model as shown in equations (5.4-5.5). Using only the 3 month follow-up of number of wheezing episodes. Model has prenatal covariates with intercept and only area-level random effects. . . . .	63
5.10	Wheezing dependent variables at 3 months aggregated by each area. Recall Wheezing severity is binary while number of wheezing episodes is a count variable. . . . .	64
5.11	Final Area-level logistic model with four covariates. Odds ratios are calculated by taking the exponent of the covariate parameter estimates. . . . .	67
5.12	Univariate area-level ZIP model with random effects and prenatal Mother asthma as a covariate. Unusually large estimates of $\gamma_0$ and $1 - \pi$ is 1 indicate this model is not optimal. . . . .	71
5.13	Univariate area-level ZIP model with Mother asthma as a covariate for both the logistic and Poisson components of the model. . . . .	71
5.14	Univariate Poisson area-specific random effects model with Mother asthma as a covariate. . . . .	72
5.15	Final area-level ZIP model with two covariates Farm and Mother Asthma for both the logistic and Poisson components of the model. . . . .	72
6.1	Child Health: Frequency table displaying distribution of wheezing for all 63 individuals who wheezed in the first follow-up (3 months). . . . .	78
A.1	Region code with associated location and small area number. . . . .	87
C.1	Univariate logistic model with random effects and Mother Wheezed as a covariate. All parameters are statistically significant. . . . .	93

C.2 Univariate logistic model with random effects and Mother Smoked as a covariate. Mother smoked is not statistically significant. . . . .	93
C.3 Univariate logistic model with random effects and Father Wheezed as a covariate. Father wheezed and $\rho$ are not statistically significant. . . . .	93
C.4 Univariate logistic model with random effects and Father Smoked as a covariate. Father smoked is statistically significant. . . . .	94
C.5 Univariate logistic model with random effects and Asthma Young as a covariate. All parameters are statistically significant. . . . .	94
C.6 Univariate logistic model with random effects and Asthma older as a covariate. All parameters are statistically significant. . . . .	94
C.7 Univariate logistic model with random effects and Mother asthma as a covariate. All parameters are statistically significant. . . . .	94
C.8 Univariate logistic model with random effects and Diagnosed Mother Asthma as a covariate. All parameters are statistically significant. . . . .	95
C.9 Univariate logistic model with random effects and mother treated for asthma as a covariate. All parameters are statistically significant. . . . .	95
C.10 Univariate logistic model with random effects and Father asthma as a covariate. Father asthma is not statistically significant. . . . .	95
C.11 Univariate logistic model with random effects and Diagnosed father asthma for as a covariate. Diagnosed father asthma is not statistically significant. . . . .	95
C.12 Univariate logistic model with random effects and Father treated for asthma for as a covariate. Father treated for asthma is not statistically significant. . . . .	96

---

C.13 Univariate logistic model with random effects and Furry pets as a covariate. Furry pets is statistically significant. . . . .	96
C.14 Univariate logistic model with random effects and SES as a covariate. SES is not statistically significant. . . . .	96
C.15 Univariate logistic model with random effects and Highway as a covariate. Highway is not statistically significant. . . . .	96
C.16 Univariate logistic model with random effects and Water as a covariate. Water is not statistically significant. . . . .	97
C.17 Univariate logistic model with random effects and Factory as a covariate. Factory is not statistically significant. . . . .	97
C.18 Univariate logistic model with random effects and Factory as a covariate. All parameters are statistically significant. . . . .	97
C.19 Univariate logistic model with random effects and Construction as a covariate. Construction is not statistically significant. . . . .	97
D.1 Model with two covariates: Asthma young and Asthma older. . . . .	98
D.2 Model with three covariates: Asthma young, Asthma older and Mother wheezed. . . . .	98
D.3 Model with four covariates: Asthma young,Asthma older, Mother wheezed and Diagnosed mother asthma. . . . .	98
D.4 Model with five covariates. Mother asthma and Diagnosed mother asthma did not converge and both have high standard errors. One of the variables will need to be removed eventually. . . . .	99

---

D.5	Model with six covariates: Asthma young,Asthma older, Mother wheezed, Diagnosed mother Asthma older,Mother asthma and newly added Farm. . . . .	99
D.6	Model with seven covariates. Asthma young,Asthma older,Mother wheezed, Diagnosed mother asthma older,Mother asthma, Farm added alongside with Furry pets. . . . .	100
D.7	Removed Mother asthma from Table D.5. After removing this vari- able, the model converged. . . . .	100
D.8	Added the covariate Furry pets back to the model in Table D.7. . . .	100
D.9	Removed Farm from the model in Table D.8. Removing Farm does not make any significant changes to the model. . . . .	101
E.1	Univariate ZIP model with random effects and Mother asthma as a covariate. All parameters are statistically significant. . . . .	102
E.2	Univariate ZIP model with random effects and Mother wheezed as a covariate. All parameters are statistically significant. . . . .	102
E.3	Univariate ZIP model with random effects and Mother smoked as a covariate. Mother smoked is not statistically significant. . . . .	103
E.4	Univariate ZIP model with random effects and Father wheezed as a covariate. Father wheezed is not statistically significant. . . . .	103
E.5	Univariate ZIP model with random effects and Father smoked as a covariate. Father smoked is not statistically significant. . . . .	103
E.6	Univariate ZIP model with random effects and Furry pets as a co- variate. Furry pets is not statistically significant. . . . .	104

---

E.7	Univariate ZIP model with random effects and SES as a covariate. SES is not statistically significant. . . . .	104
E.8	Univariate ZIP model with random effects and Highway as a covariate. Highway is not statistically significant. . . . .	104
E.9	Univariate ZIP model with random effects and Water as a covariate. Water is not statistically significant. . . . .	105
E.10	Univariate ZIP model with random effects and Factory as a covariate. Factory is not statistically significant. . . . .	105
E.11	Univariate ZIP model with random effects and Farm as a covariate. Farm is not statistically significant. . . . .	105
E.12	Univariate ZIP model with random effects and Construction as a covariate. Construction is not statistically significant. . . . .	106
F.1	ZIP model with two covariates,Mother asthma (prenatal) and Mother wheezed (prenatal). . . . .	107
F.2	ZIP model with three covariates: Mother asthma (prenatal), Mother wheezed (prenatal) and Mother smoked (prenatal). . . . .	107
F.3	ZIP model with four covariates. Father smoked added to the previous model. . . . .	108
F.4	ZIP model with five covariates. Farm added to the previous model. .	108
F.5	ZIP model with six covariates. Furry pets added to the previous model. . . . .	109
G.1	Univariate area-level logistic model with prenatal Mother smoked as a covariate. Mother smoked is not statistically significant. . . . .	110

---

G.2 Univariate area-level logistic model with prenatal Mother wheezed as a covariate. Mother smoked is statistically significant. . . . .	110
G.3 Univariate area-level logistic model with prenatal Mother asthma as a covariate. Mother smoked is statistically significant. . . . .	110
G.4 Univariate area-level logistic model with Father smoked as a covariate. Father smoked is not statistically significant. . . . .	111
G.5 Univariate area-level logistic model with Father wheezed as a covariate. Father wheezed is statistically significant. . . . .	111
G.6 Univariate area-level logistic model with Father asthma as a covariate. Father asthma is not statistically significant. . . . .	111
G.7 Univariate area-level logistic model with Highway as a covariate. Highway is not statistically significant. . . . .	111
G.8 Univariate area-level logistic model with Water as a covariate. Water is not statistically significant. . . . .	112
G.9 Univariate area-level logistic model with Factory as a covariate. Factory is not statistically significant. . . . .	112
G.10 Univariate area-level logistic model with Farm as a covariate. Farm is not statistically significant. . . . .	112
G.11 Univariate area-level logistic model with Construction as a covariate. Construction is not statistically significant. . . . .	112
G.12 Univariate area-level logistic model with SES as a covariate. SES is not statistically significant. . . . .	113

---

H.1	Area-level logistic model with two prenatal covariates, Mother wheezed and Mother asthma. Odds ratios were calculated by taking the exponent of the covariate parameter estimates. . . . .	114
H.2	Area-level logistic model with three prenatal covariates, Mother wheezed,Mother asthma and Mother smoked. . . . .	114
I.1	Univariate area-level ZIP model with random effects and prenatal Mother smoked as a covariate. Mother smoked is not statistically significant. . . . .	115
I.2	Univariate area-level ZIP model with random effects and prenatal Mother wheezed as a covariate. Mother wheezed is not statistically significant. . . . .	115
I.3	Univariate area-level ZIP model with random effects and prenatal Mother asthma as a covariate. Mother asthma is statistically significant.	115
I.4	Univariate area-level ZIP model with random effects and prenatal Father smoked as a covariate. Father smoked is not statistically significant.	116
I.5	Univariate area-level ZIP model with random effects and prenatal Father wheezed as a covariate. Father wheezed is not statistically significant. . . . .	116
I.6	Univariate area-level ZIP model with random effects and prenatal Father asthma as a covariate. Father asthma is not statistically significant.	116
I.7	Univariate area-level ZIP model with random effects and Highway as a covariate. Highway is not statistically significant. . . . .	116
I.8	Univariate area-level ZIP model with random effects and Water as a covariate. Water is not statistically significant. . . . .	117

---

I.9	Univariate area-level ZIP model with random effects and Factory as a covariate. Factory is not statistically significant. . . . .	117
I.10	Univariate area-level ZIP model with random effects and Farm as a covariate. Farm is not statistically significant. . . . .	117
I.11	Univariate area-level ZIP model with random effects and Construction as a covariate. Construction is not statistically significant. . . . .	117
I.12	Univariate area-level ZIP model with random effects and SES as a covariate. SES is not statistically significant. . . . .	118
J.1	Univariate area-level ZIP model with Mother smoked as a covariate for both the logistic and Poisson component of the model. Mother smoked is not statistically significant. . . . .	119
J.2	Univariate area-level ZIP model with Mother wheezed as a covariate for both the logistic and Poisson component of the model. Mother wheezed is not statistically significant. . . . .	119
J.3	Univariate area-level ZIP model with Mother asthma as a covariate for both the logistic and Poisson component of the model. Mother asthma is not statistically significant. . . . .	119
J.4	Univariate area-level ZIP model with Father smoked as a covariate for both the logistic and Poisson component of the model. Father smoked is not statistically significant. . . . .	120
J.5	Univariate area-level ZIP model with Father wheezed as a covariate for both the logistic and Poisson component of the model. Father wheezed is not statistically significant. . . . .	120

---

J.6 Univariate area-level ZIP model with Father asthma as a covariate for both the logistic and Poisson component of the model. Father asthma is statistically significant. . . . .	120
J.7 Univariate area-level ZIP model with Highway as a covariate for both the logistic and Poisson component of the model. Highway is not statistically significant. . . . .	120
J.8 Univariate area-level ZIP model with Water as a covariate for both the logistic and Poisson component of the model. Water is not statistically significant. . . . .	121
J.9 Univariate area-level ZIP model with Factory as a covariate for both the logistic and Poisson component of the model. Factory is not statistically significant. . . . .	121
J.10 Univariate area-level ZIP model with Farm as a covariate for both the logistic and Poisson component of the model. Farm is statistically significant. . . . .	121
J.11 Univariate area-level ZIP model with Construction as a covariate for both the logistic and Poisson component of the model. Construction is not statistically significant. . . . .	121
J.12 Univariate area-level ZIP model with SES as a covariate for both the logistic and Poisson component of the model. SES is statistically significant. . . . .	122

---

J.13 Univariate area-level ZIP model with SES (average before) as a covariate for both the logistic and Poisson component of the model. In this model the average of the individual covariates that are used to make the SES index (Father/Mother Education and household income) is taken by area and then an area level SES index is calculated. . . . .	122
J.14 Univariate area-level ZIP model with Mother education as a covariate for both the logistic and Poisson component of the model. In this model one of the covariates that are used to make the SES index is aggregated to the area-level by averaging within each area. Mother education is not statistically significant. . . . .	122
J.15 Univariate area-level ZIP model with Father education as a covariate for both the logistic and Poisson component of the model. In this model one of the covariates that are used to make the SES index is aggregated to the area level by averaging within each area. Father education is statistically significant. . . . .	122
J.16 Univariate area-level ZIP model with Income as a covariate for both the logistic and Poisson component of the model. In this model one of the covariates that are used to make the SES index is aggregated to the area level by averaging within each area. Father education is statistically significant. . . . .	123
J.17 Area-level ZIP model with no covariate on both the logistic and Poisson component of the model. . . . .	123
K.1 Area-level ZIP model with SES average before and Farm as a covariates for both the logistic and Poisson component of the model. . . . .	124

---

K.2	Area-level ZIP model with three covariates, for both the logistic and Poisson component of the model. . . . .	124
K.3	Area-level ZIP model with four covariates, added prenatal covariate Mother smoked to the previous model for both the logistic and Poisson component of the model. . . . .	125
K.4	Area-level ZIP model with two covariates Farm and Mother Asthma for both the logistic and Poisson component of the model. . . . .	125
K.5	Area-level ZIP model with three covariates: Farm, Mother Asthma and smoked for both the logistic and Poisson component of the model.	125
K.6	Area-level ZIP model with two covariates Farm and Mother smoked for both the logistic and Poisson component of the model. . . . .	126

# **Chapter 1: Introduction**

## **1.1 Background**

Asthma is characterized by airway inflammation. Asthma has symptoms such as shortness of breath, chest tightness, coughing and wheezing (Global Initiative for Asthma, 2019, pg. 16). Also, another symptom of asthma is variable expiratory airflow limitation, meaning that a patient's ability to breathe out air is reduced which can make breathing difficult (Global Initiative for Asthma, 2019, pg. 17).

The age groups with the highest prevalence of asthma in Canada seem to be ages 1-34 years (Public Health Agency of Canada, 2018). The prevalence of asthma seems to be decreasing in the older demographic but increasing in the younger demographic. Incidence rates for asthma also show a similar pattern to the prevalence of asthma in Canada (Statistics Canada, 2019). Thus, an outcome is needed which can be measured among young children to assess whether a child has asthma. It is important to note that symptoms of asthma are often triggered by exposure to allergens or irritants, changes in the weather or other respiratory infections and can begin early in life when children with asthma start to wheeze (Richelle, 2014). However, it can be challenging to make a diagnosis of asthma in children five years and younger since wheezing symptoms can be present at this age group without asthma; furthermore it is also not possible to have routine airflow limitation measurement within younger age groups (Global Initiative for Asthma, 2019).

Crucially not all children who wheeze develop asthma (Public Health Agency of Canada, 2018). In fact previous data from a birth cohort study with annual reports of wheezing (Childhood Allergy Study) was used to determine whether wheezing at particular times in early life was predictive of current asthma at the age of six years

---

(Richelle, 2014). Current asthma was defined in the study as the child ever having a physician diagnosis of asthma and either symptoms of asthma including wheezing, asthma attacks and use of asthma medication in the prior year (Wegienka et al., 2009). The results from the study indicate that wheezing at any particular time in early life may not be predictive of asthma at age six (Brick et al., 2019). However, using groupings of any early and late wheezing, wheezing at ages of four or five years was strongly associated with current asthma compared with early wheezing in the first three years for all children in the study (Wegienka et al., 2009). Another study using the PASTURE (Protection Against Allergy—Study in Rural Environments) data indicated that correct parental recall of wheezing episodes in the first year of life was associated with asthma at age six years old (Brick et al., 2019). True-positive recall of wheezing was associated with a higher risk of subsequent asthma and impaired lung function as compared to false-negative recall. Studies involving parental recall of wheezing would have analyzed with caution since this could lead to over or under estimation of asthma prevalence among children. Wheezing early on in life has complex interactions with later onsets of asthma and is still an ongoing debate and is worth exploring further (Subbarao et al., 2009).

The difficulty of diagnosing asthma earlier in life means that wheezing can be used as an indicator of ‘possible’ asthma diagnosis in the future, so wheezing is an essential factor when assessing a child’s risk of developing asthma (Richelle, 2014). Wheezing will be the outcome of interest for this study and all modeling will be done with wheezing as the dependent variable. A model to describe and predict wheezing severity by location and time would be useful for the healthcare system in Manitoba since this would allow policymakers to allocate resources to areas within Manitoba which are most at need. Also, further insight could be gained into preventive measures by analyzing the conditions in which wheezing occurs for young children. The intended model could be used in a separate study to model the risk of a child developing asthma with wheezing as a risk factor for asthma.

---

This research project aims to use the cohort data from Canadian Healthy Infant Longitudinal Development (CHILD) study to model different severities of wheezing in small areas within Manitoba. Generally, a “small area” is denoted as any domain or area for which direct estimates of adequate precision cannot be produced (Rao & Molina, 2015). In other words, the small area models will provide better estimates for areas which have sample size limitations. In a small area model, the target population is broken down in-to “areas”, the goal is then to find reliable small area statistics for these areas. Areas can be physical locations or any other type of sub-populations such as age-sex-ethnicity domains. By using an indirect model-dependent approach, instead of direct estimation, the intended model can borrow strength from other resources (e.g., covariates from an administrative dataset) to produce better prediction of severity of wheezing in small areas (Rao & Molina, 2015).

In the literature there are small area models such as area-level and unit-level models (Rao & Molina, 2015). Area-level models allow for between area variation and include area specific covariates while unit-level models take this further and incorporate unit specific covariates relevant to individuals within a small area (Rao, 1994, pg. 155). A study used Poisson and logistic regression to find covariates that predict recurrent wheezing (more than two episodes), however since there were no small areas involved in the study, the information in the breastfeeding study was effectively collapsed and just reported the overall statistics of the CHILD study (Azad et al., 2017). By using a unit-level model, strength thus can be borrowed from other small areas, and other resources like auxiliary variables (Rao, 1999, pg. 177). Breastfeeding for example could then be a unit-level variable if the Winnipeg region is broken up into small areas and within those small areas breastfeeding responses were used as covariates to predict our response which is wheezing.

---

## 1.2 Purpose and Objectives

This research project will seek to model wheezing severity using a small area estimation technique using longitudinal data. Keeping all this in mind that this project will be quantitative, and therefore the research questions will also be quantitative in nature. **Since location will be an integral part of this research project, the primary research question will be: to what extent does location affect the severity of wheezing?** Thus, an area-level model of wheezing count at different small areas will be performed to give a general indication on wheezing rates in different small areas. Impact of location will then be investigated using an area-level and individual-level model. The models mentioned use concepts of the unit-level and area-level, more of which will be explained in the methods section. It should be kept in mind that location could be a proxy for other contributing covariates such as specific home environment of the child. Thus, location will be analyzed to reveal why a location has a higher rate of wheezing severity to find hidden variables which cause certain locations to have higher rates. The hypothesis is that location will have an impact on wheezing at both area and individual levels.

**A secondary research question will be to determine how wheezing severity changes throughout childhood?** The goal of this research question is to determine if there are any cyclic patterns that could lead to better preparation for spikes in wheezing at certain age ranges? The research question will incorporate both a location and time element to answer this question. The hypothesis will be that seasonal changes and also throughout childhood will have some influence in wheezing, but the goal will be to quantify this and other covariates that cause spikes in wheezing severity.

# **Chapter 2: Literature Review and Contribution**

## **2.1 Asthma and Wheezing: Impacts and Risk Factors**

Asthma-related expenditures in Canada are around \$2 billion annually and are the leading causes of emergency treatment for the younger demographic (The Canadian Lung Association, 2011, pg. 1-2). In 2016/17 there were 55,102 asthma-related emergency room visits. Usually, there are 55,000 to 70,000 asthma-related visits annually and emergency department length of stay for people admitted to hospitals with asthma in 2016/17 was up 11% when compared to 2015 (Canadian Institute for Health Information, 2018). There are no current reports related to emergency wait times from the Canadian Institute of Health Information (CIHI), but it is known that asthma is adding to the strain of emergency services.

Although the information on the link between wheezing and asthma is limited, both share some common risk factors, including exposure to environmental tobacco smoke, low socioeconomic status, crowding, and allergen exposure, and increased risk in males (Morgan & Martinez, 1992, pg. 1199). An extensive literature review was undertaken as part of the CHILD study to examine risk factors for the development of asthma in early childhood (Subbarao et al., 2009). The study found prenatal risk factors for asthma including maternal smoking, use of antibiotics and delivery by caesarean section. Risk factors for asthma later in childhood include sex, allergic sensitization, in particular house dust mite, cat and cockroach allergens, exposure to environmental tobacco smoke, antibiotic use, and infections. Several studies have proved lower risk of asthma and wheezing with exposure to farm animals in early life. However, the findings show inconsistency between the association of asthma and exposure to domestic cats and dogs (Subbarao et al., 2009). Therefore, when choosing covariates for modeling, pet exposure and exposure to farm

---

animals should be analyzed carefully. Furthermore, the influence of breastfeeding on the risk of asthma is still an ongoing debate. Breastfeeding may initially decrease the risk of asthma development but then later on in time can increase the risk of sensitization (Quirt et al., 2018, pg. 16). A study which investigated the association of breastfeeding and wheezing using the CHILD study converted the counts into rates by dividing the number of episodes by the follow-up months (Azad et al., 2017). The study found association between wheezing in earlier life and breastfeeding from participants 12 months and younger. However, further follow-ups are needed within the CHILD study to accurately determine the association between asthma and breastfeeding.

Recent studies have also found maternal history of asthma is significantly associated with childhood development of asthma and wheezing (Venter et al., 2021). It is widely accepted, that children with a family history of atopic diseases are far more likely to develop similar conditions, although specifics on which gene or genes have effects on offspring allergy have not fully been established (Venter et al., 2021, pg. 7-8).

There is indeed a complex relationship between asthma and wheezing. It is crucial to understand that wheezing does not necessarily lead to development of asthma. In fact, many studies have shown that a large proportion of young children who wheeze, even recurrently, show remission of symptoms at the age of three or during school years (Castro-Rodriguez et al., 2019, pg. 1-2). The major difficulty in assessing childhood asthma is that it cannot be thought of as a single disease entity, but in fact several wheezing phenotypes, each of which can have different risk factors (Caui et al., 2013, pg. 1396). If adequate data is not available which describes wheezing severity, a method is needed to determine if an individual is at greater risk of developing asthma later on in life. Generally, wheezing phenotypes are characterized by timing of onset and subsequent decline or persistence of wheezing (Duijts et al., 2016). Wheeze phenotype classifications are often based on parental reports of

---

onset, severity and frequency of symptoms combined with suspected trigger factors (Brick et al., 2019, pg. 796). Thus, it should be kept in mind that parental recall is subject to recall bias. In children with asthma, three wheeze phenotypes have been identified: transient early wheezing; nonatopic wheezing; and atopic wheezing (Quirt et al., 2018). These phenotypes should be kept in mind when assessing wheezing and asthma severity and careful consideration needs to be placed on how wheezing severity is defined. The current phenotypes are time dependent and each has its own range of severity. For example, children with the nonatopic wheezing phenotype tend to have milder asthma than the atopic phenotype (Quirt et al., 2018). Particular attention is needed on how to define wheezing severity, will severity refer to the most severe exacerbation a child has ever had, or a combination of the number and severity of individual episodes or even a combination of either one of these plus persistence of asthma. Clearly defining severity leads to better models that can predict asthma.

## 2.2 Theoretical Background

Wheezing phenotypes are based on longitudinal data since frequency is important in classifying the correct phenotype. Predictive models can supplant clinical judgment when deciding on treatment since only using phenotype to predict asthma is tough with children whose future is obviously unknown (Castro-Rodriguez et al., 2019, pg. 1). Thus, asthma prediction models can be used to determine if a child is at risk of developing asthma in the future. Most existing predictive models for asthma development in children are based on the statistical method of logistic regression. The predictive models usually try to predict asthma development at certain age ranging from 6-13 years (Luo et al., 2015). The predictors in the models consist of responses from questionnaires which include sex, smoking mother, parent asthma history and other relevant medical conditions such as allergic rhinitis. The difficulty is that predictive models for asthma in children use differing asthma definitions and asthma development by various ages which may lead to different results. In general

---

models with a conservative asthma definitions help identify the predictors of and estimate the risk for true asthma (Luo et al., 2015, pg. 12).

Other popular predictive models are based on either risk score or combination of risk factors, which include clinical indexes. Asthma predictive indexes (API) often use similar predictor variables to statistical models and often are sub-divided into stringent indexes or loose indexes. Loose indexes require less strict combinations of major and minor risk factors. Crucially APIs can be included as predictors of a logistic model in order to test their capability of predicting either frequent wheezing or asthma. However, sensitivity and specificity can also be used to directly test the predictive models (Luo et al., 2015, pg. 11). The original API remains the most clinically viable. It has major risk factors parental history of asthma or eczema with minor risk factors, eosinophilia, wheezing without colds, and allergic rhinitis (Castro-Rodriguez et al., 2019). The original API is popular due since it is derived and built from a multi-ethnic population, it is also simple and cheap to recollect (Castro-Rodriguez et al., 2019). There are other promising clinical indexes but it is agreed that the orginal API should be used in a health setting to identify children at risk of developing asthma (Castro-Rodriguez et al., 2019).

The existing models in the literature have some limitations. The current predictive models for asthma in children make predictions at a time unsuitable for making clinical impact (Luo et al., 2015, pg. 10-11). Ideally a predictive model for asthma development is before children reach the age of two. Thus special consideration will be placed on predicting asthma before the age of two, this of course depends on the availability of data. Furthermore, to be clinically valuable, a predictive model for asthma development in children needs to have both high positive predictive value and high sensitivity (Savenije et al., 2012, pg. 327). There are several potential approaches for improving accuracy of the predictive models. Larger data sets could be used combined with exhaustive variable sets. Further improvements can be made by focusing on a child population with a high prevalence of future asthma develop-

---

ment. Most models for predicting asthma in the child population have low accuracy, typically with a sensitivity or positive predictive value less than 80% (Luo et al., 2015, pg. 4-11). Thus it would be desirable to develop an accurate predictive model and then use it in a clinical setting to measure its impact.

Populations with high prevalence of asthma could also be linked to specific areas. In a study performed in 2018, data were collected in Manitoba during the 2000–2010 fiscal years and analyzed for annual physician visits relating the total respiratory morbidity (TRM) condition, which included patients diagnosed with asthma, chronic or acute bronchitis (Shokoohi & Torabi, 2018, pg. 338). The study due to the presence of missing values used 217 of 222 regional health authority districts (RHADs) as small areas. A logistic mixed model was used in the study to make inferences about the rate of physician TRM visits in all the 217 RHADs throughout the different years. Overall, the study found that Winnipeg and some areas in southern Manitoba have larger rates of asthma visits compared to other parts of the province. The authors reported that the results may represent real increases or different distributions of important covariates that are unmeasured and unadjusted for in the model, since the only covariates available were age and sex. There is a limited amount of literature on geographic location and its development to asthma in childhood. Although it is understood that environment is linked with asthma development, there are limited models which incorporate the random effects into small area models. A systematic review found that although there is no direct known association between greenness and asthma, it seems to be protective when present with other factors that contribute to childhood asthma such as difficult family relationships, high traffic volume, and tobacco smoke exposure (Hartley et al., 2020, pg. 457-459). Greenness is difficult to incorporate into the model since data does not always contain specific area level information, thus an area-level model to predict wheezing or asthma would account for such important covariates which would usually be unobserved.

---

## 2.3 Research Contribution

The main point of the aforementioned indices and models is to predict outcomes and to have meaningful discussions about asthma prognosis with parents and caregivers, and to further accelerate prevention studies that will be targeted to children at the highest risk of developing persistent wheezing and asthma symptoms later in life. To that end, the contribution of this project is to preserve as much information in the CHILD study and use that information to track wheezing throughout different ages and provide its related risk factors and prediction of wheezing for as many small areas the data will allow. The current asthma and wheezing predictive models do not incorporate area as risk, thus the main goal of this project is open the way on area based asthma and wheezing predictive models.

Due to the longitudinal nature of this study, the basic small area models will be extended to allow for follow-ups between participants. When extending the unit-level model, it will create an opportunity to allow for autoregressive error components while also incorporating area-specific and unit-specific covariates. Extending beyond the time series model use in the prior study will lead to new statistical models being developed and preserve the individual-level data since follow-ups will also be part of the intended longitudinal small area model of this project. The intention is to borrow strengths from other small areas, follow-ups, and other resources, since the CHILD cohort has a rich amount of potential explanatory variables. Different outcome variables will be considered including logistic model for binary data (e.g., wheezing or not) and Poisson model for count data (e.g., number of wheezing severity episodes). The logit model will track the most severe exacerbations and the Poisson model will track the number of wheezing severity episodes. Combined the two models could be used to see how wheezing frequency and severity changes throughout for the children, thus covering two definitions of wheezing severity with hopes that this will improve prediction of asthma earlier in childhood when compared to the existing APIs.

---

There is no current research on small area methods used to model wheezing severity in Manitoba. Therefore, this project will be useful in revealing the current spread of wheezing in Manitoba and the potential causes. Policy makers will also be able to identify which areas within Manitoba are most at risk for wheezing. Using the model, we can map which areas are most at risk for asthma development with the longitudinal model and thus in future research other clinical indexes can be used on individuals in small areas to provide further credibility to parents that a child may have asthma.

## **2.4 Ethics**

All required ethical conditions by the CHILD study cohort were satisfied, and approval was granted. Ethical conditions included any privacy concerns relating to individual participants. I have additionally gained certificate of approval from the Bannatyne campus research ethics board.

# **Chapter 3: Methods**

## **3.1 Research Design**

This study is a population-based study using data from the Canadian Healthy Infant Longitudinal Development (CHILD) cohort. The CHILD cohort study is a prospective longitudinal pregnancy cohort study. Due to the nature of prospective studies the prevalence of wheezing cases among the participants could be low since participants were sampled randomly.

## **3.2 Study Area**

Between late 2008 to 2012, 3,621 pregnant mothers were recruited from four major cities across Canada (Vancouver, Edmonton, Winnipeg and Toronto) and a small rural population (Morden and Winkler) outside of Winnipeg. These mothers gave birth to 3,537 children eligible for inclusion in the study, of whom 3,455 began the study. Inclusion criteria were if the mother's age was greater than 18 years ( $>19$  years in Vancouver), living near ( $<50$  Km) and planning to deliver at a designated recruitment center participating hospital, able to read, write and speak English, willing to donate cord blood, and child born at or after 35 weeks (Subbarao et al., 2015; Takaro et al., 2015). While exclusion criteria included major congenital abnormalities or respiratory distress syndrome (RDS), expectation of moving away from a recruitment center within 1 year of recruitment, children of multiple births or resulting from in vitro fertilization, and children not spending over 80% of time in the index home (Subbarao et al., 2015; Takaro et al., 2015). Recruitment included having staff meet mothers in antenatal ultrasound clinics and physician offices as well as community "baby fairs" and included person-to-person referrals and lastly social media advertising (Takaro et al., 2015, pg. 581). Since 80% of Canadians live in urban

---

centers, the recruited population of the CHILD cohort should be representative of the Canadian population (Takaro et al., 2015, pg. 582). For this study, the Manitoba sub-cohort is considered which has 1,055 participants of which some were excluded from the study resulting in 690 individuals remaining. The reasons for excluding some participants are explained further in the data section.

For the context of this study Manitoba will be broken down into small geographical regions (areas) and thus more information will be provided into which regions have higher rates of wheezing severity. The province has five regional health authorities (RHAs) created from the initial eleven RHAs in 2012. These five RHAs are further divided to 96 regional health authority districts (RHADs), these RHADs will be used as small areas for this study (Figure 3.1). By using small areas further insight is gained onto why a specific location has a differing rate of wheezing and the associated contributing factors. A simple summary statistic that provides the population wheezing rate provides very little information to analyze, but by breaking down into small (high-resolution) areas much more insight can be gained. All CHILD participants in Manitoba were recruited within a radius of Winnipeg and Morden-Winkler. However, families who moved out of this region continued to be followed thus some areas outside of southern Manitoba may have low frequency of participants, which would misrepresent the wheezing rates in these particular regions.

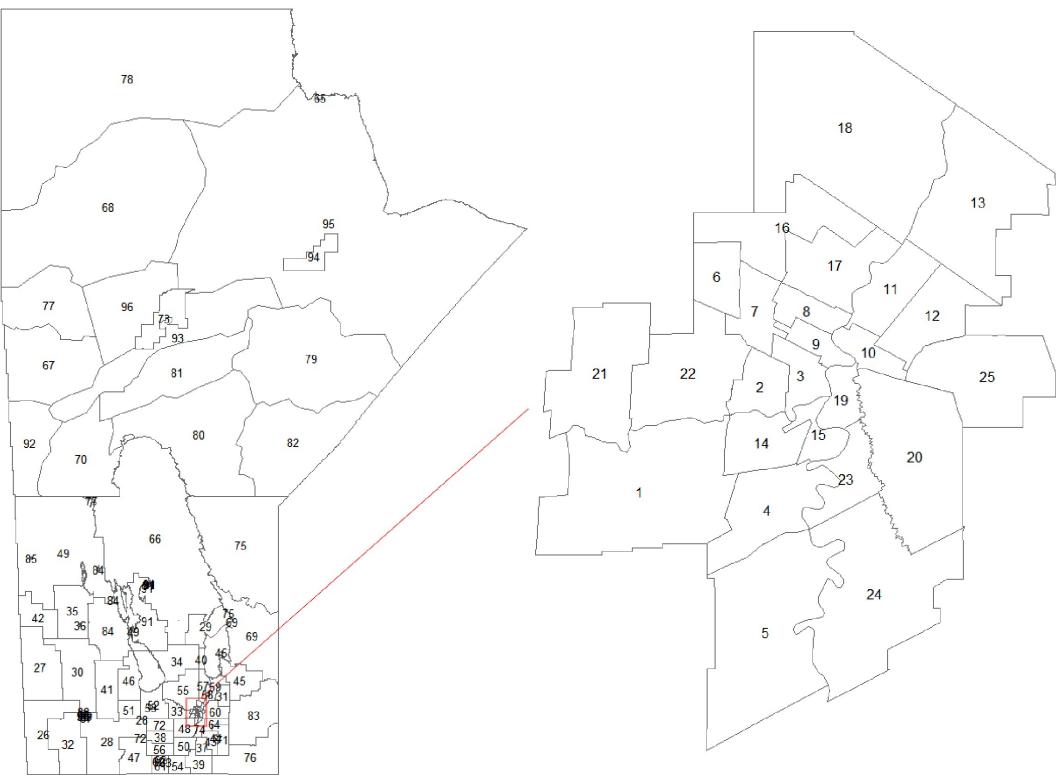


Figure 3.1: RHADs in Manitoba, with Winnipeg specific RHADs shown on right.  
(Region codes given in Appendix A)

### 3.3 Data

The CHILD study contains extensive environmental and health questionnaires at various time points. The follow-up time points in the CHILD study are prenatal, 3, 6, 12, 18 months and 2, 2.5, 3, 4, and 5 years. The primary outcome of the CHILD study is expert physician diagnosed asthma, but wheezing is one of the secondary outcomes (Subbarao et al., 2015, pg. 1-2). Since the CHILD study focuses on asthma as an outcome, the data includes exposure domains which include indoor and outdoor air pollutants, inhalation, ingestion and dermal uptake of chemicals, mold, dampness, biological allergens, pets and pests, housing structure, and living behavior, together with infections, nutrition, psychosocial environment, and medications (Takaro et al., 2015, pg. 1). The follow-ups of concern for study are those which contain wheezing responses, in total there are nine follow-ups which is the max amount possible but not all the response within the questionnaires have complete

---

follow-ups. Specifically, it should be noted that parents rated the severity of their child's worst wheezing on a scale of 1-3 and at six time points between 3 months and 2.5 years. The questionnaires in the CHILD cohort include child health, home environment, residences, socioeconomic status (SES) and parental health questionnaires. Additionally, the child clinical assessment questionnaire is used but it should be noted that although the clinical assessment was also conducted at 1 year, the questions needed for this projects were only available for two follow-ups at ages three and five in Manitoba. Data from the questionnaires will be used as covariates in the model building process based upon how much influence each variable has on wheezing. Time independent covariates in the questionnaires include sex, paternal/maternal history of asthma. Thus, follow-up independent covariates will act as baseline covariates which will explain the initial variation of wheezing. Furthermore, since not all wheezing is caused by asthma, other variables such as infections and previous medical history will also be used to predict wheezing. Time dependent covariates are participant's health, and changes in home environment and smoking since these are known from literature to be risk factors for wheezing. The time dependent covariates mentioned are asthma triggers and thus may influence wheezing in a participant. Another variable not specifically mentioned in the literature is SES which is also used in the statistical models of this study. The predictor variables or covariates are all at the individual level. Viral and environmental sensitization will also be looked at but will depend on the amount of quality available for these covariates.

The data for the baseline model is a subset of the CHILD study. The subset takes the individuals from the follow-up with the most unique postal codes (2 year follow-up) and removes individuals with incomplete follow-ups for postal codes or missing postal codes. Thus all follow-ups are assumed to have the same location throughout the study. This is due to the fact that the model will become overly complex if movement between follow-ups is present. The responses from the home environment questionnaire indicate that between 80 to 100 participants moved

---

between each follow-up in Manitoba.

The data is combined between responses and covariates after cleaning and imputing using the mean of non-missing values for each variable. The number of missing values for each covariate is shown in this section for the subcohort of 690 individuals. Missing values for the covariates are imputed by area mean imputation using the five health regions in Manitoba. Due to the objective of this study imputation by area seems the most appropriate for the context of this study. Individuals with two or more missing values of both the parental education (in years) questions and household income are removed since all three are needed to create the SES index.

### 3.3.1 Covariates

#### Diagnosis of asthma

Diagnosis of asthma is a categorical variable with three possible outcomes (0,1,2) and has two follow-ups at years 3 and 5. If the physician diagnosed the child of having no asthma in the follow-up, then the outcome is 0, if diagnosed with asthma then 1 and possible asthma was represented as 2. The outcomes 1-2 are merged to form a binary variable. This is done due to lack of information for levels 1 and 2 but also the distinction between the levels is not concrete. The two covariates are represented by AY (Asthma young; at 3 years) and AO (Asthma older; at 5 years).

Asthma young	
	Freq
No	595
Yes	26
Possible	61
Missing	8

(a) Three years

Asthma older	
	Freq
No	551
Yes	31
Possible	85
Missing	23

(b) Five years

Table 3.1: Frequency table of children diagnosed with asthma at ages three and five.

---

## History of maternal and parental wheezing

The binary response which represents if the mother or father of the child ever had a wheezing noise coming from the chest. The question was asked at the prenatal and 1-year mark for both parents but refers to at any point of time prior to the follow-up thus can be used as a baseline variable. The two covariates are represented by MW (Mother wheezed) and FW (Father wheezed).

Mother wheezed	
	Freq
No	489
Yes	200
Missing	1

(a) Prenatal (18 weeks)

Mother wheezed	
	Freq
No	520
Yes	133
Missing	37

(b) One year

Table 3.2: Maternal Health: Frequency table of maternal history of wheezing.

Father wheezed	
	Freq
No	454
Yes	198
Missing	38

Table 3.3: Parental Health: Frequency table of if father wheezed at 18 weeks (prenatal).

## Maternal and parental smoking

The binary response if mother smoked at 18 weeks and 1 year follow-up. Data for father smoking is only available for at 18 weeks. The covariates are represented by MS (Mother smoked) and FS (Father smoked).

Mother smoked	
	Freq
No	656
Yes	32
Missing	2

(a) Prenatal (18 Weeks)

Mother smoked	
	Freq
No	614
Yes	35
Missing	41

(b) One year

Table 3.4: Parental Health: Frequency table of maternal history of smoking, asked at 18 weeks (prenatal) and at 1 year follow-up.

Father smoked	
	Freq
No	526
Yes	125
Missing	39

Table 3.5: Parental Health: Frequency table of paternal history of smoking during pregnancy of mother (prenatal).

### Maternal and parental history of asthma

The binary response at the 1-year time point is if both parents have ever had asthma and if it was diagnosed by a physician while also asking if both parents ever had any treatment for asthma or wheezing. The responses are represented by MA (Mother asthma), DMA (Diagnosed mother asthma), MTFAOW (Mother treated for asthma/wheezing), FA (father asthma), DFA (Diagnosed father asthma), and FTFAOW (father treated for asthma/wheezing).

It should be clarified why three related variables are used to represent asthma for both parents. Both MA and FA were questionnaire response and are based entirely on parental recall, however DMA and DFA asked specifically if the parent was diagnosed by a physician for asthma. MTFAOW and FTFAOW ask further if both parents were treated for asthma. Ideally we would like to use DMA and DFA as covariates but we have to be mindful of the percentage of missing values and how many responses were imputed for each question. Initially all of parental asthma variables will be considered but the most significant response will be kept to avoid collinearity issues within the model.

Mother asthma	
	Freq
No	528
Yes	161
Missing	1

Table 3.6: Maternal Health: Frequency table of if mother ever had asthma asked at 18 weeks pregnant (prenatal).

Diagnosed Mother asthma		Mother asthma		Mother treated for asthma/wheezing	
	Freq		Freq		Freq
No	3	No	507	No	487
Yes	143	Yes	146	Yes	166
Missing	544	Missing	37	Missing	37

Table 3.7: Maternal Health: Frequency table of if mother was ever diagnosed with asthma or ever had asthma and if the mother was diagnosed and treated for asthma or wheezing (all asked at 1-year follow-up).

Diagnosed father asthma		Father asthma		Father treated for asthma/wheezing	
	Freq		Freq		Freq
No	12	No	532	No	529
Yes	106	Yes	119	Yes	120
Missing	572	Missing	39	Missing	41

Table 3.8: Parental Health: Frequency table of if father was ever diagnosed with asthma or ever had asthma and if the father was diagnosed and treated for asthma or wheezing at 18 weeks pregnancy (prenatal).

## Furry pets

A binary response indicating if there are any furry pets living within the participants household. In total, six follow-ups are used from 3 months to 2.5 years.

Furry pets						
	3 months	6 months	1 year	18 months	2 years	2.5 years
No	363	348	338	349	373	345
Yes	316	310	311	284	317	295
Missing	11	32	41	57	0	50

Table 3.9: Frequency table indicating if the household had furry pets at six time points throughout the study.

## Socio-economic status (SES)

Specific responses from the prenatal SES questionnaire are the number of years of education of both parents and total household income. Parental number of years of education and household income are mean imputed by health region, factor analysis is then performed on these covariates to create a SES index using Thompson's scores. Only the prenatal responses are used due to the high number of

---

missing responses on later follow-ups. Participants that responded that they would not like to disclose their income are also imputed.

Mother education (years)		Father education (years)	
Average	16	Average	15
Missing	25	Missing	32

(a) Average education of mother in years, excluding missing values.

(b) Average education of father in years, excluding missing values.

Table 3.10: SES: Paternal frequency table with statistics for number of years of education.

Household income	Frequency
\$0 - \$9,999	5
\$10,000 - \$19,999	17
\$20,000 - \$29,999	24
\$30,000 - \$39,999	38
\$40,000 - \$49,999	54
\$50,000 - \$59,999	52
\$60,000 - \$79,999	111
\$80,000 - \$99,999	115
\$100,000 - \$149,999	158
\$150,000 or over	46
Prefer not to say	90
Missing	18

Table 3.11: SES: Distribution of household income of the entire sub-cohort. Note “prefer not to say” is also imputed when cleaning this variable to calculate SES score.

## Environment factors near home

The binary response is if participant’s home is within 100 metres from any major highway/artery, body of water, factory, farm, and prolonged construction. Responses were used from the 3-month follow-up, and for the farm covariate, 1 year follow-up is also used in a later model. The other follow-ups are not used due to data unavailability and missing values.

Home environment						
	Highway	Water	Factory	Farm (3 month)	Farm (1 year)	Construction
No	529	598	665	602	589	591
Yes	151	82	15	78	62	89
Missing	10	10	10	10	39	10

Table 3.12: Environmental: Frequency table showing distribution of participants within five environmental factors.

### 3.3.2 Response Variable

#### Child wheezing

Question asks with 9 follow-ups if the child ever had a whistling noise coming from the chest with or without a cold. The question was not used as a covariate but rather used as a method to clean the response variable severity of wheezing. To clean severity of wheezing, we simply note that if a child does not wheeze then the wheezing severity should be 0. Note that code 888 means question was skipped due to a previous answer. Additional codes are also present in the data, 8888 means skipped the entire questionnaire of that particular follow-up and 999 means no response for the particular question.

Wheezed since last follow-up	3 Months	6 months	1 year	18 months	2 years	2.5 years	3 years	4 years	5 years
No	618	596	576	567	600	581	557	433	499
Yes	63	62	73	67	78	58	38	55	48
888	0	0	0	0	0	0	0	0	0
999	0	0	0	0	1	0	0	0	1
8888	9	32	41	56	11	51	95	202	142
Missing (all codes)	9	32	41	56	12	51	95	202	143

Table 3.13: Child Health: Frequency table displaying distribution of wheezing for all nine follow-ups.

The next set of wheezing variable tables will be used as dependent variables.

#### Severity of wheezing

Table 3.14 shows the frequencies of each ordinal category for the dependent variable severity of wheezing. Severity of wheezing is cleaned using the parental

---

response of whether their child wheezed by adding a fourth response to wheezing severity for children who did not wheeze as a response 0 (not shown in table). Since participants followed the required skip pattern and did not answer a question that was not relevant based on their previous response changing those skipped responses to 0 merges the variables regarding the presence and severity of wheezing. The remaining missing values of wheezing are mean imputed.

Severity of wheezing						
Severity	3 months	6 months	1 Year	18 months	2 years	2.5 years
1	34	29	26	2	31	24
2	19	16	19	6	17	12
3	8	7	12	2	13	9
888	0	0	16	57	0	0
8888	9	32	41	56	11	51
999	2	10	0	0	18	13
Missing (all codes)	11	42	57	112	29	64

Table 3.14: Child Health: Severity of wheezing frequency distribution for six follow-ups. Scale of 1-3, 1 being the least severe.

## Number of wheezing episodes

Recall childhood wheezing asked with 9 follow-ups if the child ever had a whistling noise coming from the chest with or without a cold. Also recall that this question was used to clean the response variable severity of wheezing. We simply noted that if a child does not wheeze then the wheezing severity should be 0. A follow-up question to childhood wheezing in the CHILD cohort is the number of episodes of wheezing the child experienced during last six months of the follow-up in question. This response like childhood wheezing has 9 follow-ups but is count data representing the number of episodes per follow-up. Due to the nature of how the two questions are set, if the participant answered they did not wheeze then the number of wheezing episodes question is skipped and code 888 is collected as a response. Therefore the childhood wheezing question can be used to clean the number of wheezing episodes, if the child did not wheeze then the number of wheezing episodes would be 0.

Number of wheezing episodes since last follow-up	3 months	6 months	1 year	18 months	2 years	2.5 years	3 years	4 years	5 years
0	618	596	576	567	600	581	557	433	499
1	41	33	36	31	28	19	12	18	20
2 to 6	13	15	19	23	30	23	17	26	20
6 to 11	2	1	2	2	3	0	1	0	5
11 and higher	2	1	0	0	0	0	0	1	0
888	0	0	16	10	0	0	0	9	1
999	5	12	0	1	18	16	8	1	3
8888	9	32	41	56	11	51	95	202	142
Missing (all codes)	14	44	57	67	29	67	103	212	146

Table 3.15: Child Health: Frequency table displaying the counts of wheezing episodes for all nine follow-ups. Cleaning by using the prior question, if child did not wheeze then the number of wheezing episodes would be 0.

Note from Tables 3.13 and 3.15 that episodes of wheezing still have more missing values than the prior question of childhood wheezing, even after cleaning. This may be because a small number of participants skipped the question when they were supposed to answer it. However, the crucial point to consider here is that those who had no history of wheezing should have a zero count whenever possible, even if it leads to an excess of zeros in future models. This however brings up the challenge on how to handle excess zeros in the data. The remaining missing values of wheezing episode counts will be mean imputed by the five health regions, similar to how the covariates were imputed.

We also provide maps to show number of individuals with wheezing by RHAD. The maps are useful in portraying the spread of wheezing within Manitoba for all 6 follow-ups. The maps presented are the wheezing cases of the sub-cohort of completed cases within the CHILD study. There are of course more wheezing cases in Manitoba not captured with data available for this study since recruitment of participants was done in Winnipeg and Morden-Winkler. Note that white areas in the map indicate no wheezing response from the participants, and missing areas mean recruitment was not done in those areas.

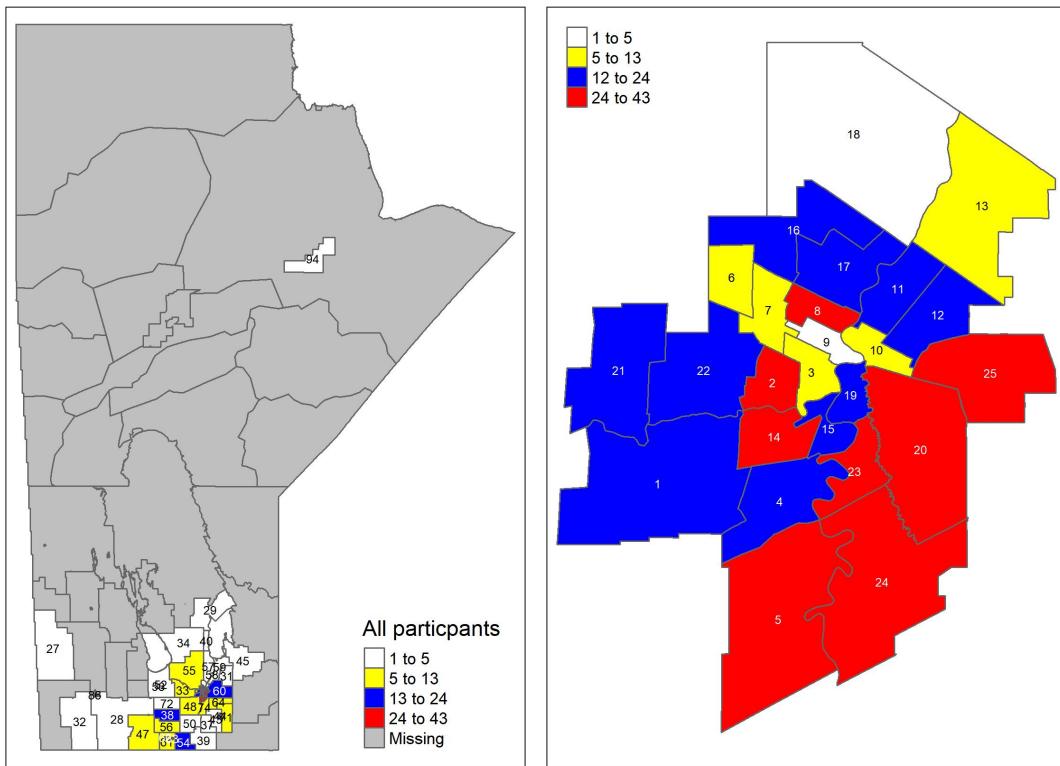


Figure 3.2: Map of all 690 participants in the study by RHAD in Manitoba (left) and Winnipeg (right).

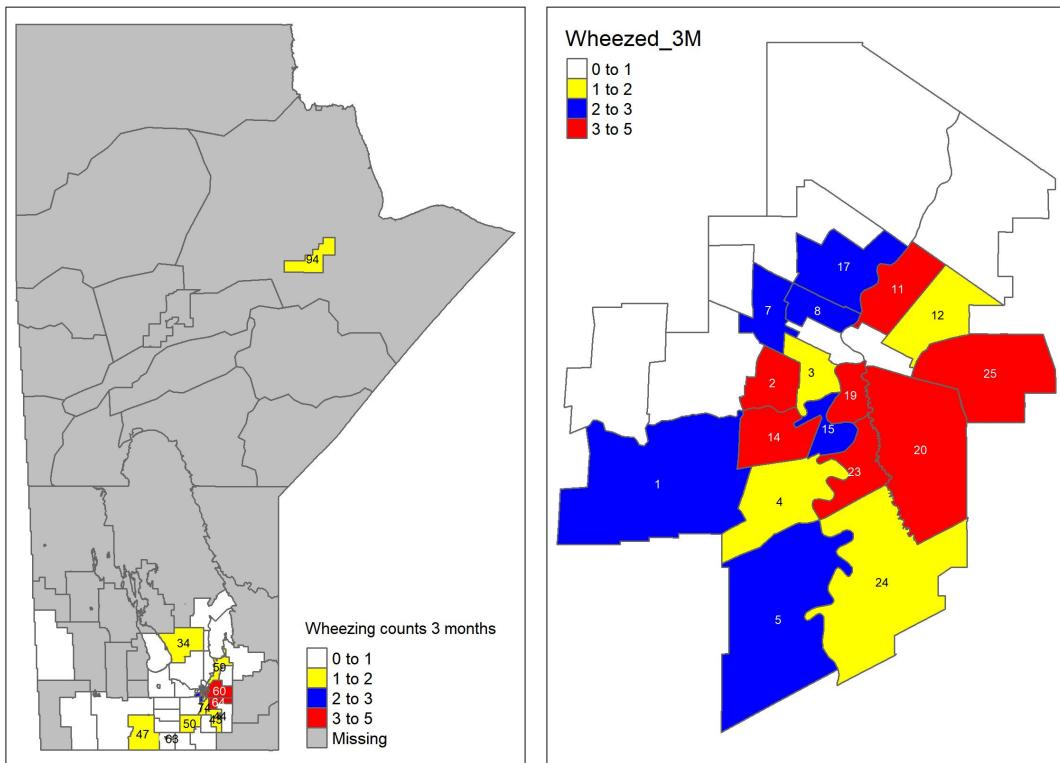


Figure 3.3: Number of children that wheezed three months after birth by RHAD in Manitoba (left) and Winnipeg (right).

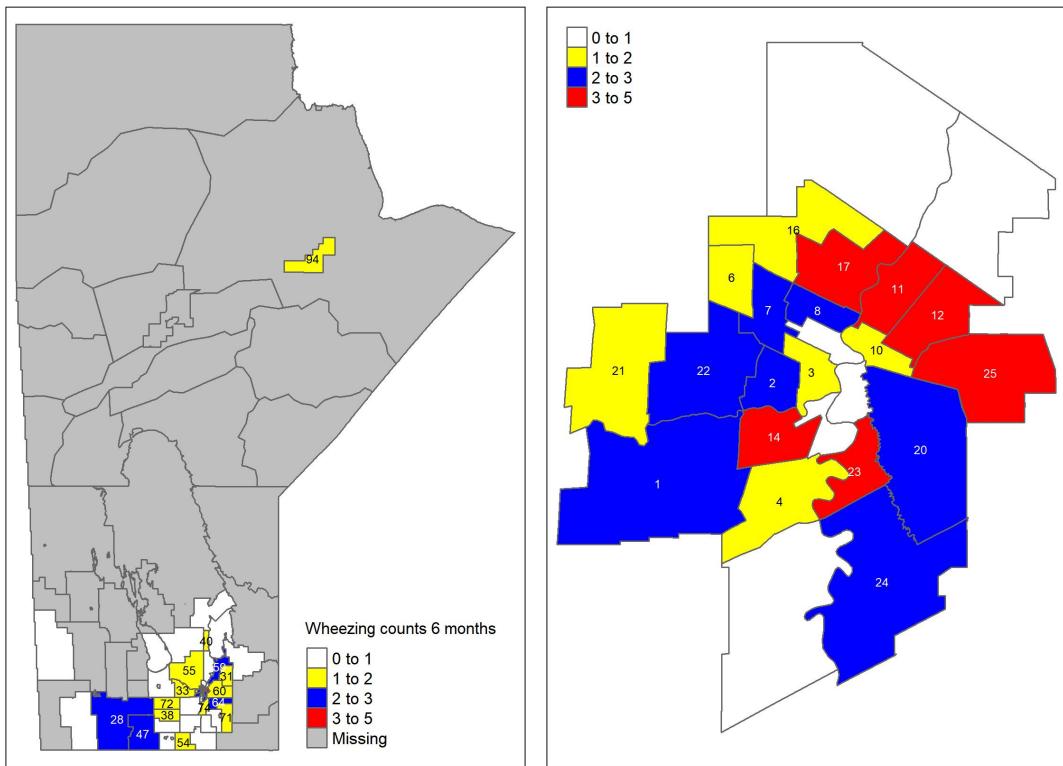


Figure 3.4: Number of children that wheezed six months after birth by RHAD in Manitoba (left) and Winnipeg (right).

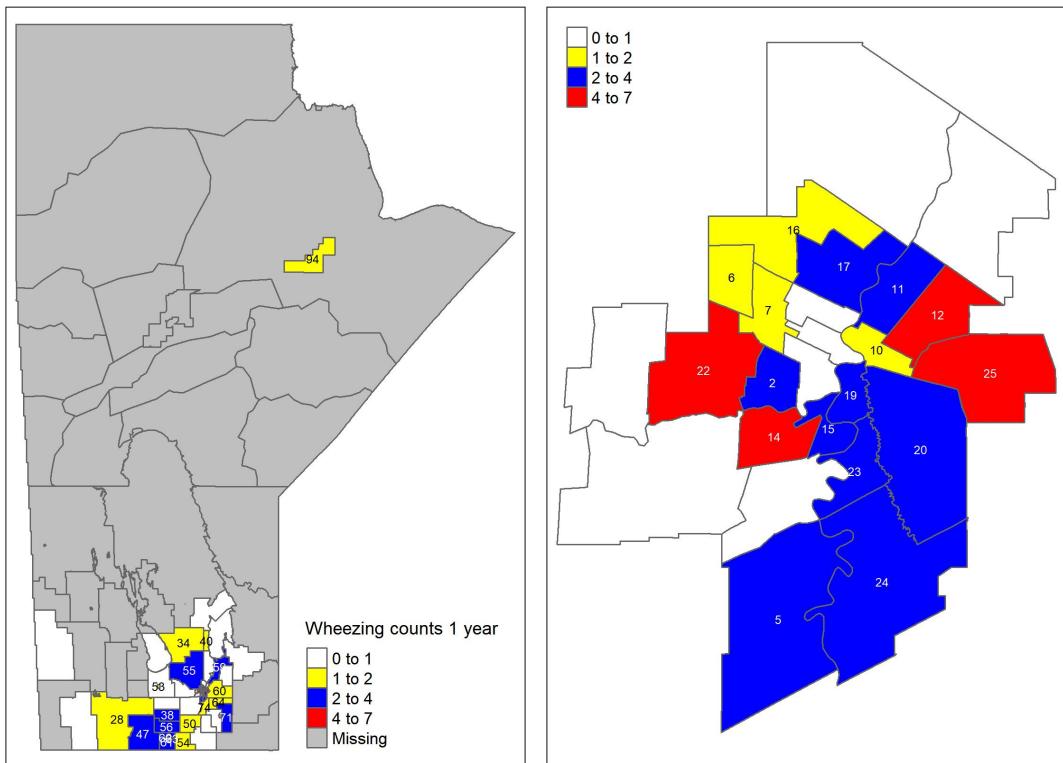


Figure 3.5: Number of children that wheezed one year after birth by RHAD in Manitoba (left) and Winnipeg (right.).

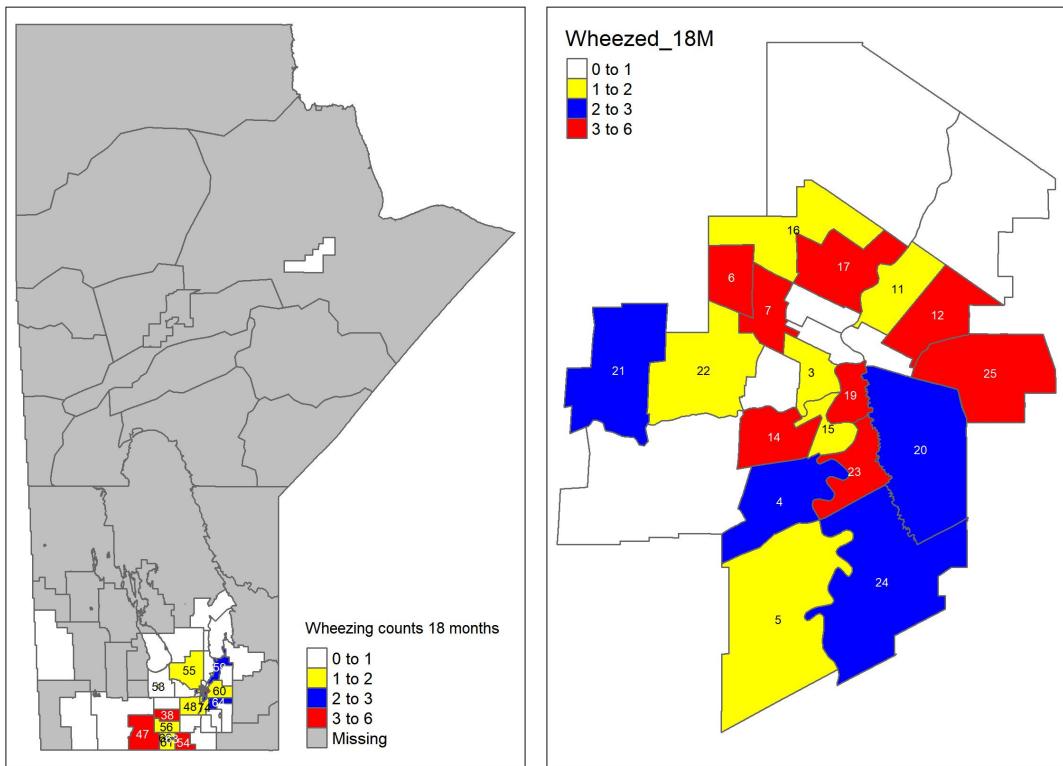


Figure 3.6: Number of children that wheezed eighteen months after birth by RHAD in Manitoba (left) and Winnipeg (right).

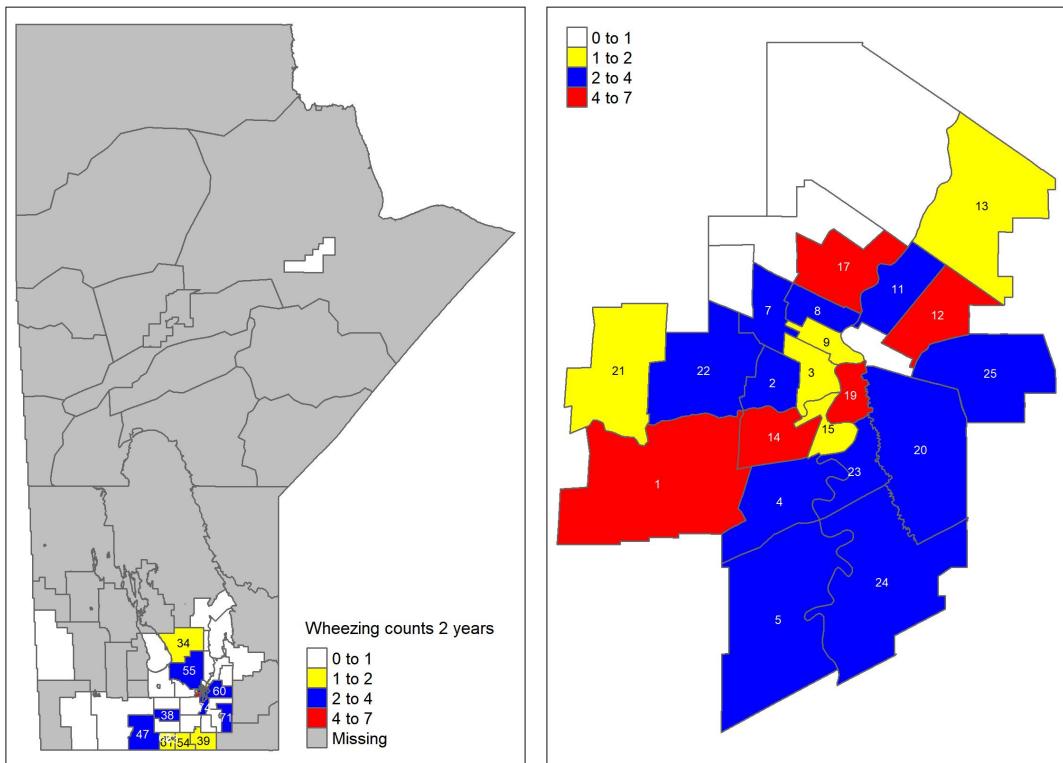


Figure 3.7: Number of children that wheezed two years after birth by RHAD in Manitoba (left) and Winnipeg (right).

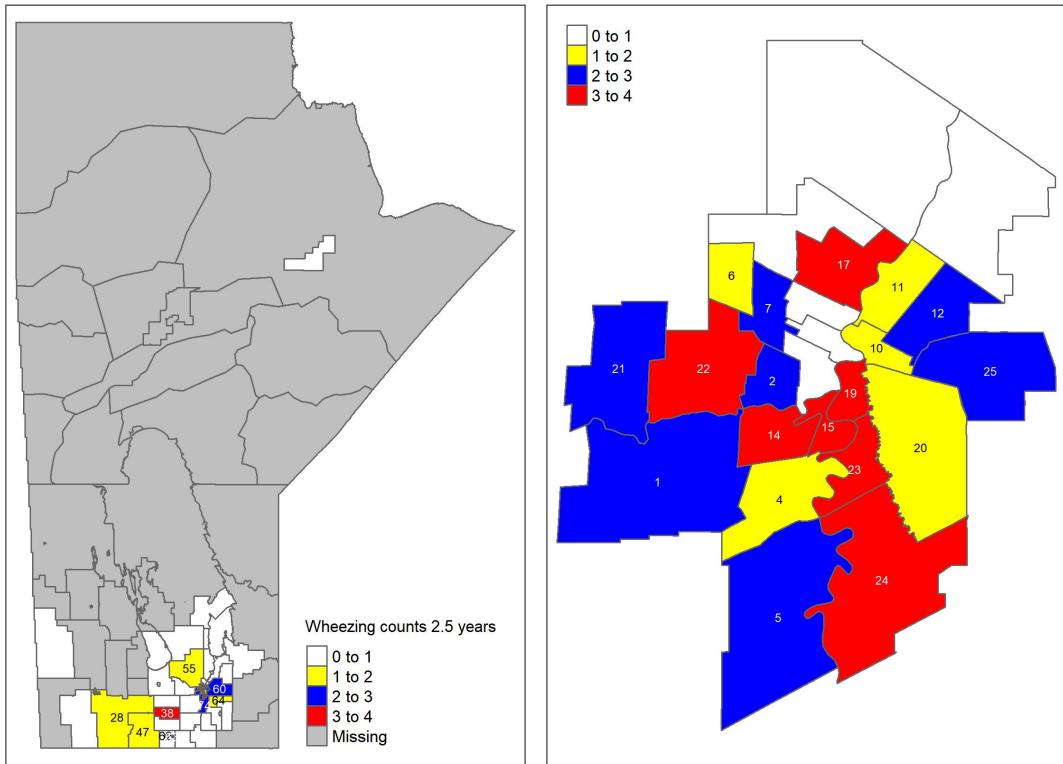


Figure 3.8: Number of children that wheezed two and half years after birth by RHAD in Manitoba (left) and Winnipeg (right).

From Figures 3.3-3.11 we can see that the highest wheezing counts are within Winnipeg. From 3 months to 1 year there is a region in the north of Manitoba, Fox Lake Cree Nation coded by region 94 which is the only northern RHAD with wheezing episodes.

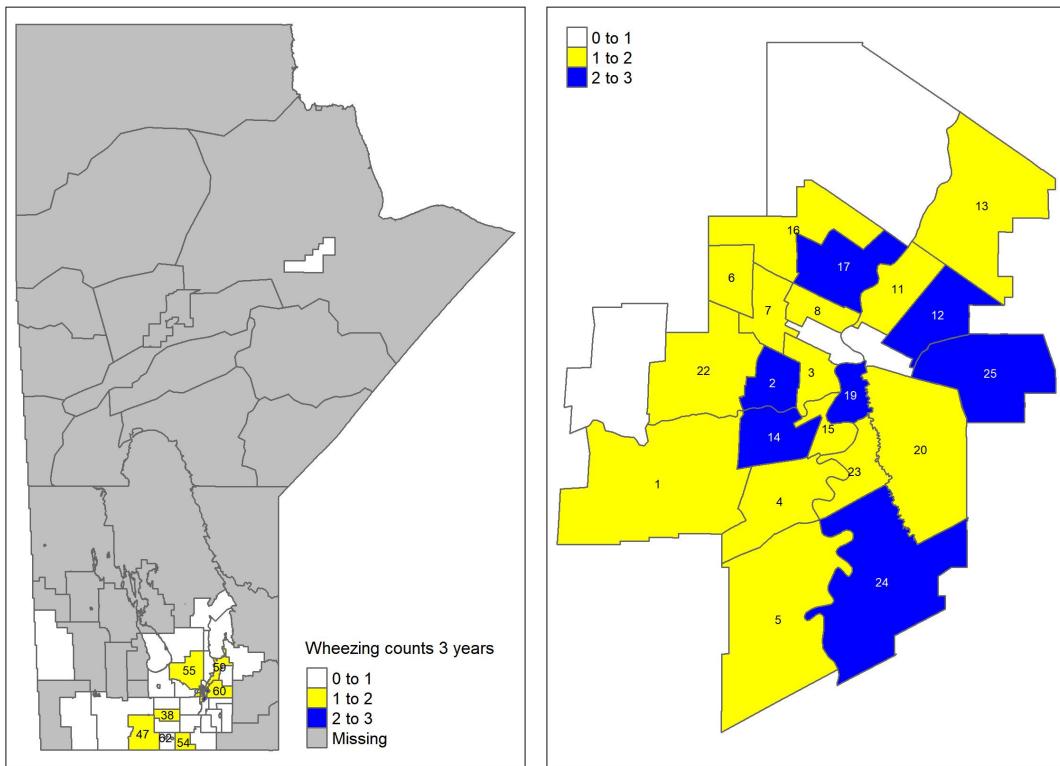


Figure 3.9: Number of children that wheezed three years after birth by RHAD in Manitoba (left) and Winnipeg (right).

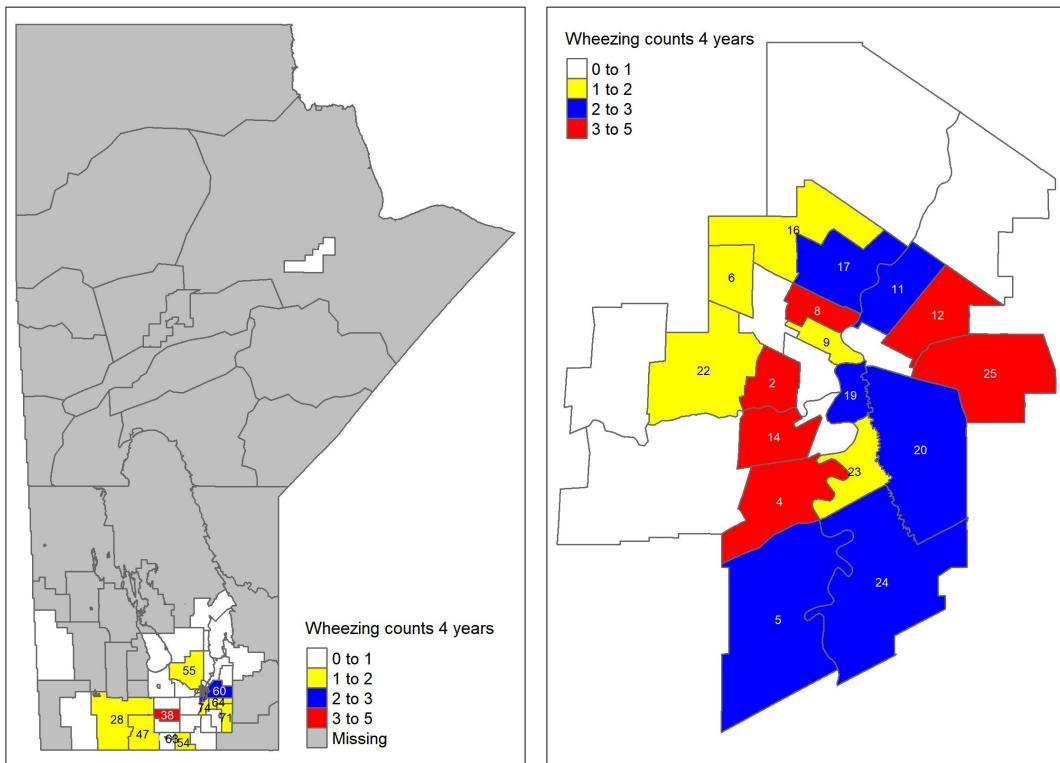


Figure 3.10: Number of children that wheezed four years after birth by RHAD in Manitoba (left) and Winnipeg (right).

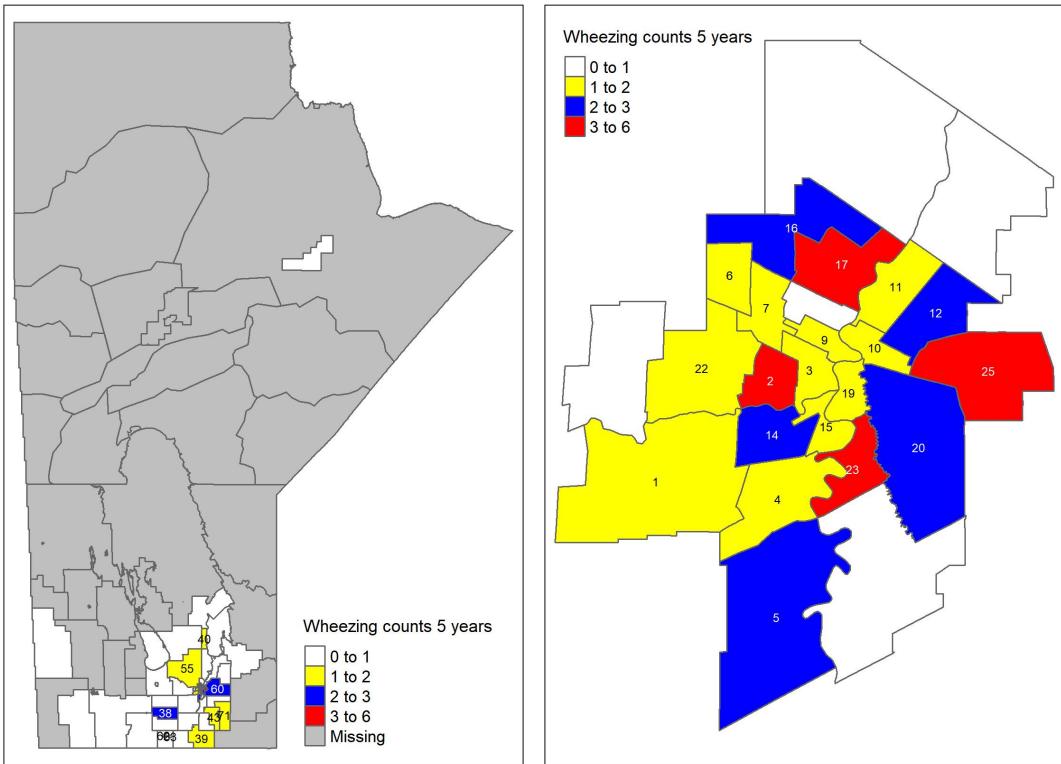


Figure 3.11: Number of children that wheezed five years after birth by RHAD in Manitoba (left) and Winnipeg (right).

### 3.4 Small Area Models

Model building using small areas is an important aspect of this study. It is important that the basic small area models are understood before building upon the more complex models needed for this study. To begin with we will discuss the difference between direct-domain and indirect-domain estimation. Domains are defined by geographic areas or socio-demographic groups or any other subpopulations. We simply refer to a domain estimator as “direct” if it is based only on the domain-specific sample data, but it still may use auxiliary data related to the outcome (Rao & Molina, 2015, pg. 24-25). An indirect estimator is used when the sample is not large enough to use direct estimator with an adequate level of precision. Thus, this makes it necessary to find indirect estimators that increase the “effective” sample size and as a result the standard error is also reduced. Indirect estimators are design-based and “borrow strength” by using values of the variable of interest as auxiliary

---

information from related areas or other resources (Rao & Molina, 2015, pg. 24). Area-level and unit-level auxiliary variables affect the bias and mean squared error (MSE) of a parameter estimator.

Small area models are model-based indirect estimators which allow for variations that cannot be fully explained by the direct variable. The two basic types of small area models are area-level and unit-level models. Area-level models are also called aggregate models which relate small area means to area-specific auxiliary covariates (Rao & Molina, 2015, pg. 97-98) . Unit-level models relate the unit values within a small area to unit-specific auxiliary variables. Unit-level auxiliary variables provide individual-level information. Area-level models are especially important when unit-level data is not available for small areas due to confidentiality or other reasons. The parameter of interest is some function of the small area means  $\bar{Y}_k$  for area  $k$  such that  $\mu_k = g(\bar{Y}_k)$ . It is crucial to understand that the end goal of small area analysis is to estimate the small area parameter  $\mu_k$ . Assume the indirect estimator of  $\mu_k$  is

$$\widehat{\mu}_k = g\left(\widehat{\bar{Y}}_k\right) = \mu_k + e_k, \quad k = 1, \dots, K, \quad (3.1)$$

where  $K$  is the number of small areas and error terms  $e_k$  ( $k = 1, \dots, K$ ) are independently distributed with mean 0 and variance  $\psi_k$ . Note that it is assumed that  $\psi_k$  is known to avoid identifiability issues. Further, if area-level auxiliary data  $\mathbf{x}_k = (x_{1k}, \dots, x_{pk})^T$  are available then a linear model can be formed as

$$\mu_k = \mathbf{x}_k^T \boldsymbol{\beta} + b_k v_k, \quad k = 1, \dots, K, \quad (3.2)$$

where  $b_k$ 's are known positive constants, these are to set optional different random effects for each small area. The  $v_k$  term represents the area-specific random effects with the assumption that all  $v_k$ 's are independent and identically distributed

---

(iid) with mean 0 and variance  $\sigma_v^2$ , as denoted by  $v_k \stackrel{iid}{\sim} \mathcal{N}(0, \sigma_v^2)$ . Note that  $v_k$ 's are assumed to be normally distributed but are not limited to this distribution. Lastly  $\beta$  is a vector of regression coefficients. Combining (3.1) with (3.2) results in

$$\widehat{\mu}_k = \mathbf{x}_k^T \beta + b_k v_k + e_k, \quad k = 1, \dots, K, \quad (3.3)$$

which is the final form of the basic area-level model or is also referred to as the Fay–Herriot (FH) model (Fay & Herriot, 1979). Now if there is unit-specific auxiliary data available within each small area, then the basic unit-level model can be used to model small areas. Assuming that  $i$  is any element say an individual within any small area  $k$  and then  $n_k$  presents the sample size in each area, then the model is

$$y_{ik} = \mathbf{x}_{ik}^T \beta + v_k + e_{ik}, \quad i = 1, \dots, n_k, \quad k = 1, \dots, K. \quad (3.4)$$

In Equation (3.4),  $\mathbf{x}_{ik} = (x_{ik1}, \dots, x_{ikp})^T$  is auxiliary data and  $\beta$  are corresponding covariates. Like the FH model,  $v_k$  represents the area-level random effects while the sampling error is  $e_{ik} \sim (0, t_{ik}\sigma_e^2)$  with known constants  $t_{ik}$  but unknown variance  $\sigma_e^2$ . It should be kept in mind that the parameters of interest like the FH model are the small area means,  $\bar{Y}_k$ . Note that one can also incorporate area-level and unit-level covariates into the unit-level model (3.4) as is the case for the CHILD cohort. It should be mentioned that (3.3) and (3.4) are special cases of the general linear mixed model (gLMM)

$$\mathbf{y} = \mathbf{X}\beta + \mathbf{Z}\mathbf{v} + \mathbf{e}, \quad (3.5)$$

where  $\mathbf{X}$  and  $\mathbf{Z}$  are known matrices of full rank. Assuming  $\mathbf{v}$  and  $\mathbf{e}$  are independently distributed with means  $\mathbf{0}$  and covariance matrices  $\mathbf{G}$  and  $\mathbf{R}$ , depending on variance parameters  $\delta = (\delta_1, \dots, \delta_q)^T$ , then the variance of  $\mathbf{y}$  is

$$\mathbf{V} = \mathbf{R} + \mathbf{Z}\mathbf{G}\mathbf{Z}^T. \quad (3.6)$$

---

If the gLMM is partitioned into  $K$  areas and covariance matrices  $\mathbf{G}$ ,  $\mathbf{R}$  and  $\mathbf{V}$  are set-up to be matrices of  $K$  diagonal matrices with block diagonal structure, the FH model is then derived. Similarly, a unit-level model can be derived from the gLMM. The FH and unit-level models like the gLMM assume continuous outcomes, that is  $y_{ik}$  are assumed to be continuous. In the case of binary and count data a different approach is needed.

Suppose that now  $y_{ik}$  is binary and thus we are now interested in the small area proportions  $\bar{Y}_k = P_k = \sum_{i=1}^{N_k} y_{ik}/N_k$ ,  $k = 1, \dots, K$ , where  $N_k$  is the number of individuals in area  $k$ . Assuming  $y_{ik}$ 's are independent Bernoulli( $p_{ik}$ ) with random area effects  $v_k$ , the model is then

$$\text{logit}(p_{ik}) = \log\left(\frac{p_{ik}}{1-p_{ik}}\right) = \mathbf{x}_{ik}^T \boldsymbol{\beta} + v_k, \quad (3.7)$$

where  $p_{ik} = \text{Prob}(y_{ik} = 1)$  and  $i = 1, \dots, N_k; k = 1, \dots, K$ .

Now we can further extend to count data by assuming that  $y_k$  are area-level counts and are independently distributed Poisson variables. The parameter of interest in model is  $\lambda_k$ , the true rate of the  $k^{th}$  area. Thus the model becomes

$$\log(\lambda_k) = \mathbf{x}_k^T \boldsymbol{\beta} + v_k, \quad (3.8)$$

where  $k = 1, \dots, K$ .

Note that (3.7) is a basic example of a logistic mixed model, in which like (3.4) unit specific auxiliary data is a logit function of  $p_{ik}$ . Similarly (3.8) is a basic example of a Poisson mixed model, in which covariates are a log function of response  $\lambda_k$ .

---

### 3.4.1 Small Area Parameters

In the context of small area estimation, the parameters of interest are the small area means represented by  $\mu_k$ . In general, to estimate small area means, a best predictor (BP) is used. The BP of a parameter of interest such as the small area mean is the conditional expectation of the parameter given the data and the model parameters. Thus, the best predictor assumes that the model parameters are known. In the context of the basic area model (3.3), the BP is

$$\mathbb{E}(\mu_k | \hat{\mu}_k, \boldsymbol{\beta}, \sigma_v^2) = \hat{\mu}_k^B = \gamma_k \hat{\mu}_k + (1 - \gamma_k) \mathbf{x}_k^T \boldsymbol{\beta}, \quad (3.9)$$

where  $\gamma_k = \sigma_v^2 b_k^2 / (\psi_k + \sigma_v^2 b_k^2)$ . In the case of the unit-level model (3.4), the BP is

$$\hat{\mu}_k^B = \bar{\mathbf{X}}_k^T \boldsymbol{\beta} + \tilde{v}_k, \quad \tilde{v}_k = \Gamma_k (\bar{y}_{ka} - \bar{\mathbf{x}}_{ka}^T \boldsymbol{\beta}), \quad (3.10)$$

where  $\Gamma_k = \sigma_v^2 / (\sigma_v^2 + \sigma_e^2/a_{k.})$ . The weighted means  $\bar{y}_{ka}$  and  $\bar{\mathbf{x}}_{ka}$  are given by

$$\bar{y}_{ka} = \sum_{i=1}^{n_k} a_{ik} y_{ik} / a_{k.}, \quad \bar{\mathbf{x}}_{ka} = \sum_{i=1}^{n_k} a_{ik} \mathbf{x}_{ik} / a_{k.},$$

in which  $a_{ik} = t_{ik}^{-2}$  and  $a_{k.} = \sum_{i=1}^{n_k} a_{ik}$ , for known constants  $t_{ik}$ .

Now in the case of when the response is binary and count values, the parameters of interest are now the proportions and rates  $P_k$  and  $\lambda_k$  respectively. We can use the same idea of the BP as in (3.9) to estimate  $P_k$  and  $\lambda_k$ , noting that in these cases, we do not have closed forms of the BP unlike the area-level and unit-level models in gLMM. Furthermore,  $\boldsymbol{\beta}$  and variance parameters are unknown and need to be estimated (Rao & Molina, 2015). Techniques of estimating the model parameters will now need be explored in the next section.

---

### 3.4.2 Model Parameter Estimation

There are two main approaches to estimate the model parameters (frequentist and Bayesian). Maximum likelihood estimation (MLE) is a popular frequentist method to estimate parameters of interest. The likelihood function is fundamental to both MLE and Bayesian approaches. The likelihood function is essentially probability distribution of the data given the model parameters are fixed. Assume that  $\boldsymbol{\theta}$  is a vector of model parameters of interest given any iid sample  $(x_1, y_1), (x_2, y_2), \dots, (x_K, y_K)$  then the definition of the likelihood function is

$$L(\boldsymbol{\theta} | \mathbf{x}, \mathbf{y}) = \prod_{k=1}^K f(y_k | x_k; \boldsymbol{\theta}) \quad (3.11)$$

The goal of MLE is to find the vector  $\boldsymbol{\theta}$  for which the function  $L(\boldsymbol{\theta} | \mathbf{x}, \mathbf{y})$  is maximized. This involves using calculus to differentiate  $L(\boldsymbol{\theta} | \mathbf{x}, \mathbf{y})$  with respects to the parameter  $\boldsymbol{\theta}$ . An alternative approach for inference is Bayesian. Note that in the MLE approach  $\boldsymbol{\theta}$  is considered a unknown fixed value, but now for Bayesian we consider  $\boldsymbol{\theta}$  as an unknown random variable. The posterior distribution is the probability distribution of the model parameters given the data. The connection between the two methods is through the Bayes rule which is defined as ,

$$p(\boldsymbol{\theta} | \mathbf{x}, \mathbf{y}) = \frac{L(\boldsymbol{\theta} | \mathbf{x}, \mathbf{y})p(\boldsymbol{\theta})}{p(\mathbf{x}, \mathbf{y})}.$$

Here  $p(\mathbf{x}, \mathbf{y}) = \int L(\boldsymbol{\theta} | \mathbf{x}, \mathbf{y})p(\boldsymbol{\theta})d\boldsymbol{\theta}$  is the probability of the data and is often called the evidence, since the probability represents the evidence, the data was generated by the model we are estimating from. Lastly  $p(\boldsymbol{\theta})$  represents the prior distribution of the model parameters.

Note that  $p(\mathbf{x}, \mathbf{y})$  is often considered as a constant since after integration, it is independent of the parameter of interest  $\boldsymbol{\theta}$ , hence the posterior is directly

---

proportional to the product of the likelihood and the prior distribution such that,

$$p(\boldsymbol{\theta}|\mathbf{x}, \mathbf{y}) \propto L(\boldsymbol{\theta}|\mathbf{x}, \mathbf{y})p(\boldsymbol{\theta}).$$

Calculating  $p(\boldsymbol{\theta}|\mathbf{x}, \mathbf{y})$  requires integrating over samples (small areas) which will require multi-dimensional integration and can be quite complex, thus a Monte Carlo Markov Chain (MCMC) numerical algorithm will be used to estimate the parameters of interest. Hierarchical Bayes (HB) maybe used with defining hyper parameters for model parameters.

There are well-established computational algorithms that can sample from the posterior directly without finding the exact analytical solution. The sampling technique of MCMC and programs such as WinBUGS and JAGS have made Bayesian estimation much more viable due to use of computing power. It should be kept in mind that Bayesian modeling needs a prior distribution  $p(\boldsymbol{\theta})$ , which can influence the parameter estimates. Careful consideration is needed in picking a prior to make sure the prior is not affecting the estimate more than the data itself.

MLE for our CHILD study is not convenient to implement due to the complexity of the likelihood function. Alternatively a Bayesian method (e.g., HB) will be appropriate. There are two components in the MCMC algorithm. Monte Carlo is technique for randomly sampling a probability distribution while a Markov chain is method of generating a sequence of random variables in which the current values are dependent on the prior value (Gamerman, 2006). Combining these two components results in MCMC and allows the random sampling of high-dimensional probability distributions. The key is that MCMC algorithms are sensitive to the starting point thus a burn-in phase is often used in which initial iterations are discarded. The idea is that the chain will settle on the desired posterior distribution, but the required number of iterations is unknown and must be determined. Diagnostic tools such as the Gelman-Rubin test can be used to determine if the model has converged to the target posterior distriubution and if it is stable. The Gelman–Rubin diagnostic

---

evaluates MCMC convergence by analyzing the differences between multiple Markov chains, small within-chain variance indicate convergence (Alkan, 2017). Thus in practice, multiple chains are sampled from, each with burn-in to ensure accurate results and to speed up convergence.

The MCMC method returns a distribution of possible effect values, the posterior  $p(\boldsymbol{\theta}|\mathbf{y})$ . Now the credible interval is the range containing a particular percentage of probable values. For example, the 95% credible interval is the portion of the posterior distribution that contains 95% of the values. Credible intervals summarise the uncertainty related to the unknown model parameters. Initially credible intervals seem similar to the frequentist Confidence Intervals. Credible intervals also allow us to use the interpretation “given the observed data, the effect (parameter) has 95% probability of falling within this range” (Kruschke, 2015). The interpretation of credible intervals are very different. Credible intervals are also straightforward to compute as compared to confidence intervals which have distributional assumptions and approximations. In practice, credible intervals can be used to test significance of model parameter estimates similar to confidence intervals. Software JAGS provides 95% credible intervals in its summary statistics output for convenience.

A reliable and well defined method for model comparison is often useful in the model building stages of analysis . Now when determining how well a model fits the data, a useful measure is the deviance which is defined as  $D(\boldsymbol{\theta}) = -2 \log p(\mathbf{y}|\mathbf{x}, \boldsymbol{\theta}) + 2 \log h(\mathbf{y})$  and is useful in both frequentist and Bayesian approaches (Chan & Grant, 2016). Here  $h(\mathbf{y})$  is a standardizing term that is a function of the data. Now the effective number of parameters is defined as  $p_D = \overline{D(\boldsymbol{\theta})} - D(\tilde{\boldsymbol{\theta}})$ , where  $\overline{D(\boldsymbol{\theta})} = -2\mathbb{E}_{\boldsymbol{\theta}}[\log p(\mathbf{y}|\mathbf{x}, \boldsymbol{\theta})|\mathbf{x}, \mathbf{y}] + 2 \log h(\mathbf{y})$  is the posterior mean deviance and  $\tilde{\boldsymbol{\theta}}$  is an estimate of  $\boldsymbol{\theta}$  (Chan & Grant, 2016). Then, the Deviance Information Criterion (DIC) is defined as

$$\text{DIC} = \overline{D(\boldsymbol{\theta})} + p_D,$$

hence DIC is often viewed as a trade-off between model adequacy and complexity

---

(Chan & Grant, 2016). JAGS can compute DIC statistics for complex Bayesian models and will be used throughout in the model building stages of this study.

We use the R package Rjags to implement the Bayesian method for our proposed model. All models are run on Rjags with the initial model having 200,000 iterations with thinning degree of 18. If the initial model does not converge according to the Gelman diagnostics for all covariates, then the model is updated each time with 10,000 iterations. The model can be updated until convergence is reached. All of the models that shall be presented henceforth are forced to converge with a Gelman-Rubin diagnostic ratio  $< 1.05$  for all model parameters. Additionally the 95% credible intervals for the posterior distributions are calculated using the equal tail interval method, which is the default method in Rjags.

### 3.5 Statistical Contribution

For the CHILD study, our interest is to determine which factors affect wheezing severity. Wheezing severity could be a binary variable or an ordinal variable. Furthermore, a model to count for the number of child's wheezing episodes is needed. Thus more complex models rather than the logit (3.7) and the Poisson (3.8) will be needed for this study.

Models mentioned so far are all cross-sectional, however, the CHILD cohort is longitudinal with a rich set of auxiliary data. There are time-series and cross-sectional models used in studies but even still there is a limited amount of research based on time series in the context of small-area estimation for non-normal responses. In general, the model can allow different link functions depending on the type of dependent variable. In the context of the CHILD study the response  $y_{ijk}$  is dependent on participant  $i$ , follow-up  $j$  and area  $k$ . Assuming without loss of generality the model has three covariates which can depend on any combination of  $i, j$  and  $k$  and

---

with response variable  $y_{ijk}$ . Then letting  $\mu_{ijk} = E(y_{ijk})$  we get,

$$g(\mu_{ijk}) = a_c + \beta_1 X_i + \beta_2 X_{ij} + \beta_3 X_{ijk} + f_{ij} + v_k + I_{ijk}. \quad (3.12)$$

In equation (3.12),  $a_c$  is level specific intercept with the restriction that  $a_1 < \dots < a_c$  such that  $c = 1, \dots, l - 1$ . Now  $f_{ij}$  is the random effect of the individual and follow-up. Usually this is the time specific disturbance but in the case of this model the disturbance is an AR(1) (autoregressive-one) model.  $I_{ijk}$  is the interaction random effect of all three levels interacting with one another and  $v_k$  is simply the random effect of each small area. Examples of  $X_i$ ,  $X_{ij}$  and  $X_{ijk}$  from our CHILD study include biological sex, SES and environmental factors, respectively.

From the model (3.12),  $g(\cdot)$  is the link function that connects  $y_{ijk}$  with the covariates and random components of the model. The link functions of interest for this study will be the logit and log, as shown below in equations (3.13-3.14). The logit link function in (3.13) is used when the response variable is assumed to follow a binomial, multinomial or ordinal distribution (e.g., a child wheezed or not in our CHILD study). While the log link function in (3.14) is primarily used when the response is assumed to follow a Poisson distribution (e.g., wheezing episodes in our CHILD study).

$$g(\mu_{ijk}) = \text{logit}(\mu_{ijk}) = \log \left( \frac{\mu_{ijk}}{1 - \mu_{ijk}} \right) \quad (3.13)$$

$$g(\mu_{ijk}) = \log(\mu_{ijk}) \quad (3.14)$$

Since the proposed model is complex and closed-formed solutions do not exist, one can use Bayesian methods for inference. Particular attention will be made to the choice of priors and structure of the random effects. Models of such kind will be a new contribution to the literature. An ordinal logistic model in which there are

---

space-time and individual level effects is of particular interest in this study.

### 3.5.1 Ordinal Logistic Model

A logistic model with the assumption that the response,  $y_{ijk}$ , is ordinal will be used to fulfil the first objective of this thesis which is to quantify how much of an impact location has on wheezing development. Here,  $y_{ijk}$  represents childhood wheezing severity in a particular area  $k$  at follow-up  $j$ . Further, a baseline model will be established to determine which covariates could be acting as proxy variables for location which will include covariates supported by literature .

Now we will describe the ordinal logit model, if such model is possible to fit with the CHILD dataset. Let  $y_{ijk}$  have three levels  $l = 1, 2, 3$  thus there are two categories  $c = 1, 2$  (with another level as a reference category), we then get

$$\log \left[ \frac{\text{Prob}(y_{ijk} \leq c)}{\text{Prob}(y_{ijk} > c)} \right] = a_c + \beta_1 X_i + \beta_2 X_{ij} + \beta_3 X_{ijk} + f_{ij} + v_k + I_{ijk}, \quad (3.15)$$

where  $i$  represents the individual,  $j$  is follow-up, and  $k$  is small area and

$$\begin{aligned} v_k &\sim \mathcal{N}(0, \sigma_v^2) \\ f_{ij} &= \rho f_{ij-1} + \epsilon_{ij}, \epsilon_{ij} \sim \mathcal{N}(0, \sigma_f^2) \\ f_{ij} \mid f_{ij-1} &\sim \mathcal{N}(\rho f_{ij-1}, \sigma_f^2) \\ I_{ijk} &\sim \mathcal{N}(0, \sigma_I^2) \\ -1 < \rho < 1. \end{aligned}$$

Thus the two logit equations with  $Q(m) = \text{Prob}(y_{ijk} \leq c)$  are

$$\begin{aligned} \text{logit}[Q(1)] &= a_1 + \beta_1 X_i + \beta_2 X_{ij} + \beta_3 X_{ijk} + f_{ij} + v_k + I_{ijk} \\ \text{logit}[Q(2)] &= a_2 + \beta_1 X_i + \beta_2 X_{ij} + \beta_3 X_{ijk} + f_{ij} + v_k + I_{ijk}. \end{aligned}$$

---

### 3.5.2 Zero-Inflated Poisson Model

So far, our response variable wheezing severity has been binomial or ordinal in this study. Now to use count data from the CHILD study a different model is needed. Now a Poisson regression model could be used to model the number of wheezing episodes as the response variable along with covariates that are statistically significant or known from literature to cause increased wheezing episodes in children. However, there are some problems with this approach. From Table 3.15 it is clear that there are many children with zero episodes of wheezing or excess zeros when compared with the number of children with wheezing episodes greater than one. Furthermore standard Poisson regression assumes that the variance of the count outcome, in this case wheezing episodes, is equal to the mean. Real data however can violate this assumption and this is referred to as overdispersion. Zero-Inflated Poisson (ZIP) models can allow for excess zeros in the dependent variable and overdispersion.

The ZIP model will allow for overdispersion assuming that there are two different types of participant in the CHILD study: Those who have a zero count with a probability of one (Always-zero), and children with counts predicted by the standard Poisson (Not always-zero). This can also be referred to as “true” zeros and “excess” zeroes. The ZIP model is a mixture of probabilities from the two groups, which allows for both the overdispersion and excess zeros that cannot be modeled by the conventional Poisson regression model. We now introduce area and follow-up level effects to ZIP model similar to the logistic model introduced in section 3.5.1. However, to reduce the number of parameters  $I_{ijk}$  is not introduced in the model. Assuming, without loss of generality, the model has three covariates as follows:

$$\Theta_{ijk} = \beta_0 + \beta_1 X_i + \beta_2 X_{ij} + \beta_3 X_{ijk} + f_{ij} + v_k, \quad (3.16)$$

where

$$\begin{aligned} v_k &\sim \mathcal{N}(0, \sigma_v^2) \\ f_{ij} &= \rho f_{ij-1} + \epsilon_{ij}, \epsilon_{ij} \sim \mathcal{N}(0, \sigma_f^2) \\ f_{ij} \mid f_{ij-1} &\sim \mathcal{N}(\rho f_{ij-1}, \sigma_f^2) \\ -1 < \rho &< 1. \end{aligned}$$

Then the ZIP model becomes

$$\text{Prob}(y_{ijk} = m) = \begin{cases} \pi_{ijk} + (1 - \pi_{ijk}) \exp(-\lambda_{ijk}) & \text{if } m = 0 \\ (1 - \pi_{ijk}) \frac{\lambda_{ijk}^{y_{ijk}} \exp(-\lambda_{ijk})}{y_{ijk}!} & \text{if } m > 0 \end{cases} \quad (3.17)$$

such that  $\log(\lambda_{ijk}) = \Theta_{ijk}$  and  $\text{logit}(\pi_{ijk}) = \Theta_{ijk}$ .

Then the expected value of the ZIP model is

$$E(y_{ijk}) = 0 + \sum_{m=1}^{\infty} m(1 - \pi_{ijk})e^{-\lambda_{ijk}} \frac{\lambda_{ijk}^m}{m!} = (1 - \pi_{ijk})e^{-\lambda_{ijk}} \sum_{m=1}^{\infty} m \frac{\lambda_{ijk}^m}{m!}.$$

Now let  $t = m - 1$  and then,

$$\begin{aligned} \sum_{m=1}^{\infty} m \frac{\lambda_{ijk}^m}{m!} &= \sum_{m=1}^{\infty} \frac{\lambda_{ijk}^m}{(m-1)!} \\ &= \sum_{m=1}^{\infty} \frac{\lambda_{ijk} \lambda_{ijk}^{m-1}}{(m-1)!} \\ &= \lambda_{ijk} \sum_{t=0}^{\infty} \frac{\lambda_{ijk}^t}{t!} \\ &= \lambda_{ijk} e^{\lambda_{ijk}}. \end{aligned}$$

Thus,  $E(y_{ijk}) = (1 - \pi_{ijk})\lambda_{ijk}$ .

Similarly the variance of the ZIP model is  $\text{Var}(y_{ijk}) = E(y_{ijk}^2) - [E(y_{ijk})]^2$  such

---

that

$$\begin{aligned} E(y_{ijk}^2) &= (1 - \pi_{ijk})\lambda_{ijk}e^{-\lambda_{ijk}} \sum_{t=0}^{\infty} (t+1)\frac{\lambda^t}{t!} \\ &= (1 - \pi_{ijk})(\lambda_{ijk}^2 + \lambda_{ijk}). \end{aligned}$$

Then,

$$Var(y_{ijk}) = (1 - \pi_{ijk})(\lambda_{ijk}^2 + \lambda_{ijk}) - (1 - \pi_{ijk})^2 \lambda_{ijk}^2 = \lambda_{ijk}(1 - \pi_{ijk})(1 + \pi_{ijk}\lambda_{ijk}).$$

To summarize,

$$\begin{aligned} E(y_{ijk}) &= (1 - \pi_{ijk})\lambda_{ijk} \\ \text{Var}(y_{ijk}) &= \lambda_{ijk}(1 - \pi_{ijk})(1 + \pi_{ijk}\lambda_{ijk}) \end{aligned} \tag{3.18}$$

In ZIP models, it is possible to choose different predictors for the true and excess zeros if it is expected different variables to be driving presence or absence of the outcome in the data. For simplicity the same covariates in both cases of ZIP model are considered as shown by in the expression,  $\Theta_{ijk}$ . Typically however we are less interested in the modeling of false zeros thus a diffuse Bayesian prior for  $\pi_{ijk}$  will be implemented such that  $logit(\pi_{ijk}) = logit(\pi) = \gamma_0$ ,  $\gamma_0 \sim \mathcal{N}(0, 10000)$ .

### 3.5.3 Area-Level Proportions and Means

In Sections 3.5.1 and 3.5.2, we were interested to study the effects of covariates over time in our longitudinal set-ups. In addition, we are also interested to study which areas (regions) are more at risk in terms of wheezing episodes and also its severity. Our parameters are now small area proportions or means. To that end, in the Bayesian framework, following Rao and Molina (2015, pg. 408), we consider the unit-level binary logit model as

- 
- (i)  $y_{ik} \mid p_{ik} \sim \text{Bernoulli}(p_{ik})$ ,
- (ii)  $\text{logit}(p_{ik}) = \mathbf{x}_{ik}^T \boldsymbol{\beta} + v_k$ , with  $v_k \stackrel{\text{iid}}{\sim} N(0, \sigma_v^2)$  and priors,
- (iii)  $f(\boldsymbol{\beta}) \propto 1$  and  $\sigma_v^{-2} \sim G(0.001, 0.001)$ ,

where  $y_{ik}$  is e.g. whether child  $i$  at area  $k$  wheezed or not at the first follow-up.

Now consider the case that the covariates  $\mathbf{x}_{ik}$  are at area level such that  $\mathbf{x}_{ik} = \mathbf{x}_k$ .

The area-level logit model is described as

- (i)  $y_k \mid p_k \sim \text{Binomial}(p_k, n_k)$ ,
- (ii)  $\text{logit}(p_k) = \mathbf{x}_k^T \boldsymbol{\beta} + v_k$ , with  $v_k \stackrel{\text{iid}}{\sim} N(0, \sigma_v^2)$  and priors,
- (iii)  $f(\boldsymbol{\beta}) \propto 1$  and  $\sigma_v^{-2} \sim G(0.001, 0.001)$ ,

where  $y_k$  is the number of children who wheezed at area  $k$ . Now the parameter of interest is  $p_k = h(\mathbf{x}_k^T \boldsymbol{\beta} + v_k)$ , where  $h(t) = 1/\{1 + \exp(-t)\}$ . Since the true population parameters are not known they can be estimated using the logit-model, then we get  $\hat{p}_k^{HB} = h(\mathbf{x}_k^T \hat{\boldsymbol{\beta}} + \hat{v}_k)$ . Here  $\hat{\boldsymbol{\beta}}$  is the vector of estimated fixed effects and  $\hat{v}_k$  is predicted area-specific random effect. Note that  $\hat{p}_k^{HB}$  represents the probability or proportion of wheezing in area  $k$ .

Similarly for the ZIP model, our interest is the average count per area. For instance, we can restrict our ZIP model to one follow-up. The model is similar to the previously mentioned ZIP model in equations (3.15-3.16) and is described as

$$\Theta_{ik} = \mathbf{x}_{ik}^T \boldsymbol{\beta} + v_k \tag{3.19}$$

where  $\mathbf{x}_{ik}$  are unit-level covariates and  $v_k \sim \mathcal{N}(0, \sigma_v^2)$ .

---

Then, the ZIP model becomes

$$\text{Prob}(y_{ik} = m) = \begin{cases} \pi_{ik} + (1 - \pi_{ik}) \exp(-\lambda_{ik}) & \text{if } m = 0 \\ (1 - \pi_{ik}) \frac{\lambda_{ik}^{y_{ik}} \exp(-\lambda_{ik})}{y_{ik}!} & \text{if } m > 0 \end{cases} \quad (3.20)$$

such that  $\log(\lambda_{ik}) = \Theta_{ik}$  and  $\text{logit}(\pi_{ik}) = \Theta_{ik}$ . Recall that we are typically less interested in the modelling of false zeros thus a diffuse prior for  $\pi_{ik}$  will be implemented such that  $\text{logit}(\pi_{ik}) = \text{logit}(\pi) = \gamma_0$ ,  $\gamma_0 \sim \mathcal{N}(0, 10000)$ . Now consider the simple case with the covariates  $\mathbf{x}_{ik}$  are at area-level such that  $\mathbf{x}_{ik} = \mathbf{x}_k$ . Then, the area-level ZIP model is

$$\Theta_k = \mathbf{x}_k^T \boldsymbol{\beta} + v_k, \quad (3.21)$$

where  $\mathbf{x}_k$  are area-level covariates and  $v_k \sim \mathcal{N}(0, \sigma_v^2)$ .

Then,

$$\text{Prob}(y_k = m) = \begin{cases} \pi_k + (1 - \pi_k) \exp(-\lambda_k) & \text{if } m = 0 \\ (1 - \pi_k) \frac{\lambda_k^{y_k} \exp(-\lambda_k)}{y_k!} & \text{if } m > 0 \end{cases} \quad (3.22)$$

Now our parameter of interest is  $\lambda_k = \exp(\mathbf{x}_k^T \boldsymbol{\beta} + v_k)$ . Using the parameter estimates of the ZIP model this becomes  $\hat{\lambda}_k^{HB} = \exp(\mathbf{x}_k^T \hat{\boldsymbol{\beta}} + \hat{v}_k)$ . Here,  $\hat{\lambda}_k^{HB}$  represents the average number of wheezing episodes in area  $k$ .

# Chapter 4: Simulation Study

## 4.1 Ordinal Logistic Model

A simulation study is performed in which data was simulated with an ordinal response; the response was represented the wheezing severity scale found on the CHILD study questionnaires. The dependent variable or response variable  $y_{ijk}$  was randomly generated with 1600 random values all ranging from 1-3 using the integer uniform distribution or trinomial distribution. The probability of generating each integer 1, 2, 3 is  $\frac{1}{3}$ , thus equal probabilities. The aim of the simulation is to verify that the estimates produced by an ordinal model with data like the CHILD cohort would be viable. For the simulation there are 20 individuals, 4 follow-ups and 20 small areas. The covariates  $X_i, X_{ij}$  and  $X_{ijk}$  are generated from binary distribution with probability of generating each integer 0, 1 being equal to  $\frac{1}{2}$ . The simulation is conducted by fitting the simulated data with the model in equation (3.15) using Rjags. Diffuse priors are used for all the model parameters. Coefficients from the fitted model are then to be used as true parameters since a starting point is needed to begin the simulations.

Parameter	Estimate
$a_1$	-0.685
$a_2$	0.772
$\beta_1$	-0.090
$\beta_2$	0.068
$\beta_3$	0.170
$\rho$	0.378
$\sigma_v$	0.145
$\sigma_f$	0.162
$\sigma_I$	0.169

Table 4.1: Ordinal Simulation Study: Three chains are used with 60,000 samples from each chain. Furthermore, the Gelman diagnostics indicated convergence of model parameters. These estimates are to be used as the true parameters for the ordinal logit model.

---

We now use these estimates as true parameters for our simulation study. The initial goal is to evaluate performance of the model parameters. The simulation procedure is as follows:

Algorithm A:

Step 1: Consider the estimates from Table 4.1 as the true parameters, represented by the vector  $\boldsymbol{\delta}$

Step 2: Use the true parameters to generate the random effects which in turn generate the new response values

Step 3: Refit the model using the new generated response values

Step 4: Extract the estimates from the newly fitted model

Step 5: Repeat the cycle by going to step 2

In the above steps the model is essentially reversed in the sense that the aim was to simulate values of the response  $y_{ijk}$  using the ordinal model described before and thus replacing the original values of  $y_{ijk}$  with the simulated values of  $y_{ijk}$  and then refitting the same ordinal model; this was done 500 times. Each simulation run produced it's own estimates ( $\hat{\boldsymbol{\delta}}_r; r = 1, \dots, 500$ ) of the model parameters. Then the bias, relative bias (R.Bias), and MSE (Mean Square Error) were calculated using the three formulas below, assuming element wise operation of vectors (Table 4.2). The bias and relative bias are used to quantify the average difference between the true parameters and the estimates within the simulations runs. The MSE is used to assess, on average, how far an estimator is from the true parameter within the simulation runs.

---


$$\text{Bias} = \frac{1}{500} \sum_{r=1}^{500} (\boldsymbol{\delta} - \hat{\boldsymbol{\delta}}_r)$$

$$\text{Relative Bias}(\%) = \frac{1}{500} \sum_{r=1}^{500} \frac{(\boldsymbol{\delta} - \hat{\boldsymbol{\delta}}_r)}{\boldsymbol{\delta}} \times 100$$

$$\text{MSE} = \frac{1}{500} \sum_{r=1}^{500} (\boldsymbol{\delta} - \hat{\boldsymbol{\delta}}_r)^2$$

Parameter	True	Bias	R.Bias(%)	MSE
$a_1$	-0.685	-0.022	3.21	0.037
$a_2$	0.772	0.010	1.24	0.038
$\beta_1$	-0.090	0.003	-3.33	0.011
$\beta_2$	0.068	-0.002	-2.94	0.009
$\beta_3$	0.170	-0.010	-5.88	0.009
$\rho$	0.378	-0.019	-4.95	0.015
$\sigma_v$	0.145	0.028	19.24	0.015
$\sigma_f$	0.162	0.011	6.48	0.001
$\sigma_I$	0.169	-0.006	-3.55	0.001

Table 4.2: Bias, relative bias, and MSE of model parameter estimates based on 500 simulation runs

The results from Table 4.2 indicate the model parameter estimates have very small bias and MSE. The relative bias for  $\sigma_v$  is higher than the other parameters but overall is acceptable. A possible explanation is that the number of small areas may not be enough in this simulation set-up to reflect the performance of small area variation in the proposed model. Thus, the results indicate that the estimated parameters are working well in the proposed model. To further evaluate the proposed model, probabilities for each category can be compared to each other to determine if the model fits according to the true distribution of the data (Appendix B).

Further simulation studies are performed which assume the responses are binary and count data , thus logit and log link functions will be used as shown in equation (3.12).

---

## 4.2 Binary Logistic Model

The model is described as

$$y_{ijk} \mid p_{ijk} \sim \text{Bernoulli}(p_{ijk})$$

where  $y_{ijk}$  is the response of wheezing for individual  $i$ , follow-up  $j$ , and small area  $k$ . Then  $p_{ijk} = \text{Prob}(y_{ijk} = 1)$  is the probability that the individual  $i$  has a wheezing at follow-up  $j$  and small area  $k$ . Now the logit model is

$$\log \left[ \frac{p_{ijk}}{1 - p_{ijk}} \right] = \beta_0 + \beta_1 X_i + \beta_2 X_{ij} + \beta_3 X_{ijk} + f_{ij} + v_k \quad (4.1)$$

where

$$\begin{aligned} v_k &\sim \mathcal{N}(0, \sigma_v^2) \\ f_{ij} &= \rho f_{ij-1} + \epsilon_{ij}, \epsilon_{ij} \sim \mathcal{N}(0, \sigma_f^2) \\ f_{ij} \mid f_{ij-1} &\sim \mathcal{N}(\rho f_{ij-1}, \sigma_f^2) \\ -1 < \rho &< 1. \end{aligned}$$

Now  $f_{ij}$  is the random effect of the individual and follow-up. Usually this is the time specific disturbance but in the case of this model the disturbance is an autoregressive-one or AR(1) model and  $v_k$  is simply the random effect of each small area. A simulation study is performed in which data was simulated with a binary response; the response represents whether each child within the cohort wheezed or not. The aim of the simulation is to verify that the estimates produced by an binomial logit model with data representing the CHILD cohort would be viable. For the simulation there are 20 individuals, 4 follow-ups and 20 small areas. The dependent variable  $y_{ijk}$  is randomly generated with 1600 random binary values from a bernoulli distribution. The probability of generating each integer 0, 1 is  $\frac{1}{2}$ , thus equal probabilities. The covariates  $X_i, X_{ij}$  and  $X_{ijk}$  are also generated from binary distribution with probability of generating each integer 0, 1 being equal to  $\frac{1}{2}$ . The

---

estimated parameters are reported in Table 4.3.

Parameter	Estimate
$\beta_0$	-0.298
$\beta_1$	-0.118
$\beta_2$	0.270
$\beta_3$	0.0168
$\rho$	0.0117
$\sigma_v$	0.111
$\sigma_f$	0.128

Table 4.3: Binary Logit Simulation Study: Three chains are used with 60,000 samples from each chain. Furthermore, the Gelman diagnostics indicated convergence of model parameters. These estimates are to be used as the true parameters for the binary logit model.

The goal is now to evaluate performance of the model parameters in the proposed logit model. The simulation procedure is as followed from Algorithm A. In the simulation steps the model is essentially reversed in the sense that the aim was to simulate values of the response  $y_{ijk}$  using the ordinal model described before and thus replacing the original values of  $y_{ijk}$  with the simulated values of  $y_{ijk}$  and then refitting the same binary logit model; this was done 500 times. Each simulation run produced it's own estimates  $(\hat{\delta}_r; r = 1, \dots, 500)$  of the model parameters. Then the bias, relative bias, and MSE are calculated (Table 4.4).

Parameter	True	Bias	R.Bias(%)	MSE
$\beta_0$	-0.298	0.009	-3.00	< 0.0001
$\beta_1$	-0.118	-0.002	-1.80	< 0.0001
$\beta_2$	0.270	-0.001	-0.41	< 0.0001
$\beta_3$	0.0168	-0.0002	-8.40	< 0.0001
$\rho$	0.0117	0.0025	22.70	< 0.0001
$\sigma_v$	0.111	0.0005	0.47	< 0.0001
$\sigma_f$	0.128	0.0006	0.62	< 0.0001

Table 4.4: Bias, relative bias, and MSE of model parameter estimates based on 500 simulation runs from the proposed logit model

The results from Table 4.4 indicate the model parameter estimates have small bias and MSE. The relative bias for  $\rho$  is higher than the other parameters but is still acceptable. Note that the model is sensitive to the number of follow-ups in

---

the dataset. Ultimately the number of areas, follow-ups and individuals chosen will dictate how well the simulation performs. Thus, the overall results indicate that the estimated parameters are working well in the proposed logit model.

### 4.3 Zero-Inflated Poisson Model

We now simulate the ZIP model with the response being number of wheezing episodes. We consider the same covariates as discussed in the logit model. For this simulation we shall assume we are less interested in the modelling of false zeros therefore a diffuse Bayesian prior for  $\pi_{ijk}$  will be implemented such that  $\text{logit}(\pi_{ijk}) = \text{logit}(\pi) = \gamma_0$ ,  $\gamma_0 \sim \mathcal{N}(0, 10000)$ . Again for the simulation there are 20 individuals, 4 follow-up,s and 20 small areas. The dependent variable  $y_{ijk}$  is randomly generated with 1600 random values from ZIP distribution with  $\pi = 0.4$  and  $\lambda = 4$  which gives the range of possible values for  $y_{ijk}$  between 0-13. The covariates  $X_i, X_{ij}$  and  $X_{ijk}$  are also generated from binary distribution with probability of generating each integer 0, 1 being equal to  $\frac{1}{2}$ . The statistical measures used to test the credibility of the models are bias and MSE. Now let

$$\Theta_{ijk} = \beta_0 + \beta_1 X_i + \beta_2 X_{ij} + \beta_3 X_{ijk} + f_{ij} + v_k \quad (4.2)$$

where

$$\begin{aligned} v_k &\sim \mathcal{N}(0, \sigma_v^2) \\ f_{ij} &= \rho f_{ij-1} + \epsilon_{ij}, \epsilon_{ij} \sim \mathcal{N}(0, \sigma_f^2) \\ f_{ij} \mid f_{ij-1} &\sim \mathcal{N}(\rho f_{ij-1}, \sigma_f^2) \\ -1 < \rho &< 1. \end{aligned}$$

Then the ZIP model becomes,

$$\text{Prob}(y_{ijk} = m) = \begin{cases} \pi + (1 - \pi) \exp(-\lambda_{ijk}) & \text{if } m = 0 \\ (1 - \pi) \frac{\lambda_{ijk}^{y_{ijk}} \exp(-\lambda_{ijk})}{y_{ijk}!} & \text{if } m > 0 \end{cases} \quad (4.3)$$

---

such that  $\log(\lambda_{ijk}) = \Theta_{ijk}$  and  $\text{logit}(\pi) = \gamma_0$  where  $\gamma_0 \sim \mathcal{N}(0, 10000)$ .

Parameter	Estimate
$\beta_0$	1.085
$\beta_1$	-0.034
$\beta_2$	0.066
$\beta_3$	0.085
$\gamma_0$	-0.304
$\rho$	0.657
$1 - \pi$	0.575
$\sigma_v$	0.639
$\sigma_f$	0.042

Table 4.5: ZIP Simulation Study: Three chains are used with 60,000 samples from each chain. Furthermore, the Gelman diagnostics indicated convergence of model parameters. These estimates are to be used as the true parameters for the ZIP model.

Then, the simulation procedure follows Algorithm A. Similar to prior simulations the model is essentially reversed. Each simulation run produced it's own estimates  $(\hat{\delta}_r; r = 1, \dots, 500)$  of the model parameters. Then the bias, relative bias, and MSE are calculated (Table 4.6).

Parameter	True	Bias	R.Bias(%)	MSE
$\beta_0$	1.085	0.270	24.9	0.252
$\beta_1$	-0.034	-0.003	8.40	< 0.0001
$\beta_2$	0.066	0.002	3.30	< 0.0001
$\beta_3$	0.085	-0.003	-3.50	< 0.0001
$\gamma_0$	-0.304	< 0.0001	0.10	< 0.0001
$\rho$	0.657	-0.224	-34.10	0.194
$1 - \pi$	0.575	< 0.0001	< 0.0001	< 0.0001
$\sigma_v$	0.639	-0.210	-34.00	0.355
$\sigma_f$	0.042	< 0.0001	0.30	< 0.0001

Table 4.6: Bias, relative bias, and MSE of model parameter estimates based on 500 simulation runs from the proposed ZIP model.

The results from Table 4.6 indicate the model parameter estimates behave reasonably well however some parameters have relative bias above 30 percent which after further investigation is due to  $\rho$  not be statistically significant for this particular simulated dataset. When using such model with real data, particular attention

---

should be placed on the AR(1) and should be removed if not significant.

# Chapter 5: CHILD Study: Data Analysis

## 5.1 Logistic Model

Due to the complexity of the ordinal model in equation (3.15) and lack of viable data, the response variable severity of wheezing is transformed to run a binary logistic model by merging responses 2 and 3 into one level since the numbers in these two levels are small compared to the level 1. After cleaning and imputation, the dataset follows  $I = 690$  individuals,  $K = 60$  small areas and  $J = 6$  follow-ups.

The model is described as

$$y_{ijk} \mid p_{ijk} \sim \text{Bernoulli}(p_{ijk})$$

where  $y_{ijk}$  is the response of wheezing severity for individual  $i$ , follow-up  $j$ , and small area  $k$ . Then  $p_{ijk} = \text{Prob}(y_{ijk} = 1)$  is the probability that the individual  $i$  has a wheezing severity at follow-up  $j$  and small area  $k$ . Now the logistic model is

$$\begin{aligned} \log \left[ \frac{p_{ijk}}{1 - p_{ijk}} \right] &= \beta_0 + \beta_1 \text{MW}_{ik} + \beta_2 \text{MS}_{ik} + \beta_3 \text{FW}_{ik} \\ &\quad + \beta_4 \text{FS}_{ik} + \beta_5 \text{AY}_{ik} + \beta_6 \text{AO}_{ik} \\ &\quad + \beta_7 \text{MA}_{ik} + \beta_8 \text{DMA}_{ik} \\ &\quad + \beta_9 \text{MTFAOW}_{ik} + \beta_{10} \text{FA}_{ik} \\ &\quad + \beta_{11} \text{DFA}_{ik} + \beta_{12} \text{FTFAOW}_{ik} \\ &\quad + \beta_{13} \text{FurryPets}_{ijk} + \beta_{14} \text{SE}_{ik} + f_{ij} + v_k, \end{aligned} \tag{5.1}$$

---

where

$$\begin{aligned}
 v_k &\sim \mathcal{N}(0, \sigma_v^2) \\
 f_{ij} &= \rho f_{ij-1} + \epsilon_{ij}, \epsilon_{ij} \sim \mathcal{N}(0, \sigma_f^2) \\
 f_{ij} \mid f_{ij-1} &\sim \mathcal{N}(\rho f_{ij-1}, \sigma_f^2) \\
 -1 < \rho &< 1.
 \end{aligned}$$

First the random effects model is run to get a general representation of the variability within the model without the covariates.

Parameter	Estimate	95% credible interval (lower, upper)
$\rho$	0.06	(0.03,0.11)
$\sigma_v^2$	98.30	(94.10,99.90)
$\sigma_f^2$	91.60	(82.20,99.50)

Table 5.1: Binary logistic random effects model.

Table 5.1 shows the variance components estimate of the random effects model. From Table 5.1, the area-specific random effect is statistically significant as shown by the 95% credible interval of  $\hat{\sigma}_v^2$ . Similarly, the AR(1) variance components  $\hat{\sigma}_f^2$  and  $\hat{\rho}$  are also statistically significant. Thus, we can interpret this as there is a difference of wheezing severity between small areas. The significance of the AR(1) model indicates that there is some dependence of wheezing severity over follow-ups. This result is desired since it is intuitively expected that previous history of wheezing would result in wheezing in future follow-ups.

We use the bottom-up approach for model building. Thus, now univariate analysis is performed on covariates that are considered important according to the literature. The previously mentioned Bernoulli Bayesian model is used with one covariate one at a time, and thus only  $\beta_1$  and the variance components are estimated by RJags. The results of the individual univariate models can be found in Appendix C.

From the results of the univariate analysis, we then add significant variables

---

one by one based upon the statistical significance. The goal is to build a model step by step and keep track of the impact each covariate has on the overall model. Note that we assume that all the covariates considered above are clinically important. Note that the model building steps are shown in Appendix D with summary statistics. Based upon the results of the model building process, the final model is selected and the odds ratio is calculated for each covariate (Table 5.2) [Including mother wheezed, mother asthma and/or diagnosed mother asthma in the same model could introduce some co-linearity to the model likewise for asthma young and asthma older. However, for now these variables will be kept in the baseline model but may be removed if more covariates like sex of the child at birth are added, depending on how the variables interact with each other].

Parameter	Estimate	95% credible interval(lower, upper)	Odds Ratio
Intercept	-187.10	(-202.40, -172.20)	-
Asthma at 3 years	66.00	(46.87, 86.89)	$4.80 \times 10^{28}$
Asthma at 5 years	27.10	(9.74, 44.89)	$6.03 \times 10^{11}$
Mother wheezed (3 months)	35.80	(15.74, 56.31)	$3.40 \times 10^{15}$
Diagnosed mother asthma	19.80	(0.72, 43.07)	$3.89 \times 10^8$
Farm	1.90	(-24.17, 27.46)	6.57
Furry pets	4.00	(-11.63, 17.29)	54.00
$\rho$	0.04	(0.01, 0.08)	-
$\sigma_v^2$	17.70	(2.40, 31.86)	-
$\sigma_f^2$	98.60	(9.47, 99.97)	-

Table 5.2: Final logistic model based upon model building steps with intercept and random effects. The model has six covariates; odds ratios are calculated by taking the exponent of the covariate parameter estimates. Credible intervals are provided for the parameter estimates.

The model parameters related to the covariates are all statistically significant except for Farm and Furry pets. However, Farm and Furry pets are kept in the model due to their clinical importance. It should be mentioned that Asthma at 3 and 5 years are only included to test if the model is working properly since there is high correlation between asthma and wheezing thus these covariates are expected to be significant. Asthma at 3 and 5 years will be removed further into the analysis. All

---

the covariates from their respective odds ratio indicate that their presence increases the odds of developing wheezing versus no wheezing. The odds ratios are very high which may be because of the limited number of wheezing cases present within the CHILD sub-cohort. In order to determine why the odds ratios are so over inflated we could look again at the data. Now we can find  $p_{ijk} = \text{Prob}(y_{ijk} = 1)$  from the severity wheezing binary variable, which is cleaned and imputed as shown in Tables 3.13 and 3.14, we calculate  $p_{.j.}$  which is the probability that an individual in the study at follow-up  $j$  develops wheezing severity. Note that  $p_{.j.}$  is not individual or area dependent and simply sums across the hierarchical levels.

Parameter	Follow-up	Estimate
$p_{.1.}$	3 months	$\frac{39}{629} = 0.06$
$p_{.2.}$	6 months	$\frac{62}{628} = 0.10$
$p_{.3.}$	1 year	$\frac{73}{617} = 0.12$
$p_{.4.}$	18 months	$\frac{67}{623} = 0.11$
$p_{.5.}$	2 years	$\frac{78}{612} = 0.13$
$p_{.6.}$	2.5 years	$\frac{58}{632} = 0.09$

Table 5.3: Probabilities of wheezing severity for all six follow-ups.

In particular,

$$p_{.j.} = \frac{\sum_{i=1}^I \sum_{k=1}^K y_{ijk}}{\sum_{i=1}^I \sum_{k=1}^K n_{ijk}},$$

where  $n_{ijk}$  is the total participants for individual  $i$ , follow-up  $j$  and small area  $k$ . Then summing across all follow-ups in Table 5.3

$$p_{...} = \frac{\text{children with wheezing severity}}{\text{children in the cohort}} = \frac{286}{3854} = 0.07.$$

Now when looking at the intercept coefficient in Table 5.2 we can see that intercept is the only negative number in the model and greater in absolute value than the other coefficients. This indicates that on average there are a great number of

---

individuals in the study with no wheezing severity. That could over-inflate the odds of the covariates when using complex models with multiple parameters. If a child developed asthma or if there is an evidence of maternal asthma and wheezing, there is an increase in odds of the child developing wheezing. Interestingly, if a participant in the study had furry pets or lived near a farm, there seems to an increase in the odds of developing wheezing. The large number of zero cell counts of wheezing may be the result of the large coefficient estimates and will need to be investigated.

Although the model selected has statistically significant variables we need to make sure that all the significant covariates based on the literature are added such as maternal smoking and asthma at 18 weeks, since it is known from studies to be contributor to childhood wheezing and asthma. At the same time some covariates need to be removed in order to reduce the co-linearity in the model. Thus Asthma older will be removed along with diagnosed Mother asthma and Mother wheezed. The goal is to make a baseline model which confirms the literature but can also be easily interpreted, as the extreme values for the odds ratios are cause for concern which could be due to possible co-linearity. In an effort to reduce the high odds ratios, more models are introduced still using all six follow-ups of the two-stage wheezing severity variable.

Parameter	Estimate	95% credible interval(lower, upper)	Odds Ratio
Intercept	-176.00	(-191, -161 )	-
Asthma at 3 years	34.10	(23.1, 45.10 )	$6.77 \times 10^{14}$
Mother asthma (prenatal)	39.80	(24.0, 55.20 )	$1.94 \times 10^{17}$
Mother smoked (prenatal)	13.50	(-20.20 ,42.40)	$7.25 \times 10^{05}$
Farm	1.12	(-22.30, 23.20 )	3.05
Furry pets	2.82	(-13.90, 16.00)	168
$\rho$	0.05	(0.02, 0.09 )	-
$\sigma_v^2$	18.10	(3.18, 33.00 )	-
$\sigma_f^2$	98.70	(94.80, 100.00 )	-

Table 5.4: Logistic model with prenatal covariates, based upon model building steps with intercept and random effects.

---

The model in Table 5.4 is similar to the model in Table 5.2 in terms of performance of covariates, and their respective odds ratios indicate that their presence increases the odds of developing wheezing versus no wheezing. The prenatal covariate of maternal asthma is statistically significant while prenatal smoking is not. However, the odds ratios are still very high for most of the covariates, this could be due to the high estimate of  $\sigma_f^2$ . Removing  $\sigma_f^2$  from the model may reduce the over inflated odds ratios and make a simpler model with only area-level random effects.

In order to investigate if the AR(1) effect should be removed, a model similar to Equation (5.1) but with two follow-ups is considered. To further investigate and find the best model for the CHILD dataset, we consider the following covariates: if mother had asthma and smoked (prenatal and 1-year follow-ups), if the child lived near a farm (3-months and 1-year follow-ups) and furry pets (3-months and 1-year follow-ups). The reason that these time points are chosen is to try and make a model which would predict wheezing before the age of two and also considering the availability of the data.

Parameter	Estimate	95% credible interval(lower, upper)	Odds Ratio
Intercept	-300.00	( -405.00, -167.00 )	-
Mother asthma	103.00	( 26.2, 167.00 )	$4.78 \times 10^{44}$
Mother smoked	28.80	( -59.8, 117.00 )	$3.25 \times 10^{12}$
Farm	-5.96	( -97.5, 63.70 )	$2.59 \times 10^{-3}$
Furry pets	-2.12	( -56.1, 39.90 )	0.12
$\rho$	-0.10	( -0.20, 0.03 )	-
$\sigma_v^2$	78.00	( 20.5, 99.80 )	-
$\sigma_f^2$	98.80	( 95.7, 100.00 )	-

Table 5.5: Logistic model with two follow-ups, based upon model building steps with with intercept and random effects.

From Table 5.5, the covariate Mother asthma is only statistically significant. Also,  $\rho$  is not statistically significant indicating that the AR(1) term should be removed from the model. Based on these results we should fit a simpler model.

---

Now a simpler Hierarchical Bayes unit-level model is fitted (Rao & Molina, 2015, pg. 408). The 3-month follow-up for wheezing severity is used as the response along with covariates in the prenatal and 3-month time points. The model result is reported in Table 5.6.

Parameter	Estimate	95% credible interval(lower, upper)	Odds Ratio
Intercept	-2.74	(-3.25,-2.27)	-
Mother asthma (prenatal)	1.20	(0.63, 1.76 )	3.31
Mother smoked (prenatal)	1.05	(-0.02, 2.01 )	2.85
Farm (3 months)	-0.36	(-1.47, 0.60)	0.70
Furry pets (3 months)	-0.28	(-0.86, 0.28 )	0.76
$\sigma_v^2$	0.37	(0.04, 0.95)	-

Table 5.6: Unit-level logistic model with using only the 3 month follow-up of wheezing. Model has prenatal covariates with intercept and only area-level random effects.

The estimated parameters in Table 5.6 are now more sensible. Interestingly, in this particular model prenatal maternal asthma and smoking increase the odds of developing wheezing while living near a farm and having any furry pet decrease the odds of developing wheezing, noting that only the covariate Mother asthma has a statistically significant contribution in the model.

## 5.2 Zero-Inflated Poisson Model

Now to further understand the data we fit the ZIP model with the response being number of wheezing episodes. The model is similar to the previously mentioned ZIP model in equations (3.15-3.16). We consider the same covariates as discussed in

---

the logistic model (Section 5.1). Let

$$\begin{aligned}
\Theta_{ijk} = & \beta_0 + \beta_1 \text{MW}_{ik} + \beta_2 \text{MS}_{ik} + \beta_3 \text{FW}_{ik} \\
& + \beta_4 \text{FS}_{ik} + \beta_5 \text{AY}_{ik} + \beta_6 \text{AO}_{ik} \\
& + \beta_7 \text{MA}_{ik} + \beta_8 \text{DMA}_{ik} \\
& + \beta_9 \text{MTFAOW}_{ik} + \beta_{10} \text{FA}_{ik} \\
& + \beta_{11} \text{DFA}_{ik} + \beta_{12} \text{FTFAOW}_{ik} \\
& + \beta_{13} \text{FurryPets}_{ijk} + \beta_{14} \text{SES}_{ik} + f_{ij} + v_k
\end{aligned} \tag{5.2}$$

where

$$\begin{aligned}
v_k & \sim \mathcal{N}(0, \sigma_v^2) \\
f_{ij} & = \rho f_{ij-1} + \epsilon_{ij}, \epsilon_{ij} \sim \mathcal{N}(0, \sigma_f^2) \\
f_{ij} \mid f_{ij-1} & \sim \mathcal{N}(\rho f_{ij-1}, \sigma_f^2) \\
-1 < \rho & < 1.
\end{aligned}$$

Then the ZIP model becomes,

$$\text{Prob}(Y_{ijk} = m) = \begin{cases} \pi_{ijk} + (1 - \pi_{ijk}) \exp(-\lambda_{ijk}) & \text{if } m = 0 \\ (1 - \pi_{ijk}) \frac{\lambda_{ijk}^{y_{ijk}} \exp(-\lambda_{ijk})}{y_{ijk}!} & \text{if } m > 0 \end{cases} \tag{5.3}$$

such that  $\log(\lambda_{ijk}) = \Theta_{ijk}$ ,  $\text{logit}(\pi_{ijk}) = \text{logit}(\pi) = \gamma_0$ ,  $\gamma_0 \sim \mathcal{N}(0, 10000)$ . We run different univariate/multivariable analyses as part of model building process (see details in Appendix E). Based upon the results of the model building process, the final model is selected and the rate ratio (RR) is calculated for each covariate for interpretation by taking the exponent of the covariate parameter estimates (Table 5.7).

Parameter	Estimate	95% credible interval(lower, upper)	Rate Ratio
$\beta_0$	-1.33	(-1.83, -0.87 )	-
Mother asthma (prenatal)	0.00	(-0.42, 0.41 )	1.00
Mother wheezed (prenatal)	1.07	(0.70, 1.47 )	2.92
Mother smoked (prenatal)	-0.23	(-1.22, 0.75 )	0.79
Father smoked	-0.65	(-1.59, 0.28 )	0.52
Farm	0.61	(0.12, 1.11 )	1.85
Furry pets	0.07	(-0.24, 0.39 )	1.07
$\gamma_0$	1.43	(1.19, 1.65 )	-
$1 - \pi$	0.19	(0.16, 0.23 )	-
$\rho$	0.00	(0.00, 0.00 )	-
$\sigma_v^2$	0.34	(0.02, 0.63 )	-
$\sigma_f^2$	1.01	(0.80, 1.24 )	-
DIC	3066	-	-

Table 5.7: ZIP model with six covariates from the model building process.

The parameter estimates in Table 5.7 indicate that there are a high number of false zeros in the data since the estimate of  $1 - \pi$  indicates that only 19% of individuals have true zero counts. This is expected since many of the individuals did not wheeze thus their number of wheezing episodes is 0. The model building process gives in the indication to remove the auto correlated component of the ZIP model since  $\rho$  is estimated to be 0 in all models. Rate ratio estimates are also given for covaritates in the model. Rate ratios represent the multiplicative factor that number of wheezing episodes would increase/decrease for a one unit increase in the covariate of interest (for binary covariates presence vs no presence), given the other variables are held constant in the model. From Table 5.7 Mother asthma seems to have no effect on the number of wheezing episodes, while Mother wheezed, Farm and Furry pets seem to increase the number of wheezing episodes. Surprisingly parental smoking seems to expect a decrease in the number of wheezing episodes a child has, which is going against the literature. Referring to Appendix F, the models with lowest DIC values are preferred. Taking this into mind the most optimal model is fitted and then the DIC value is compared.

Parameter	Estimate	95% credible interval(lower, upper)	Rate Ratio
$\beta_0$	1.33	(-1.81, -0.90)	-
Mother asthma (prenatal)	0.05	(-0.37, 0.46)	1.05
Mother wheezed (prenatal)	1.04	(0.66, 1.44)	2.83
Mother smoked (prenatal)	-0.59	(-1.40, 0.20)	0.55
Farm	0.61	(0.13, 1.09 )	1.84
$\gamma_0$	1.41	(1.19, 1.64)	-
$1 - \pi$	0.20	(0.16, 0.23)	-
$\rho$	0.00	(0.00, 0.00 )	-
$\sigma_v^2$	0.31	(0.04, 0.60)	-
$\sigma_f^2$	1.03	(0.81, 1.27 )	-
DIC	3087	-	-

Table 5.8: ZIP model with four covariates. Covariates chosen based upon the models in Appendix F.

As the previous models suggested that there is no auto-correlation between follow-ups, thus  $\rho$  should be removed. For further model comparison a simple unit model with the first follow-up is considered. The model is described as

$$\Theta_{ik} = \beta_0 + \beta_1 MA_{ik} + \beta_2 MW_{ik} + \beta_3 MS_{ik} + \beta_4 Farm_{ik} + v_k \quad (5.4)$$

where

$$v_k \sim \mathcal{N}(0, \sigma_v^2).$$

Then, the ZIP model becomes

$$\text{Prob}(Y_{ik} = m) = \begin{cases} \pi_{ik} + (1 - \pi_{ik}) \exp(-\lambda_{ik}) & \text{if } m = 0 \\ (1 - \pi_{ik}) \frac{\lambda_{ik}^{y_{ik}} \exp(-\lambda_{ik})}{y_{ik}!} & \text{if } m > 0 \end{cases} \quad (5.5)$$

such that  $\log(\lambda_{ik}) = \Theta_{ik}$  and  $\text{logit}(\pi_{ik}) = \Theta_{ik}$ .

Parameter	Estimate	95% credible interval(lower, upper)	Rate Ratio
$\beta_0$	-1.22	(-2.17, -0.35 )	-
Mother asthma (prenatal)	0.84	(0.18, 1.52 )	2.32
Mother wheezed (prenatal)	0.02	(-0.63, 0.65 )	1.02
Mother smoked (prenatal)	1.01	(-0.38, 2.44)	2.75
Farm (3 months)	0.90	(0.10, 1.78 )	2.45
$\gamma_0$	1.45	(1.01, 1.88)	-
$1 - \pi$	0.19	(0.13, 0.27)	-
$\sigma_v^2$	1.37	(0.86, 2.07)	-
DIC	630.4	-	-

Table 5.9: Simple unit-level ZIP model as shown in equations (5.4-5.5). Using only the 3 month follow-up of number of wheezing episodes. Model has prenatal covariates with intercept and only area-level random effects.

From Table 5.9 the simple unit-level ZIP model seems the most reliable model. The model indicates that maternal wheezing does seem to affect the expected rate wheezing episodes in Manitoba. However, maternal smoking and asthma greatly increase the expected rate of wheezing episodes while only maternal asthma is statistically significant. Interestingly, living near a farm seems to also increase the rate of wheezing episodes among those who already have wheezing symptoms.

### 5.3 Area-Level Models

As mentioned before our parameters of interests are small-area proportions or means in the study depending on the type of response variable we have. We now need to use area-level covariates, this means that we need to convert the individual-level covariates into area-level. First we look at the spread of the dependent area-level variables and the number of individuals we have for each area, which will give us the sample size  $n_k$  for each area. For this study 60 areas are used in Manitoba including the Winnipeg region.

Table 5.10: Wheezing dependent variables at 3 months aggregated by each area. Recall Wheezing severity is binary while number of wheezing episodes is a count variable.

<b>Area</b>	<b>Wheezing severity at 3 months</b>	<b>Wheezing episodes at 3 months</b>	<b>Sample size</b>
1	1	10	19
2	3	6	36
3	1	5	9
4	1	1	20
5	2	2	40
6	0	0	8
7	2	4	6
8	2	1	24
9	0	0	1
10	0	0	6
11	4	24	18
12	1	0	23
13	0	0	5
14	4	16	43
15	2	2	19
16	0	0	13
17	2	2	19
18	0	0	2
19	5	12	17
20	5	6	30
21	0	0	13
22	0	0	18
23	3	4	29
24	1	1	26
25	5	9	30
26	0	0	1

Continued on next page

**Table 5.10 – continued from previous page**

<b>Area</b>	<b>Wheezing severity 3 months</b>	<b>Wheezing episodes 3 months</b>	<b>sample size</b>
27	0	0	3
28	0	0	1
29	0	0	1
30	0	0	1
31	0	0	6
32	1	1	1
33	0	0	1
34	0	0	16
35	0	0	4
36	0	0	1
37	1	2	1
38	1	1	1
39	0	0	1
40	1	1	8
41	0	0	7
42	1	1	4
43	0	0	1
44	0	0	3
45	0	0	15
46	0	0	9
47	0	0	12
48	0	0	1
49	0	0	1
50	1	2	5
51	3	3	17
52	0	0	6
53	2	2	30
54	1	1	28

Continued on next page

---

**Table 5.10 – continued from previous page**

<b>Area</b>	<b>Wheezing severity 3 months</b>	<b>Wheezing episodes 3 months</b>	<b>sample size</b>
55	3	9	8
56	0	0	7
57	0	0	3
58	1	1	10
59	0	0	1
60	1	3	1

As shown from Table 5.10, the sample of 690 individuals are spread across 60 RHADs in Manitoba. Some of the RHADs contain only one individual meaning the sample size is one. However, proportions can still be set-up for each area using the sample size as the denominator (eg. in area 1, 1 out of 19 children wheezed at 3 months).

The covariates which indicate environment factors near home are also converted to area-level. These environmental covariates indicate if a participant's home was within 100 metres from any major highway/artery, body of water, factory, farm, and prolonged construction. Then, we replace the individual-level covariates with these proportion covariates to establish area-level covariates.

Prenatal baseline covariates such as parental smoking and history of smoking are also included in the model given their statistical significance and also following the literature. In particular, prenatal Mother asthma, Mother smoked, Father smoked and Father asthma are converted to area-level covariates. Also, history of parental wheezing covariates Mother wheezed and Father wheezed are converted to area-level. Since the aforementioned covariates are binary the approach for this is to create proportions to represent the proportion of for example smokers in a given small-area.

---

Another covariate to be considered is the SES index. Since the SES is a continuous variable, a reasonable approach is to calculate average SES per area using the individual-level SES index derived previously. Another approach to calculating area-level SES is to average the individual covariates that are used to make the SES index (Father/Mother Education and household income) taken by area and then area-level SES indexes are calculated. This alternative technique is referred to as SES (average before).

### 5.3.1 Proportion of Wheezing Severity

We now begin univariate analysis using the logistic area-level model as shown in Section 3.5.3. The results of the univariate analysis are shown in Appendix G. From the results of the model building process in Appendix H, we obtain the final model. In the final model we keep covariates which have contributions in the literature even though they are not statistically significant. When looking at the modeling building steps in Appendix H, it is observed that the DIC values increase with addition of covariates (which are not statistically significant but clinically important) at the expense of slightly higher DIC.

Parameter	Estimate	95% credible interval(lower, upper)	Odds Ratio
Intercept	-3.27	(-4.09, -2.52 )	-
Mother wheezed (prenatal)	0.01	(-0.01, 0.04)	1.01
Mother asthma (prenatal)	0.02	(0.01, 0.05)	1.02
Mother smoked (prenatal)	0.00	(-0.03, 0.03)	1.00
Farm	-0.01	(-0.03, 0.00)	0.99
$\sigma_v^2$	0.25	(0.03, 0.84)	-
DIC	139.7	-	-

Table 5.11: Final Area-level logistic model with four covariates. Odds ratios are calculated by taking the exponent of the covariate parameter estimates.

The final area-level logistic model reported in Table 5.11 which has a statistically significant covariate Mother asthma. In particular, the increase in proportion

---

of prenatal maternal asthma and wheezing in an area slightly increases the odds of developing wheezing while the percentage living near a farm slightly decreases the odds of developing wheezing. Maternal smoking interestingly has no effect on the odds of developing wheezing. It is also worth mentioning that aggregating the sample data loses some information hence why the covariates have only a marginal contribution in the odds of wheezing. However, the main purpose of the area-level model is the prediction of wheezing at the area-level. Using the results of Table 5.11, the proportion of predicted wheezing in each area  $k$  ( $= 1, \dots, K$ ) is obtained by  $\hat{p}_k^{HB} = h(\mathbf{x}_k^T \hat{\boldsymbol{\beta}} + \hat{v}_k)$ . These predicted proportions are then mapped to their corresponding RHADs in Manitoba (Figure 5.1). Note that in our CHILD study, we had information for 60 (out of 96) RHADs thus the rest will appear as missing in the maps. For comparison purposes, the area-level wheezing proportions obtained from the data by dividing wheezing severity at 3 months by the sample size in Table 5.10 are also mapped (Figure 5.2). The proportions obtained from the raw data indicate some areas with almost no proportion of wheezing while the predicted proportions obtained from the model slightly increase these proportions. The area-level binary logistic model in Section 3.5.3 is predicting the proportions of wheezing severity for each small-area while taking into account the auxiliary information available for each area. This results in smoothed predictions towards the mean (Figure 5.1).

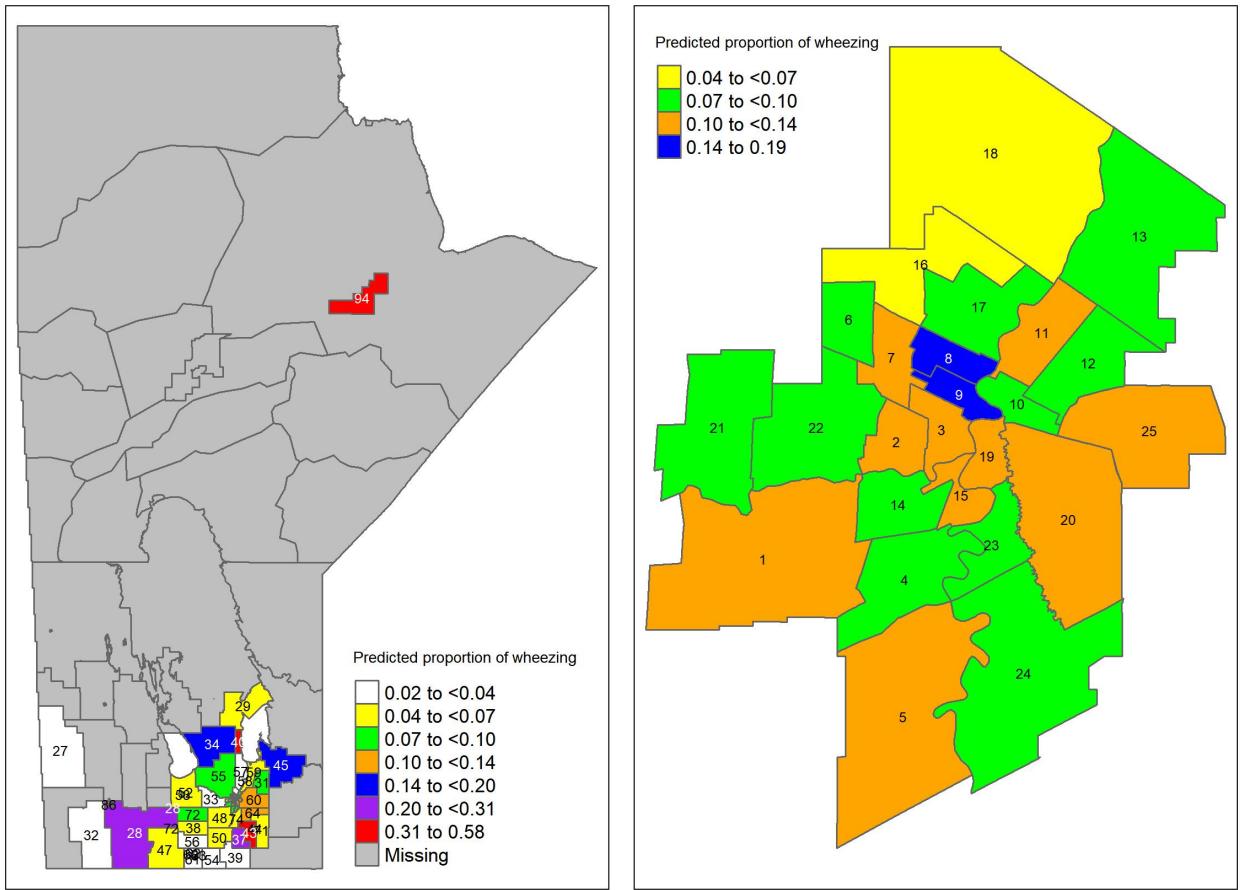


Figure 5.1: Predicted proportion of wheezing severity using the final area-level logistic model in Manitoba. Winnipeg specific are proportions shown on the right panel (Region codes are given in Appendix A).

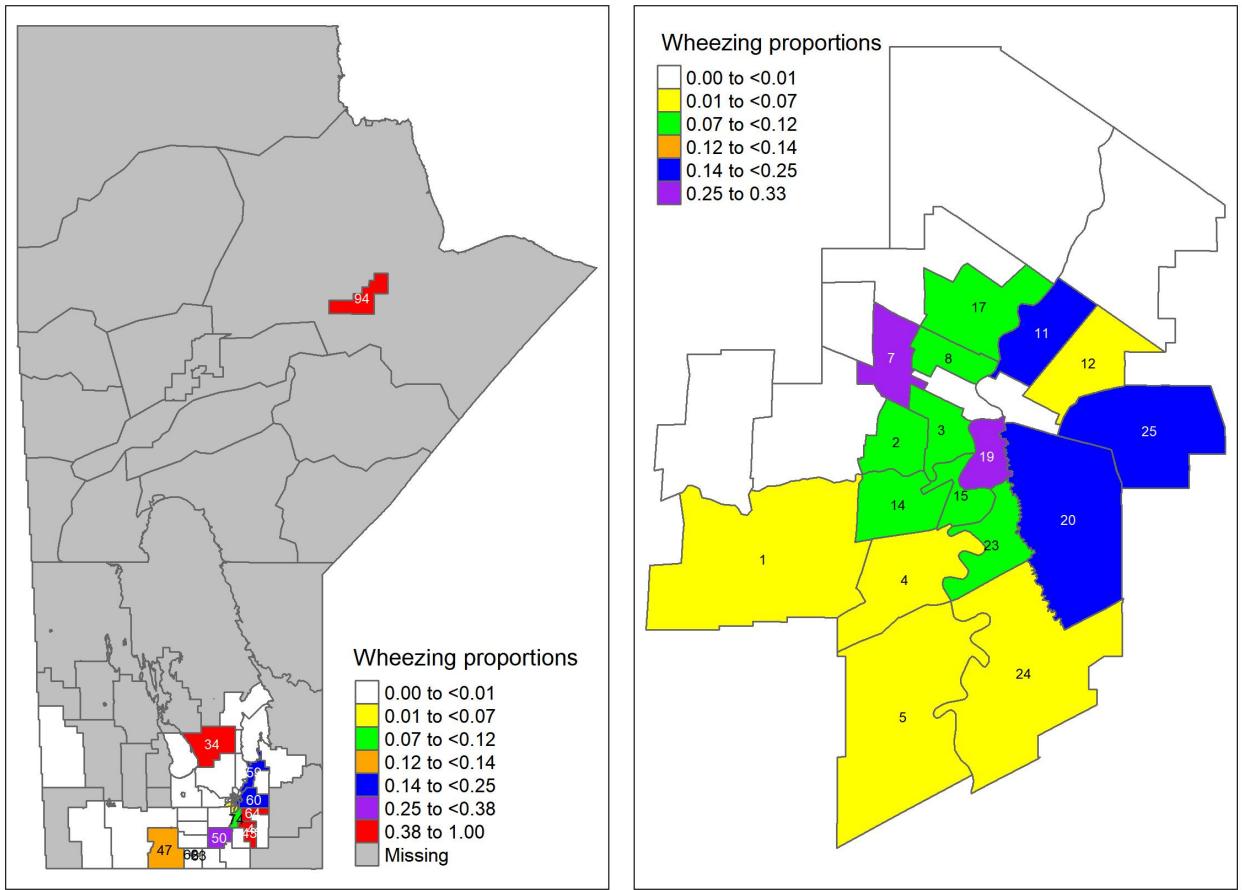


Figure 5.2: Proportion of wheezing severity obtained from the data in Manitoba (aggregated wheezing severity at 3 months divided by sample size in each area). Winnipeg specific proportions are shown on the right panel (Region codes are given in Appendix A).

### 5.3.2 Average Rate of Wheezing Episodes

We now proceed to perform the same steps for model building as we did for the area-level logistic model but now the response is a count variable, number of wheezing episodes as shown in Table 5.10. We consider the ZIP model studied in Section 3.5.3. The results of the univariate analysis are shown in Appendix I. From the univariate analysis we can see that none of the area-level covariates are statistically significant. In addition, the results suggest that the proportion of true zeros is almost 100% which is highly unusual. Thus we need to investigate by running a variation of the ZIP model with covariates for the logistic component of the ZIP model. This model is shown in equations (3.20-3.21), now we will not be using a

---

prior for the logistic component but instead the same covariates as used in the Poisson component. In particular, we assume  $\log(\lambda_k) = \Theta_k$ ,  $\text{logit}(\pi_k) = \Theta_k$ , where  $\Theta_k = \mathbf{x}_k^T \boldsymbol{\beta} + v_k$  as shown in equation (3.20). Univariate analyses are run for the old ZIP model (ZIP model with a prior for the logistic component) (Table 5.12) and the new ZIP model (Table 5.13).

Parameter	Estimate	95% credible interval (lower, upper)
$\beta_0$	-1.22	(-2.38, -0.31 )
Mother asthma (prenatal)	0.02	(0.00, 0.05 )
$\gamma_0$	-796.00	(-2250, -29.90 )
$1 - \pi$	1.00	(1.00, 1.00 )
$\sigma_v^2$	1.91	(1.34, 2.72 )
DIC	158.4	-

Table 5.12: Univariate area-level ZIP model with random effects and prenatal Mother asthma as a covariate. Unusually large estimates of  $\gamma_0$  and  $1 - \pi$  is 1 indicate this model is not optimal.

Parameter	Estimate	95% credible interval (lower, upper)
$\beta_0$	1.02	(0.24, 1.79 )
Mother asthma (prenatal)	-0.01	(-0.04, 0.01 )
$\sigma_v^2$	1.32	(0.82, 2.07 )
DIC	176.0	-

Table 5.13: Univariate area-level ZIP model with Mother asthma as a covariate for both the logistic and Poisson components of the model.

Furthermore, a conventional Poisson model with area-specific random effects is run with the same covariate for comparison purposes (Table 5.14). In particular, the Poisson model is described as

- (i)  $y_k | \lambda_k \sim \text{Poisson}(\lambda_k)$ ,
- (ii)  $\log(\lambda_k) = \mathbf{x}_k^T \boldsymbol{\beta} + v_k$ , with  $v_k \stackrel{\text{iid}}{\sim} N(0, \sigma_v^2)$  and priors,
- (iii)  $f(\boldsymbol{\beta}) \propto 1$  and  $\sigma_v^{-2} \sim G(0.001, 0.001)$ .

Parameter	Estimate	95% credible interval (lower, upper)
$\beta_0$	-1.25	(-2.35, -0.30 )
Mother asthma (prenatal)	0.02	(0.00, 0.05 )
$\sigma_v^2$	1.91	(1.35, 2.68 )
DIC	188.7	-

Table 5.14: Univariate Poisson area-specific random effects model with Mother asthma as a covariate.

Comparing the three models reported in Tables (5.12, 5.13, 5.14), we believe the second model (new ZIP model) has a merit to proceed even though it has the second largest DIC value. Thus, we proceed with full univariate analysis for this model (Appendix J). All the univariate models claim a decrease in the mean number of wheezing episodes in each area. When looking at the multivariable models in Appendix K it seems that the best model is to use Farm and prenatal Mother asthma as covariates (Table 5.15 ) since this results in relatively low DIC when compared to the other models and the interpretation of the rate ratio makes clinical sense. Table 5.15 shows that living within a 100 metres from a farm decreases the expected rate of wheezing episodes by 3%, while prenatal Mother asthma increases the expected rate of wheezing episodes by 1%.

Parameter	Estimate	95% credible interval(lower, upper)	Rate Ratio
$\beta_0$	0.95	(-0.02, 1.90 )	-
Farm	-0.03	(-0.6, 0.01)	0.97
Mother asthma (prenatal)	0.01	(-0.02, 0.03 )	1.01
$\sigma_v^2$	1.17	(0.76, 1.75 )	-
DIC	164.7	-	-

Table 5.15: Final area-level ZIP model with two covariates Farm and Mother Asthma for both the logistic and Poisson components of the model.

Using the results of Table 5.15, the predicted average number of wheezing episodes in each area  $k$  ( $= 1, \dots, K$ ) is calculated using  $\hat{\lambda}_k^{HB} = \exp(\mathbf{x}_k^T \hat{\boldsymbol{\beta}} + \hat{v}_k)$ .

These predicted average wheezing episodes are then mapped to their corresponding RHADs in Manitoba (Figure 5.3). For comparison purposes the area-level wheezing episodes at 3 months obtained from the Table 5.10 are also mapped (Figure 5.4). The average number of wheezing episodes obtained directly from the data indicate that many areas have none or small number of wheezing episodes. In particular, regions 27, 28 and 45 (Appendix A) have no wheezing episodes while the smoothed predictions using the proposed new ZIP model (Figure 5.3) have at least 4 episodes of wheezing at those areas.

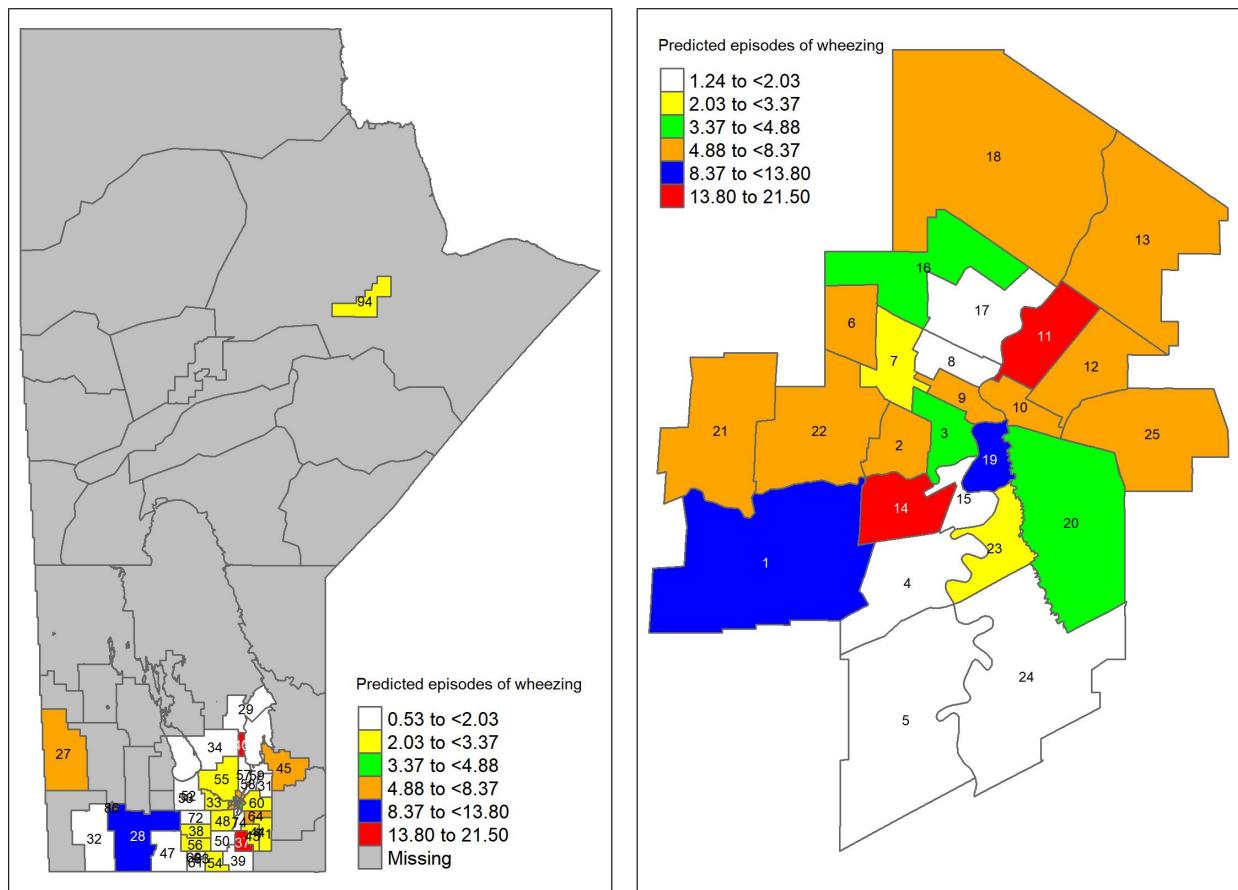


Figure 5.3: Predicted average number of wheezing episodes in Manitoba using the final area-level (new) ZIP model. Winnipeg specific proportions are shown on the right panel (Region codes are given in Appendix A).

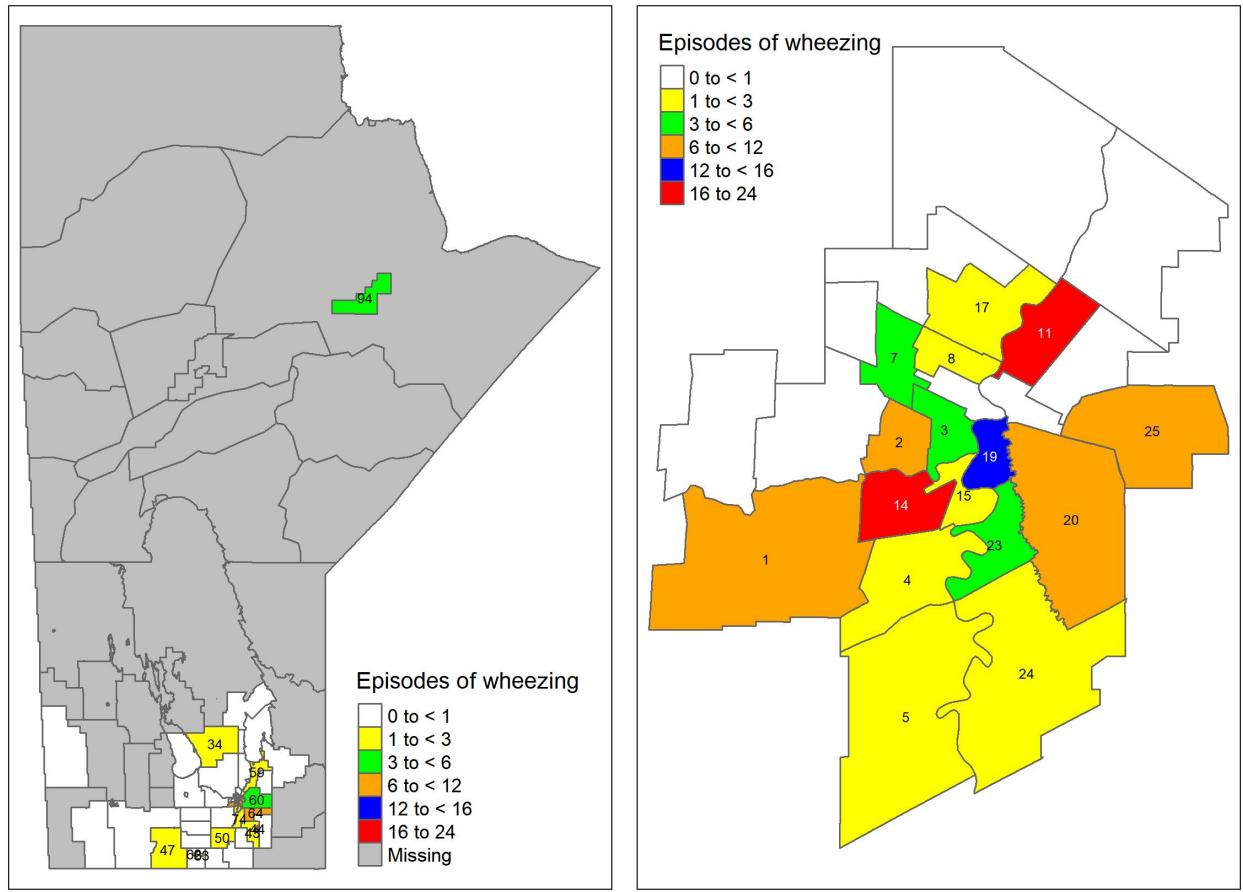


Figure 5.4: Number of wheezing episodes obtained from the data in Manitoba (aggregated wheezing episodes at 3 months). Winnipeg specific proportions are shown on the right panel (Region codes are given in Appendix A).

# **Chapter 6: Discussion & Conclusions**

This research project has been focused on wheezing development for children under five years. Since asthma is difficult to diagnose, wheezing was used as a outcome variable for all the statistical models constructed for this study. The majority of this project was to analyze the role geographic location plays on wheezing in children within Manitoba. To answer this question, logistic regression and ZIP models were used at both the individual and area level. The aim here was to determine which individual-level covariates affect wheezing of children. Estimates of the wheezing proportions and counts using the area-level models provided some insight into answering this objective. The second objective of this study was to determine how wheezing rates change throughout childhood and how early future chronic wheezing conditions will be predicted for children under five.

## **6.1 Wheezing in Small Areas in Manitoba**

Our objective was to investigate to what extent location affects the severity of wheezing and if there are underlying covariates acting as proxy variables of location. To begin to answer this question, we used a Bayesian hierarchical model to fit a binary logistic model for wheezing severity of our CHILD study. The model was intended to use 96 RHADs as small areas but due to how the participants were recruited we were only able to use 60 RHADs. From the model building process, the model shown in Table 5.6 which used a combination of prenatal and 3-month time points seemed the most optimal model since it contains interpretations that make the most clinical sense according to the literature. The model chosen was a unit-level model which only uses one follow-up at three months and prenatal maternal covariates. Furthermore, the random effect representing small areas was also statistically significant indicating there is a difference in wheezing severity between

---

areas. The interpretation of the model in Table 5.6 is that if the child's mother had asthma and smoked then this results in increased odds of the child in the cohort developing wheezing. The absence of association with lower proportion of wheezing in some of the interim logistic models may reflect the relatively low rates of smoking in this cohort. It is known from the literature, prenatal family history of atopic diseases is associated with similar developments in children, therefore it indicates that the model and the data within CHILD cohort support these findings. The model also suggested that living near a farm and having furry pets decrease the odds of developing wheezing. There is evidence from the literature to support that living near farm animals decreases the likelihood of developing wheezing and asthma but can have the opposite effect when exposed to furry pets such as cats and dogs. It is important to note however that the logistic model says living near a farm affects the odds of developing wheezing severity while the ZIP model says the expected change in the rate of wheezing episodes if living near farm among those children who already have wheezing cases. Hence, the ZIP model suggests a protective relationship between farm and rate of wheezing. There also may be some possible overlap between the covariates and since children living near farms were more prone to exposure of furry animals, however, in this case both covariates had the same effect. Another possible explanation is that children exposed to animals with a history of wheezing include exposure to furred animal allergen to which the child is allergic may aggravate wheezing in children with asthma. We are also interested in predicting the small-area proportions and average counts of wheezing. First the data was aggregated to the area-level which essentially lost some information regarding aggregated covariates but the goal was to predict area-level of wheezing. The final area-level model (Table 5.11) was used to map the predicted proportion of wheezing within Manitoba. The highest proportions of wheezing based on the predictions provided were Gimli, Hanover, Spruce Woods and St Pierre, which are interesting since in the raw data the proportion of wheezing in Gimli, Spruce Woods and St Pierre is zero. Hanover does have a high proportion of wheezing when looking at the raw

---

data. Fox Lake Cree Nation also has one of the highest proportions of wheezing in both the predictions and the raw data. It should be noted that the sample sizes for these areas were only one and thus the smoothed model also takes into account the covariates which are relevant to the outcome. Thus, the predicted proportions were heavily affected by maternal health history and if the child lived near a farm (which is effectively an indication of living near greenspace). The difficulty is that since participants in Manitoba were recruited within a radius of Winnipeg and Morden-Winkler, it is hard to draw conclusions regarding if the current environment caused wheezing or prenatal environment particularly with a sample size of one. However, the small area-level models technique accounted for the number of participants in each area and even areas with only one participant were not removed since auxiliary information related to the outcome of wheezing was still available to us. The proportion of wheezing in Winnipeg is fairly consistent with predictions ranging from mostly 0.10-0.19 with the highest wheezing proportions concentrated in the center of Winnipeg, including downtown and River East areas.

An area-level ZIP model was introduced for wheezing counts in Manitoba. The ZIP model also allowed to incorporate the excess zeros present in the wheezing outcome of the study. A final area-level model was chosen with the same covariates in both the logistic false zero component and the count component. This model was found easier to interpret compared to letting the false zero component be a diffuse prior. This final model agreed with the aforementioned logistic model and was used to map average wheezing counts per area in Manitoba. The final area-level ZIP model in Table 5.15 provided marginal contribution of covariates (Mother asthma and Farm) but had a significant area-level random effect. The final area-level ZIP model was used to map the average number of wheezing episodes within Manitoba. From the map (Figure 5.3), the highest predicted average number of wheezing episodes were Gimli, Spruce Woods and St Pierre. When compared to raw data this is the opposite since these areas have low rates of wheezing. However as previously discussed when comparing the predicted proportions these areas have

---

very low sample sizes and thus very low counts of wheezing episodes. Thus the proposed (smoothed) model was again used to account for the covariates and also for prediction of wheezing episodes for these small areas.

## 6.2 Wheezing Throughout Childhood

Models were used to address the second goal of this study which is to find how wheezing changes throughout childhood and determine if there are any patterns in wheezing at certain age groups. Wheezing phenotypes are crucial to understanding asthma development within children, wheezing phenotypes are defined by time of onset and persistence, thus it is important to note that the variation among follow-ups was significant and showed great variation between follow-ups. However, the high odds ratios of the longitudinal models make the models not very useful to make interpretations. Restricting follow-ups did not improve the interpretation of models, this could still be the result of trying to fit a complex model with a low number of wheezing cases as the outcome causing the odds ratio to be overinflated. The children who initially wheezed at the first follow-up of 3 months (17%) had consistent wheezing throughout the study while the rest had no cases or few cases. There were however some missing data within this subset of 63 individuals who initially wheezed with the later follow-ups containing the most missing data. There is an evidence to suggest that children started wheezing at different time points through the study. When looking at the other follow-ups, those children began wheezing at different follow-up points mostly all stop wheezing by the end of the study. In fact only one child wheezed consistently at every follow-up in the entire study.

Wheezed since last follow-up		3 Months	6 months	1 year	18 months	2 years	2.5 years	3 years	4 years	5 years
No	0	41	44	36	41	45	42	34	42	
Yes	63	17	17	17	19	14	12	11	11	
Missing (all codes)	0	5	2	10	3	4	9	18	10	

Table 6.1: Child Health: Frequency table displaying distribution of wheezing for all 63 individuals who wheezed in the first follow-up (3 months).

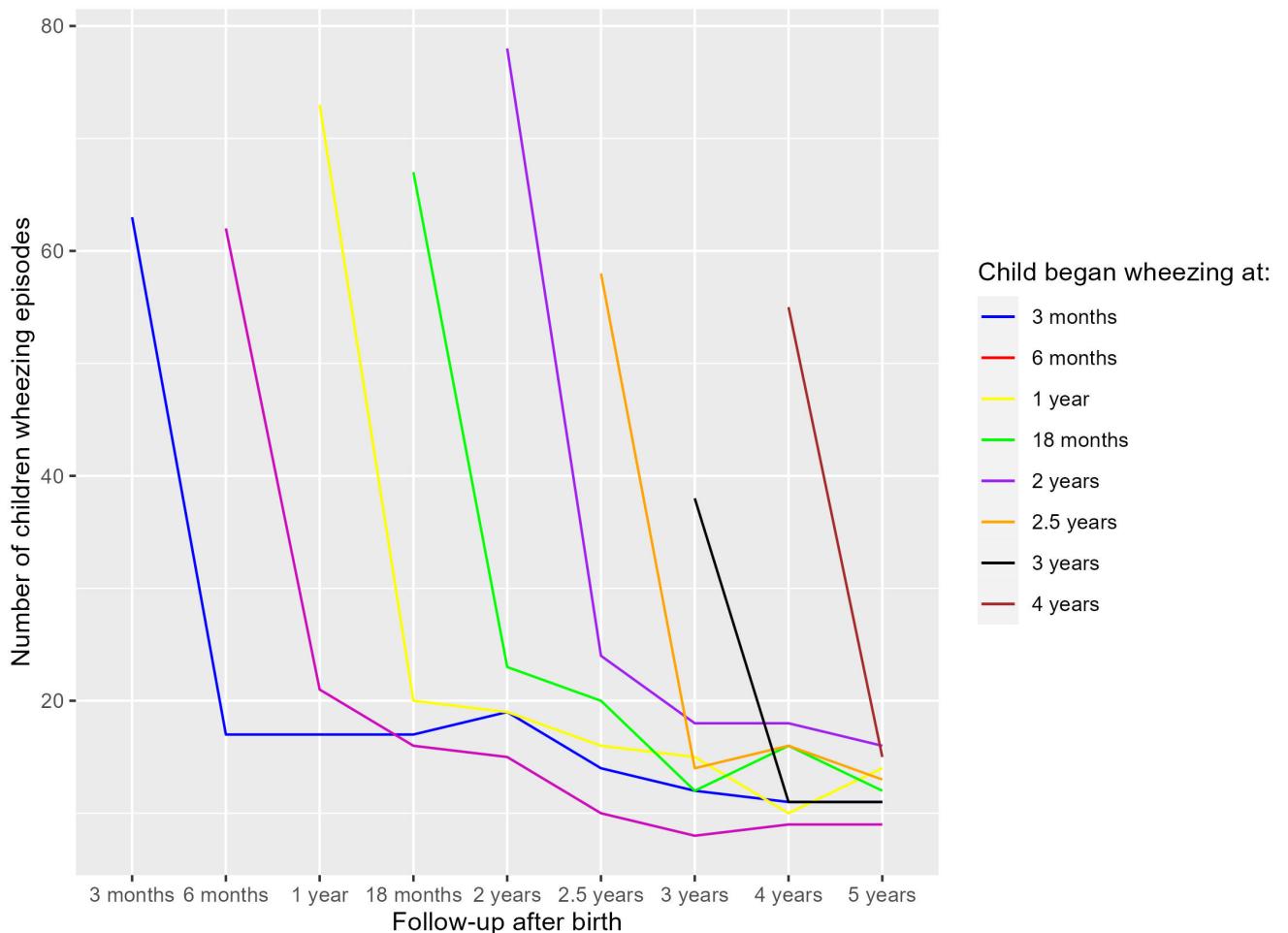


Figure 6.1: Number of wheezing episodes for children who started wheezing in every follow-up and if they continued wheezing until the last follow-up of 5 years after birth.

From Figure 6.1 the data suggests that most children in the cohort that started wheezing in the study stop wheezing within the next follow-up. In total more than two thirds of children who wheezed during the study stopped wheezing by the end of the study. An explanation could be the effect of the area since as discussed previously there was movement between areas between the follow-ups. When looking at the number of children that wheezed at two years by area in Figure 3.7 (pg.26), it suggests that this follow-up had the most spread of children wheezing throughout Manitoba but the other follow-ups had higher concentrations of children wheezing within specific areas. Another aspect that cannot be ignored is the timing of the follow-ups since the decrease in wheezing could be due to seasonality. However,

---

further time points would be required to test this assumption. It is clear however that there was a definite spike in cases of wheezing when the children reached 2 years of age.

When trying to determine if future chronic wheezing conditions be predicted before the age of two it should be kept in mind which area child currently resides and previously resided in. Additionally with further knowledge of parental health history, one could make an accurate prediction if the child will develop persistent wheezing which could ultimately result in asthma.

### **6.3 Strengths and Limitations**

One of the major strengths of this study is that data collection was from a prospective cohort study. Prospect studies ensure accuracy of data collection with regard to exposures and confounders. Since it is a prospect cohort study there is minimal selection bias. This study also demonstrated that reliable area-level wheezing rate predictions can be produced which can be useful in predicting wheezing given areas with small sample sizes. Additionally the study made use of the numerous variables within the CHILD study to further explain the causes of wheezing within the cohort. It was while using the variables as covariates it became clear that there are underlying proxy variables which are related to area like living near a farm which can decrease the chances of a child developing persistent wheezing. Ultimately this study showed that wheezing could be used as an indicator for asthma since asthma when used as a covariate in the logistic models was highly significant.

However, this study is limited due to the fact that recruitment for CHILD cohort was mainly from Winnipeg and a small rural population outside of Winnipeg. This limits the study to a small selection of areas within Manitoba. Although Winnipeg has the largest population density it would be beneficial to explore other areas with different environmental exposures. The other considerable limitation of this study was the inability to account for movement between small areas since this

---

would require a more complex model which would ultimately require more data. Although there is information in the movement of children from one area to another area, the missing data would also not allow for such model. Similarly, the data currently contains a substantial amount of missing data with some covariates having very few follow-ups, that led to longitudinal models with over-inflated odds ratios. In the future when there is more data available, longitudinal models with more complexities could build upon on the foundations of this study.

## 6.4 Conclusions

In conclusion, our study has demonstrated that wheezing is heavily affected by maternal health history, particularly maternal history of asthma and smoking. Living near a farm seems to decrease the odds of wheezing but tends to increase exacerbations of wheezing among those children already wheezing. It was shown that children who lived with furry pets had decreased risks of developing wheezing but due to the potential overlap between children who live near farms and own furry pets, this result is questionable. When looking at the small-areas outside Manitoba, Gimli, Spruce Woods and St Pierre were found to have the highest proportions of wheezing and persistent wheezing episodes. However, the models made these predictions were based on the auxiliary covariates provided and further recruitment would be needed to confirm these results. When looking at small areas within Winnipeg, it is clear that Winnipeg had the largest concentration of wheezing cases. In particular Inkster, River East, St.Boniface and River Heights had the largest number of wheezing cases within Winnipeg.

## References

- Alkan, N. (2017). Assessing convergence diagnostic tests for bayesian cox regression. *Communications in statistics. Simulation and computation*, 46(4), 3201–3212.
- Azad, M. B., Vehling, L., Lu, Z., Dai, D., Subbarao, P., Becker, A. B., Mandhane, P. J., Turvey, S. E., Lefebvre, D. L., Sears, M. R., Anand, S., Befus, A., Brauer, M., Brook, J., Chen, E., Cyr, M., Daley, D., Dell, S., Denburg, J., ... To, T. (2017). Breastfeeding, maternal asthma and wheezing in the first year of life: A longitudinal birth cohort study. *The European respiratory journal*, 49(5), 1602019.
- Brick, T., Hose, A., Wawretzka, K., Mutius, E., Roduit, C., Lauener, R., Riedler, J., Karvonen, A. M., Pekkanen, J., Divaret-Chauveau, A., Dolphin, J.-C., Ege, M. J., Kalayci, Ö., Roponen, M., Hirvonen, M. R., Hyvärinen, A., Kirjavainen, P. V., Tittanen, P., Remes, S., ... Schmaußer-Hechfellner, E. (2019). Parents know it best: Prediction of asthma and lung function by parental perception of early wheezing episodes. *Pediatric allergy and immunology*, 30(8), 795–802.
- Canadian Institute for Health Information. (2018). *Emergency department wait times in canada continuing to rise* (Report). <https://www.cihi.ca/en/emergency-department-wait-times-in-canada-continuing-to-rise>
- Castro-Rodriguez, J. A., Cifuentes, L., & Martinez, F. D. (2019). Predicting asthma using clinical indexes. *Frontiers in pediatrics*, 7, 320–320.
- Caui, D., Savenije, O. E. M., Smit, H. A., Postma, D. S., Koppelman, G. H., Wijga, A. H., Kerkhof, M., Gehring, U., Hoekstra, M. O., Brunekreef, B., & de Jongste, J. C. (2013). Perinatal risk factors for wheezing phenotypes in the first 8 years of life. *Clinical and experimental allergy*, 43(12), 1395–1405.

- 
- Chan, J. C. C., & Grant, A. L. (2016). On the Observed-Data Deviance Information Criterion for Volatility Modeling. *Journal of Financial Econometrics*, 14(4), 772–802. <https://doi.org/10.1093/jjfinec/nbw002>
- Duijts, L., Granell, R., Sterne, J. A., & Henderson, A. J. (2016). Childhood wheezing phenotypes influence asthma, lung function and exhaled nitric oxide fraction in adolescence. *The European respiratory journal*, 47(2), 510–519.
- Fay, R. E., & Herriot, R. A. (1979). Estimates of income for small places: An application of james-stein procedures to census data. *Journal of the American Statistical Association*, 74(366), 269–277.
- Gamerman, D. (2006). *Markov chain monte carlo : Stochastic simulation for bayesian inference*. (2nd ed. / Dani Gamerman, Hedibert F. Lopes.). Chapman & Hall/CRC.
- Global Initiative for Asthma. (2019). *Global strategy for asthma management and prevention* (Report). <https://ginasthma.org/wp-content/uploads/2019/06/GINA-2019-main-report-June-2019-wms.pdf>
- Hartley, K., Ryan, P., Brokamp, C., & Gillespie, G. L. (2020). Effect of greenness on asthma in children: A systematic review. *Public health nursing (Boston, Mass.)*, 37(3), 453–460.
- Kruschke, J. K. (2015). *Doing bayesian data analysis : A tutorial with r, jags, and stan* (Edition 2.). Academic Press.
- Luo, G., Nkoy, F. L., Stone, B. L., Schmick, D., & Johnson, M. D. (2015). A systematic review of predictive models for asthma development in children. *BMC medical informatics and decision making*, 15(1), 99–99.
- Morgan, W. J., & Martinez, F. D. (1992). Risk factors for developing wheezing and asthma in childhood [Asthma]. *Pediatric Clinics of North America*, 39(6), 1185–1203. [https://doi.org/https://doi.org/10.1016/S0031-3955\(16\)38440-1](https://doi.org/https://doi.org/10.1016/S0031-3955(16)38440-1)
- Public Health Agency of Canada. (2018). *Report from the canadian chronic disease surveillance system : Asthma and chronic obstructive pulmonary disease*

- 
- (*copd*) in canada, 2018. Public Health Agency of Canada = Agence de la santé publique du Canada.
- Quirt, J., Hildebrand, K. J., Mazza, J., Noya, F., & Kim, H. (2018). Asthma. *Allergy, asthma, and clinical immunology : official journal of the Canadian Society of Allergy and Clinical Immunology*, 14(Suppl 2), 50–50. <https://doi.org/10.1186/s13223-018-0279-0>
- Rao, J. (1994). Estimating totals and distribution functions using auxiliary information at the estimation stage. *Journal of official statistics*, 10(2), 153.
- Rao, J. (1999). Some recent advances in model-based small area estimation. *Survey Methodology*, 25, 175–186.
- Rao, J., & Molina, I. (2015). *Small area estimation*. John Wiley & Sons, Ltd. <https://doi.org/https://doi.org/10.1002/9781118735855>
- Richelle, J. (2014). *Wheezing in early life and validating the asthma predictive index* (Thesis). <http://hdl.handle.net/1993/30555>
- Savenije, O. E., Kerkhof, M., Koppelman, G. H., & Postma, D. S. (2012). Predicting who will have asthma at school age among preschool children. *Journal of allergy and clinical immunology*, 130(2), 325–331.
- Shokoohi, F., & Torabi, M. (2018). Semi-parametric small-area estimation by combining time-series and cross-sectional data methods. *Australian & New Zealand journal of statistics*, 60(3), 323–342.
- Statistics Canada. (2019). Asthma, by age group. <https://www150.statcan.gc.ca/t1/tbl1/en/tv.action?pid=1310009608>
- Subbarao, P., Anand, S. S., Becker, A. B., Befus, A. D., Brauer, M., Brook, J. R., Denburg, J. A., HayGlass, K. T., Kobor, M. S., Kollmann, T. R., Kozyrskyj, A. L., Lou, W. Y. W., Mandhane, P. J., Miller, G. E., Moraes, T. J., Pare, P. D., Scott, J. A., Takaro, T. K., Turvey, S. E., . . . Sears, M. R. (2015). The canadian healthy infant longitudinal development (child) study: Examining developmental origins of allergy and asthma. *Thorax*, 70(10), 998–1000.

- 
- Subbarao, P., Mandhane, P. J., & Sears, M. R. (2009). Asthma: Epidemiology, etiology and risk factors. *Canadian Medical Association journal (CMAJ)*, 181(9), E181–E190.
- Takaro, T. K., Scott, J. A., Allen, R. W., Anand, S. S., Becker, A. B., Befus, A. D., Brauer, M., Duncan, J., Lefebvre, D. L., Lou, W., Mandhane, P. J., McLean, K. E., Miller, G., Sbihi, H., Shu, H., Subbarao, P., Turvey, S. E., Wheeler, A. J., Zeng, L., . . . Brook, J. R. (2015). The canadian healthy infant longitudinal development (child) birth cohort study: Assessment of environmental exposures. *Journal of exposure science & environmental epidemiology*, 25(6), 580–592.
- The Canadian Lung Association. (2011). *Pre-budget consultations house of commons standing committee on finance* (Report). [https://www.ourcommons.ca/Content/Committee/411/FINA/WebDoc/WD5138047/411\\_FINA\\_PBC2011-Briefs/Canadian%20Lung%20Association%20E.pdf](https://www.ourcommons.ca/Content/Committee/411/FINA/WebDoc/WD5138047/411_FINA_PBC2011-Briefs/Canadian%20Lung%20Association%20E.pdf)
- Venter, C., Palumbo, M. P., Sauder, K. A., Glueck, D. H., Liu, A. H., Yang, I. V., Ben-Abdallah, M., Fleischer, D. M., & Dabelea, D. (2021). Incidence and timing of offspring asthma, wheeze, allergic rhinitis, atopic dermatitis, and food allergy and association with maternal history of asthma and allergic rhinitis. *The World Allergy Organization journal*, 14(3), 100526–100526.
- Wegienka, G., Havstad, S., Zoratti, E. M., Ownby, D. R., & Johnson, C. C. (2009). Association of early life wheeze and lung function. *Annals of allergy, asthma, & immunology*, 102(1), 29–34.

# Appendices

## Appendix A: Region Codes

Table A.1: Region code with associated location and small area number.

Code	Region	Small area
W002	Assiniboine South	1
W11A	Downtown	2
W11B	Downtown	3
W03A	Ft. Garry	4
W03B	Ft. Garry	5
W09A	Inkster	6
W09B	Inkster	7
W10A	Point Douglas	8
W10B	Point Douglas	9
W07A	River East	10
W07B	River East	11
W07C	River East	12
W07D	River East	13
W12A	River Heights	14
W12B	River Heights	15
W08A	Seven Oaks	16
W08B	Seven Oaks	17
W08C	Seven Oaks	18
W05A	St. Boniface	19
W05B	St. Boniface	20
W01A	St. James -Assiniboia	21
Continued on next page		

---

**Table A.1 – continued from previous page**

Code	Region	Small area
W01B	St. James - Assiniboia	22
W04A	St. Vital	23
W04B	St. Vital	24
W006	Transcona	25
WE34	Souris River	26
WE31	Asessippi	27
WE36	Spruce Woods	28
IE42	Arborg/Riverton	29
WE32	Little Saskatchewan	30
IE31	Beausejour	31
WE33	Turtle Mountain	32
SO14	Cartier/SFX	33
IE43	St. Laurent	34
WE13	Riding Mountain	35
WE15	Dauphin	36
SO25	St Pierre/DeSalaberry	37
SO22	Carman	38
SO26	Red River South	39
IE41	Gimli	40
WE35	Whitemud	41
WE11	Duck Mountain	42
SO45	Hanover	43
SO44	Steinbach	44
IE32	Pinawa/LDB	45
SO11	Seven Regions	46
SO31	Lorne/Louise/Pembina	47

Continued on next page

---

**Table A.1 – continued from previous page**

Code	Region	Small area
SO23	MacDonald	48
WE12	Porcupine Mountain	49
SO24	Morris	50
SO12	MacGregor	51
SO13	Rural Portage	52
SO15	City of Portage la Prair	53
SO33	Altona	54
IE21	Stonewall/Teulon	55
SO36	Roland/Thompson	56
IE22	Wpg Beach/St. Andrews	57
IE11	Selkirk	58
IE23	St. Clements	59
IE24	Springfield	60
SO32	Stanley	61
SO34	Morden	62
SO35	Winkler	63
SO42	Tache	64
WP21	Winnipeg Churchill	65
IE52	Fisher/Peguis	66
NO11	Flin,Snow,Cran,Sher	67
NO13	LL/MC,LR,O-P(SIL),PN(GVL)	68
IE51	Powerview/PF	69
NO21	GR/Mis,ML/Mos,Eas/Che	70
SO43	Ste Anne/LaBroquerie	71
SO21	Nortre Dame/St. Claude	72
NO14	Thomp,Myst Lake	73

Continued on next page

---

**Table A.1 – continued from previous page**

Code	Region	Small area
SO41	Niverville/Richot	74
IE61	Northern Remote	75
SO46	Rural East	76
NO22	Puk/Mat Col CN	77
NO23	SayD(TL),Bro/BL,NoL(Lac)	78
NO26	Bu(OH),MS(GR),GLN/GLFN	79
NO28	Norway House/NH CN	80
NO27	Cross Lake/Pimi CN	81
NO31	Island Lake	82
IE33	Whiteshell	83
WE14	Agassiz Mountain	84
WE16	Swan River	85
WE23	Bdn Downtown	86
WE24	Bdn South End	87
WE22	Bdn North Hill	88
WE25	Bdn East End	89
WE21	Bdn West End	90
IE53	Eriksdale/Ashern	91
NO12	The Pas/OCN,Kels	92
NO15	Bay Line	93
NO16	Gillam,Fox Lake CN	94
NO25	Sham,YorkF,Tat(SPL)	95
NO24	Nelson House/NCN	96

## Appendix B: Ordinal Logistic Simulation Probabilities

Plots were made using the mean of the responses  $y_{ijk}$  and probabilities  $\pi_{m,k} = \text{Prob}(y_{ijk} \leq c)$  using 500 simulation runs. These plots are shown for areas  $k=1$  and  $k=4$  but note that all 20 areas show similar patterns. There are 80 equal observations for each area. The plots and tables below show that in the 500 simulations, the probabilities of  $y_{ijk}$  being 1-3 are approximately equal. Although there is minor variation between areas, the distributions between areas behave similarly due to the low bias and MSE of the 500 simulation runs.

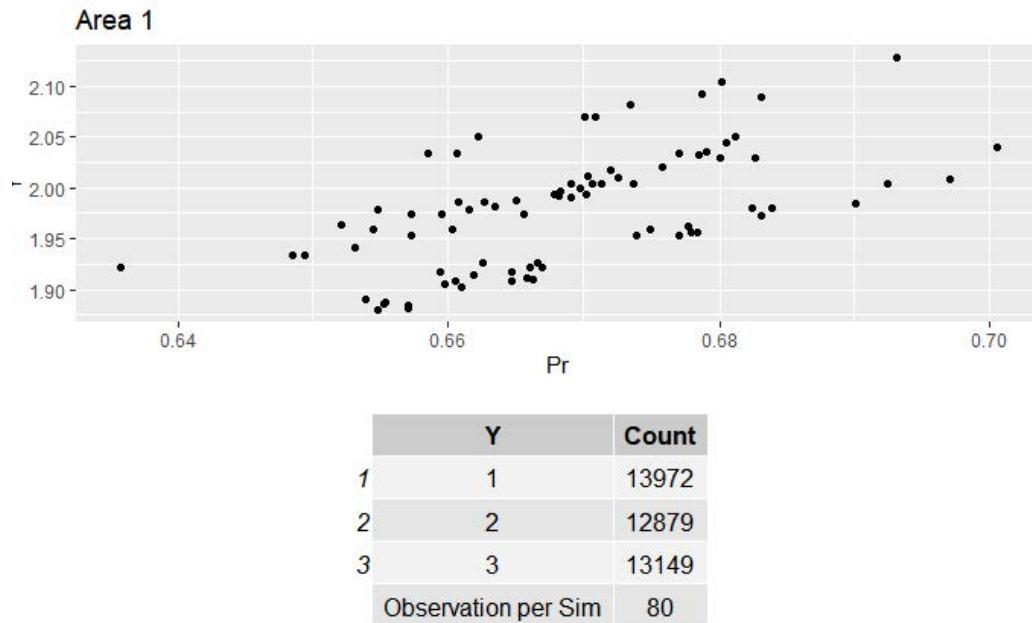
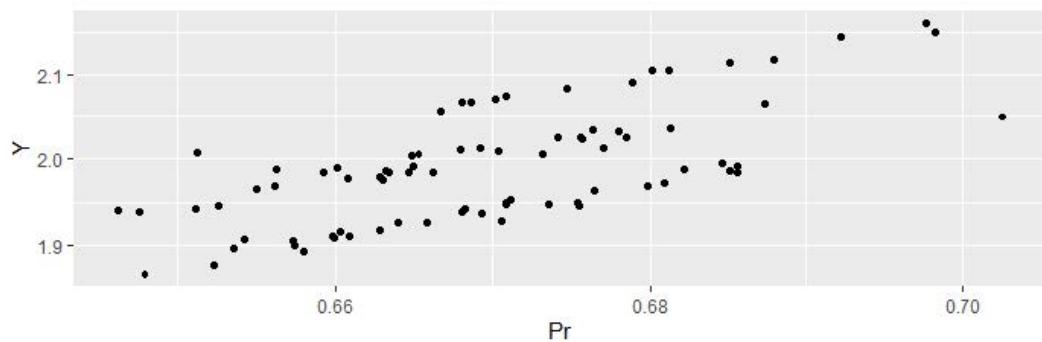


Figure B.1: Plot of average  $y_{ijk}$ 's vs corresponding average probabilities from 500 simulation runs for area one.

---

**Area 4**

	<b>Y</b>	<b>Count</b>
1	1	13728
2	2	12905
3	3	13367
Observation per Sim		80

Figure B.2: Plot of average  $y_{ijk}$ 's vs corresponding average probabilities from 500 simulation runs for area four.

## Appendix C: Logistic Model - Univariate Analysis

Parameter	Estimate	95% credible interval (lower, upper)
Mother wheezed	35.70	(19.66, 50.10)
$\rho$	0.04	(0.02, 0.09)
$\sigma_v^2$	98.70	(94.80, 99.90)
$\sigma_f^2$	91.20	(76.20, 99.70)

Table C.1: Univariate logistic model with random effects and Mother Wheezed as a covariate. All parameters are statistically significant.

Parameter	Estimate	95% credible interval (lower, upper)
Mother smoked	-17.70	(-50.80, 14.20)
$\rho$	0.06	(0.03, 0.10)
$\sigma_v^2$	98.50	(94.60, 99.90)
$\sigma_f^2$	95.50	(86.60, 99.90)

Table C.2: Univariate logistic model with random effects and Mother Smoked as a covariate. Mother smoked is not statistically significant.

Parameter	Estimate	95% credible interval (lower, upper)
Father wheezed	-13.30	(-30.30, 1.97)
$\rho$	0.06	(-30.30, 1.97)
$\sigma_v^2$	98.40	(94.20, 100.00)
$\sigma_f^2$	95.10	(84.60, 99.80)

Table C.3: Univariate logistic model with random effects and Father Wheezed as a covariate. Father wheezed and  $\rho$  are not statistically significant.

---

Parameter	Estimate	95% credible interval (lower, upper)
Father smoked	-21.70	(-54.00, -32.00)
$\rho$	0.06	(0.03, 0.10)
$\sigma_v^2$	98.40	(94.40, 99.90)
$\sigma_f^2$	94.00	(84.20, 99.80)

Table C.4: Univariate logistic model with random effects and Father Smoked as a covariate. Father smoked is statistically significant.

Parameter	Estimate	95% credible interval (lower, upper)
Asthma at 3 years	61.50	(46.30, 81.90)
$\rho$	0.05	(0.01, 0.10)
$\sigma_v^2$	98.00	(94.7, 99.9)
$\sigma_f^2$	88.0	(76.1, 99.0)

Table C.5: Univariate logistic model with random effects and Asthma Young as a covariate. All parameters are statistically significant.

Parameter	Estimate	95% credible interval (lower, upper)
Asthma at 5 years	37.30	(23.50, 51.50)
$\rho$	0.02	(0.02, 0.08)
$\sigma_v^2$	94.40	(94.35, 99.90)
$\sigma_f^2$	77.10	(77.05, 99.40)

Table C.6: Univariate logistic model with random effects and Asthma older as a covariate. All parameters are statistically significant.

Parameter	Estimate	95% credible interval (lower, upper)
Mother asthma (3 months)	29.40	(12.07, 45.67)
$\rho$	0.04	(0.02, 0.07)
$\sigma_v^2$	98.40	(94.40, 99.96)
$\sigma_f^2$	88.68	(78.09, 98.68)

Table C.7: Univariate logistic model with random effects and Mother asthma as a covariate. All parameters are statistically significant.

---

Parameter	Estimate	95% credible interval (lower, upper)
Diagnosed mother asthma	31.80	(16.96, 48.80)
$\rho$	0.05	(0.02, 0.08)
$\sigma_v^2$	98.30	(94.01, 99.96)
$\sigma_f^2$	87.00	(72.82, 99.37)

Table C.8: Univariate logistic model with random effects and Diagnosed Mother Asthma as a covariate. All parameters are statistically significant.

Parameter	Estimate	95% credible interval (lower, upper)
Mother treated for asthma	26.00	(13.28, 39.47)
$\rho$	0.05	(0.02, 0.08)
$\sigma_v^2$	98.50	(94.58, 99.96)
$\sigma_f^2$	90.10	(79.88, 98.69)

Table C.9: Univariate logistic model with random effects and mother treated for asthma as a covariate. All parameters are statistically significant.

Parameter	Estimate	95% credible interval (lower, upper)
Father asthma	-9.40	(-26.81, 7.81)
$\rho$	0.06	(0.03, 0.1)
$\sigma_v^2$	98.0	(95.57, 99.96)
$\sigma_f^2$	95.0	(82.86, 99.85)

Table C.10: Univariate logistic model with random effects and Father asthma as a covariate. Father asthma is not statistically significant.

Parameter	Estimate	95% credible interval (lower, upper)
Diagnosed father asthma	-5.83	(-24.98, 10.88)
$\rho$	0.06	(0.03, 0.09)
$\sigma_v^2$	98.00	(94.05, 99.96)
$\sigma_f^2$	93.00	(81.57, 99.71)

Table C.11: Univariate logistic model with random effects and Diagnosed father asthma as a covariate. Diagnosed father asthma is not statistically significant.

---

Parameter	Estimate	95% credible interval (lower, upper)
Father treated for asthma	2.10	(-13.48, 19.42)
$\rho$	0.06	(0.30, 0.09)
$\sigma_v^2$	98.00	(94.16, 99.97)
$\sigma_f^2$	92.60	(81.99, 96.65)

Table C.12: Univariate logistic model with random effects and Father treated for asthma for as a covariate. Father treated for asthma is not statistically significant.

Parameter	Estimate	95% credible interval (lower, upper)
Furry pets	-13.40	(-25.65, -0.55)
$\rho$	0.06	(0.03, 0.10)
$\sigma_v^2$	98.40	(94.19, 99.94)
$\sigma_f^2$	95.70	(87.96, 99.85)

Table C.13: Univariate logistic model with random effects and Furry pets as a covariate. Furry pets is statistically significant.

Parameter	Estimate	95% credible interval (lower, upper)
SES	-2.46	(-11.28, 7.09)
$\rho$	0.06	(0.03, 0.1)
$\sigma_v^2$	98.00	(94.24, 99.96)
$\sigma_f^2$	93.00	(82.72, 99.68)

Table C.14: Univariate logistic model with random effects and SES as a covariate. SES is not statistically significant.

Parameter	Estimate	95% credible interval (lower, upper)
Highway	-8.40	(-26.36, 9.63)
$\rho$	0.07	(0.03, 0.14)
$\sigma_v^2$	98.00	(93.84, 99.97)
$\sigma_f^2$	92.00	(81.59, 99.52)

Table C.15: Univariate logistic model with random effects and Highway as a covariate. Highway is not statistically significant.

---

Parameter	Estimate	95% credible interval (lower, upper)
Water	-3.88	(-26.79, 17.07)
$\rho$	0.05	(0.02, 0.09)
$\sigma_v^2$	98.30	(93.94, 99.97)
$\sigma_f^2$	92.50	(83.05, 99.62)

Table C.16: Univariate logistic model with random effects and Water as a covariate. Water is not statistically significant.

Parameter	Estimate	95% credible interval (lower, upper)
Factory	8.90	(-41.84, 51.49)
$\rho$	0.06	(0.03, 0.010)
$\sigma_v^2$	98.38	(94.07, 99.97)
$\sigma_f^2$	92.00	(78.53, 99.75)

Table C.17: Univariate logistic model with random effects and Factory as a covariate. Factory is not statistically significant.

Parameter	Estimate	95% credible interval (lower, upper)
Farm	-31.00	(-57.29, -5.99)
$\rho$	0.10	(0.03, 0.10)
$\sigma_v^2$	98.00	(93.96, 99.95)
$\sigma_f^2$	94.00	(85.52, 99.82)

Table C.18: Univariate logistic model with random effects and Factory as a covariate. All parameters are statistically significant.

Parameter	Estimate	95% credible interval (lower, upper)
Construction	-9.50	(-29.88, 8.96)
$\rho$	0.07	(0.03, 0.12)
$\sigma_v^2$	98.00	(94.40, 99.95)
$\sigma_f^2$	94.00	(85.75, 99.82)

Table C.19: Univariate logistic model with random effects and Construction as a covariate. Construction is not statistically significant.

## Appendix D: Logistic Model Building Tables

Parameter	Estimate	95% credible interval (lower, upper)
Asthma young	52.40	(35.43, 70.64)
Asthma older	17.50	(1.98, 34.15)
$\rho$	0.05	(0.02, 0.08)
$\sigma_v^2$	98.00	(94.50, 99.97)
$\sigma_f^2$	86.00	(75.23, 96.68)

Table D.1: Model with two covariates: Asthma young and Asthma older.

Parameter	Estimate	95% credible interval (lower, upper)
Asthma young	51.00	(34.78, 72.80)
Asthma older	13.00	(-2.03, 29.12)
Mother wheezed	31.00	(19.12, 43.73)
$\rho$	0.03	(0.01, 0.07)
$\sigma_v^2$	98.0	(94.20, 99.96)
$\sigma_f^2$	79.0	(67.42, 91.79)

Table D.2: Model with three covariates: Asthma young, Asthma older and Mother wheezed.

Parameter	Estimate	95% credible interval (lower, upper)
Asthma young	53.00	(34.59, 72.33)
Asthma older	14.00	(-1.03, 28.63)
Mother wheezed	27.00	(7.78, 42.55)
Diagnosed mother asthma	8.70	(-8.84, 26.52)
$\rho$	0.03	(0.01, 0.07)
$\sigma_v^2$	99.00	(94.55, 99.96)
$\sigma_f^2$	82.00	(68.60, 96.08)

Table D.3: Model with four covariates: Asthma young, Asthma older, Mother wheezed and Diagnosed mother asthma.

---

Parameter	Estimate	95% credible interval (lower, upper)
Asthma at 3 years	50.00	(32.11, 66.23)
Asthma at 5 years	15.00	(1.65, 29.54)
Mother wheezed (3 months)	30.00	(1.33, 45.15)
Diagnosed mother asthma	340.00	(112.40, 748.22)
Mother asthma	-340.00	(-743.30, -106.41)
$\rho$	0.03	(0.01, 0.069)
$\sigma_v^2$	98.40	(93.58, 99.96)
$\sigma_f^2$	79.79	(59.83, 93.93)

Table D.4: Model with five covariates. Mother asthma and Diagnosed mother asthma did not converge and both have high standard errors. One of the variables will need to be removed eventually.

Parameter	Estimate	95% credible interval (lower, upper)
Asthma young	53.00	(35.04, 72.70)
Asthma older	13.00	(-1.48, 29.78)
Mother wheezed	28.00	(8.00, 47.42)
Diagnosed mother asthma	-330.00	(184.98, 515.38)
Mother asthma	-320.00	(-507.44, -177.83)
Farm	-28.0	(-57.42, 0.06)
$\rho$	0.04	(0.01, 0.07)
$\sigma_v^2$	98.0	(94.32, 99.96)
$\sigma_f^2$	84.3	(65.92, 97.79)

Table D.5: Model with six covariates: Asthma young, Asthma older, Mother wheezed, Diagnosed mother Asthma older, Mother asthma and newly added Farm.

Parameter	Estimate	95% credible interval (lower, upper)
Asthma young	59.90	(39.98, 78.24)
Asthma older	11.60	(-4.26, 26.70)
Mother wheezed	30.00	(10.31, 47.92)
Diagnosed mother asthma	543.00	(243.54, 804.80)
Mother asthma	-535.32	(-797.84, -230.98)
Farm	-30.29	(-61.26, 3.57)
Furry pets	-13.61	(-26.37, -1.45)
$\rho$	0.04	(0.01, 0.09)
$\sigma_v^2$	99.55	(94.87, 99.96)
$\sigma_f^2$	89.74	(79.20, 99.12)

Table D.6: Model with seven covariates. Asthma young,Asthma older,Mother wheezed, Diagnosed mother asthma older,Mother asthma, Farm added alongside with Furry pets.

Parameter	Estimate	95% credible interval (lower, upper)
Asthma young	55.30	(35.19, 74.22)
Asthma older	14.00	(-1.52, 27.76)
Mother wheezed	29.24	(11.25, 48.91)
Diagnosed mother asthma	5.90	(-10.86, 4.90)
Farm	-30.20	(-56.82, -4.57)
$\rho$	0.04	(0.01,0.07)
$\sigma_v^2$	98.50	(94.66, 99.96)
$\sigma_f^2$	85.40	(69.39, 99.08)

Table D.7: Removed Mother asthma from Table D.5. After removing this variable, the model converged.

Parameter	Estimate	95% credible interval (lower, upper)
Asthma young	56.00	(38.06, 76.68)
Asthma older	12.00	(-6.07, 27.17)
Mother wheezed	30.00	(12.18, 49.03)
Diagnosed mother asthma	7.30	(-16.01, 28.20)
Farm	-34.00	(-65.74, -6.41)
Furry pets	-15.00	(-31.08, -2.24)
$\rho$	0.04	(0.01,0.08)
$\sigma_v^2$	99.00	(94.70, 99.96)
$\sigma_f^2$	90.00	(76.18, 99.42)

Table D.8: Added the covariate Furry pets back to the model in Table D.7.

---

Parameter	Estimate	95% credible interval (lower, upper)
Asthma young	56.00	(39.54, 72.83)
Asthma older	13.00	(-3.78, 28.59)
Mother wheezed	32.00	(14.31, 49.24)
Diagnosed mother asthma	6.05	(-9.71, 23.63)
Furry pets	-15.00	(-29.34, -0.98)
$\rho$	0.04	(0.01, 0.08)
$\sigma_v^2$	99.00	(95.02, 99.97)
$\sigma_f^2$	88.00	(75.23, 98.90)

Table D.9: Removed Farm from the model in Table D.8. Removing Farm does not make any significant changes to the model.

## Appendix E: ZIP Model - Univariate Analysis

Parameter	Estimate	95% credible interval (lower, upper)
Intercept	-1.32	(-1.78, -0.90 )
Mother asthma (prenatal)	1.03	(0.72, 1.36 )
$\gamma_0$	1.38	(1.15, 1.60 )
$1 - \pi$	0.20	(0.17, 0.24 )
$\rho$	0.00	(0.00, 0.00 )
$\sigma_v^2$	0.40	(0.09, 0.67 )
$\sigma_f^2$	1.07	(0.86, 1.29 )

Table E.1: Univariate ZIP model with random effects and Mother asthma as a covariate. All parameters are statistically significant.

Parameter	Estimate	95% credible interval (lower, upper)
Intercept	-1.16	(-1.62, -0.68 )
Mother wheezed (prenatal)	0.95	(0.60, 1.32 )
$\gamma_0$	1.40	(1.16, 1.65 )
$1 - \pi$	0.20	(0.16, 0.24 )
$\rho$	0.00	(0.00, 0.00 )
$\sigma_v^2$	0.39	(0.11, 0.66 )
$\sigma_f^2$	1.08	(0.83, 1.30 )

Table E.2: Univariate ZIP model with random effects and Mother wheezed as a covariate. All parameters are statistically significant.

---

Parameter	Estimate	95% credible interval (lower, upper)
Intercept	-0.61	(-1.03, -0.22 )
Mother smoked (prenatal)	-0.56	(-1.32, 0.19 )
$\gamma_0$	1.59	(1.35, 1.82 )
$1 - \pi$	0.17	(0.14, 0.20 )
$\rho$	0.00	(0.00, 0.00 )
$\sigma_v^2$	0.37	(0.08, 0.65 )
$\sigma_f^2$	0.98	(0.77, 1.19 )

Table E.3: Univariate ZIP model with random effects and Mother smoked as a covariate. Mother smoked is not statistically significant.

Parameter	Estimate	95% credible interval (lower, upper)
Intercept	-0.66	(-1.09, -0.27)
Father wheezed	0.20	(-0.14, 0.54 )
$\gamma_0$	1.61	(1.37, 1.84 )
$1 - \pi$	0.17	(0.14, 0.20 )
$\rho$	0.00	(0.00, 0.00 )
$\sigma_v^2$	0.37	(0.08, 0.64 )
$\sigma_f^2$	0.96	(0.74, 1.18 )

Table E.4: Univariate ZIP model with random effects and Father wheezed as a covariate. Father wheezed is not statistically significant.

Parameter	Estimate	95% credible interval (lower, upper)
Intercept	-0.62	(-1.04, -0.23 )
Father smoked	-0.57	(-1.33, 0.19 )
$\gamma_0$	1.58	(1.34, 1.81 )
$1 - \pi$	0.17	(0.14, 0.21 )
$\rho$	0.00	(0.00, 0.00 )
$\sigma_v^2$	0.37	(0.04, 0.64 )
$\sigma_f^2$	0.99	(0.79, 1.23 )

Table E.5: Univariate ZIP model with random effects and Father smoked as a covariate. Father smoked is not statistically significant.

---

Parameter	Estimate	95% credible interval (lower, upper)
Intercept	-0.75	(-1.24, -0.33 )
Furry pets	0.14	(-0.18, 0.47 )
$\gamma_0$	1.56	(1.31, 1.79 )
$1 - \pi$	0.17	(0.14, 0.21 )
$\rho$	0.00	(0.00, 0.00 )
$\sigma_v^2$	0.40	(0.12, 0.66 )
$\sigma_f^2$	1.00	(0.80, 1.25 )

Table E.6: Univariate ZIP model with random effects and Furry pets as a covariate. Furry pets is not statistically significant.

Parameter	Estimate	95% credible interval (lower, upper)
Intercept	-0.66	(-1.08, -0.25 )
SES	-0.13	(-0.32, 0.07 )
$\gamma_0$	1.58	(1.33, 1.81 )
$1 - \pi$	0.17	(0.14, 0.21 )
$\rho$	0.00	(0.00, 0.00 )
$\sigma_v^2$	0.36	(0.04, 0.63 )
$\sigma_f^2$	0.99	(0.75, 1.23 )

Table E.7: Univariate ZIP model with random effects and SES as a covariate. SES is not statistically significant.

Parameter	Estimate	95% credible interval (lower, upper)
Intercept	-0.71	(-1.14, -0.33 )
Highway	0.24	(-0.13, 0.62 )
$\gamma_0$	1.58	(1.34, 1.80 )
$1 - \pi$	0.17	(0.14, 0.21 )
$\rho$	0.00	(0.00, 0.00 )
$\sigma_v^2$	0.38	(0.08, 0.66 )
$\sigma_f^2$	0.99	(0.79, 1.21 )

Table E.8: Univariate ZIP model with random effects and Highway as a covariate. Highway is not statistically significant.

---

Parameter	Estimate	95% credible interval (lower, upper)
Intercept	-0.67	(-1.10, -0.27 )
Water	0.23	(-0.23, 0.72 )
$\gamma_0$	1.59	(1.35, 1.82 )
$1 - \pi$	0.17	(0.14, 0.21 )
$\rho$	0.00	(0.00, 0.00 )
$\sigma_v^2$	0.38	(0.14, 0.64)
$\sigma_f^2$	0.98	(0.77, 1.22 )

Table E.9: Univariate ZIP model with random effects and Water as a covariate. Water is not statistically significant.

Parameter	Estimate	95% credible interval (lower, upper)
Intercept	-0.68	(-1.12, -0.29 )
Factory	0.36	(-0.62, 1.39 )
$\gamma_0$	1.57	(1.32, 1.80 )
$1 - \pi$	0.17	(0.14, 0.21 )
$\rho$	0.00	(0.00, 0.00 )
$\sigma_v^2$	0.37	(0.10, 0.64 )
$\sigma_f^2$	1.00	(0.79, 1.25 )

Table E.10: Univariate ZIP model with random effects and Factory as a covariate. Factory is not statistically significant.

Parameter	Estimate	95% credible interval (lower, upper)
Intercept	-0.70	(-1.12, -0.32 )
Farm	0.44	(-0.05, 0.92 )
$\gamma_0$	1.58	(1.35, 1.81 )
$1 - \pi$	0.17	(0.14, 0.21 )
$\rho$	0.00	(0.00, 0.00 )
$\sigma_v^2$	0.34	(0.03, 0.63)
$\sigma_f^2$	0.98	(0.77, 1.22)

Table E.11: Univariate ZIP model with random effects and Farm as a covariate. Farm is not statistically significant.

---

Parameter	Estimate	95% credible interval (lower, upper)
Intercept	-0.60	(-1.01, -0.22 )
Construction	-0.20	(-0.67, 0.28 )
$\gamma_0$	1.60	(1.36, 1.82 )
$1 - \pi$	0.17	(0.14, 0.20 )
$\rho$	0.00	(0.00, 0.00 )
$\sigma_v^2$	0.36	(0.08, 0.63 )
$\sigma_f^2$	0.98	(0.77, 1.19 )

Table E.12: Univariate ZIP model with random effects and Construction as a covariate. Construction is not statistically significant.

## Appendix F: ZIP Model Building Tables

Parameter	Estimate	95% credible interval(lower, upper)
Intercept	-1.34	(-1.82,-0.91 )
Mother asthma (prenatal)	0.08	(-0.34, 0.50 )
Mother wheezed (prenatal)	1.00	(0.61, 1.39 )
$\gamma_0$	1.38	(1.14, 1.60 )
$1 - \pi$	0.20	(0.17, 0.24 )
$\rho$	0.00	(0.00, 0.00 )
$\sigma_v^2$	0.40	(0.12, 0.68 )
$\sigma_f^2$	1.07	(0.88, 1.30 )
DIC	3076	-

Table F.1: ZIP model with two covariates,Mother asthma (prenatal) and Mother wheezed (prenatal).

Parameter	Estimate	95% credible interval (lower, upper)
Intercept	-1.28	(-1.71, -0.84 )
Mother asthma (prenatal)	0.06	(-0.37, 0.48 )
Mother wheezed (prenatal)	1.00	(0.62, 1.40 )
Mother smoked (prenatal)	-0.60	(-1.41, 0.21 )
$\gamma_0$	1.40	(1.18, 1.62 )
$1 - \pi$	0.20	(0.16, 0.24 )
$\rho$	0.00	(0.00, 0.00 )
$\sigma_v^2$	0.38	(0.06, 0.65 )
$\sigma_f^2$	1.06	(0.85, 1.26 )
DIC	2995	-

Table F.2: ZIP model with three covariates: Mother asthma (prenatal), Mother wheezed (prenatal) and Mother smoked (prenatal).

---

Parameter	Estimate	95% credible interval (lower, upper)
Intercept	-1.22	(-1.71, -0.80 )
Mother asthma (prenatal)	0.01	(-0.41, 0.42 )
Mother wheezed (prenatal)	1.03	(0.65, 1.43 )
Mother smoked (prenatal)	-0.23	(-1.22, 0.72 )
Father smoked	-0.62	(-1.56, 0.33 )
$\gamma_0$	1.42	(1.18, 1.64 )
$1 - \pi$	0.20	(0.16, 0.24 )
$\rho$	0.00	(0.00, 0.00 )
$\sigma_v^2$	0.39	(0.07, 0.67 )
$\sigma_f^2$	1.03	(0.82, 1.27 )
DIC	2954	-

Table F.3: ZIP model with four covariates. Father smoked added to the previous model.

Parameter	Estimate	95% credible interval (lower, upper)
Intercept	-1.33	(-1.80, -0.90)
Mother asthma (prenatal)	0.00	(-0.42, 0.42 )
Mother wheezed (prenatal)	1.09	(0.68, 1.49 )
Mother smoked (prenatal)	-0.20	(-1.16, 0.77 )
Father smoked	-0.67	(-1.61, 0.25 )
Farm	0.62	(0.15, 1.11 )
$\gamma_0$	1.41	(1.18, 1.64 )
$1 - \pi$	0.20	(0.16, 0.24 )
$\rho$	0.00	(0.00, 0.00 )
$\sigma_v^2$	0.33	(0.05, 0.61 )
$\sigma_f^2$	1.03	(0.83, 1.25 )
DIC	3142	-

Table F.4: ZIP model with five covariates. Farm added to the previous model.

---

Parameter	Estimate	95% credible interval (lower, upper)
Intercept	-1.33	(-1.83, -0.87 )
Mother asthma (prenatal)	0.00	(-0.42, 0.41 )
Mother wheezed (prenatal)	1.07	(0.70, 1.47 )
Mother smoked (prenatal)	-0.23	(-1.22, 0.75 )
Father smoked	-0.65	(-1.59, 0.28 )
Farm	0.61	(0.12, 1.11 )
Furry pets	0.07	(-0.24, 0.39 )
$\gamma_0$	1.43	(1.19, 1.65 )
$1 - \pi$	0.19	(0.16, 0.23 )
$\rho$	0.00	(0.00, 0.00 )
$\sigma_v^2$	0.34	(0.02, 0.63 )
$\sigma_f^2$	1.01	(0.80, 1.24 )
DIC	3066	-

Table F.5: ZIP model with six covariates. Furry pets added to the previous model.

## Appendix G: Area-Level Logistic Model- Univari- ate Analysis

Parameter	Estimate	95% credible interval (lower, upper)
Intercept	-2.44	(-2.85, -2.11)
Mother smoked (prenatal)	0.01	(-0.02, 0.03 )
$\sigma_v^2$	0.34	(0.03, 0.98)

Table G.1: Univariate area-level logistic model with prenatal Mother smoked as a covariate. Mother smoked is not statistically significant.

Parameter	Estimate	95% credible interval (lower, upper)
Intercept	-3.06	(-3.71, -2.44)
Mother wheezed (prenatal)	0.02	(0.01, 0.04)
$\sigma_v^2$	0.28	(0.03, 0.85)

Table G.2: Univariate area-level logistic model with prenatal Mother wheezed as a covariate. Mother smoked is statistically significant.

Parameter	Estimate	95% credible interval (lower, upper)
Intercept	-3.17	(-3.83, -2.57)
Mother asthma (prenatal)	0.03	(0.01, 0.05)
$\sigma_v^2$	0.25	(0.03, 0.76)

Table G.3: Univariate area-level logistic model with prenatal Mother asthma as a covariate. Mother smoked is statistically significant.

---

Parameter	Estimate	95% credible interval (lower, upper)
Intercept	-2.33	(-2.72, -1.98)
Father smoked	-0.02	(-0.06, 0.02)
$\sigma_v^2$	0.33	(0.03, 0.92)

Table G.4: Univariate area-level logistic model with Father smoked as a covariate. Father smoked is not statistically significant.

Parameter	Estimate	95% credible interval (lower, upper)
Intercept	-2.89	(-3.53, -2.28)
Father wheezed	0.02	(0.00, 0.03)
$\sigma_v^2$	0.26	(0.03, 0.86)

Table G.5: Univariate area-level logistic model with Father wheezed as a covariate. Father wheezed is statistically significant.

Parameter	Estimate	95% credible interval (lower, upper)
Intercept	-2.37	(-2.96, -1.82)
Father asthma	0.00	(-0.03, 0.02)
$\sigma_v^2$	0.35	(0.04, 0.91)

Table G.6: Univariate area-level logistic model with Father asthma as a covariate. Father asthma is not statistically significant.

Parameter	Estimate	95% credible interval (lower, upper)
Intercept	-2.63	(-3.32, -2.08)
Highway	0.01	(-0.01, 0.03)
$\sigma_v^2$	0.38	(0.03, 1.01)

Table G.7: Univariate area-level logistic model with Highway as a covariate. Highway is not statistically significant.

---

Parameter	Estimate	95% credible interval (lower, upper)
Intercept	-2.30	(-2.81, -1.83)
Water	-0.01	(-0.04, 0.02)
$\sigma_v^2$	0.37	(0.03, 1.02)

Table G.8: Univariate area-level logistic model with Water as a covariate. Water is not statistically significant.

Parameter	Estimate	95% credible interval (lower, upper)
Intercept	-2.41	(-2.82, -2.09)
Factory	0.00	(-0.07, 0.05)
$\sigma_v^2$	0.37	(0.03, 1.02)

Table G.9: Univariate area-level logistic model with Factory as a covariate. Factory is not statistically significant.

Parameter	Estimate	95% credible interval(lower, upper)
Intercept	-2.24	(-2.61, -1.90)
Farm	-0.02	(-0.04, 0.00)
$\sigma_v^2$	0.32	(0.03, 0.94)

Table G.10: Univariate area-level logistic model with Farm as a covariate. Farm is not statistically significant.

Parameter	Estimate	95% credible interval (lower, upper)
Intercept	-2.35	(-2.90, -1.88)
Construction	0.00	(-0.03, 0.02)
$\sigma_v^2$	0.35	(0.04, 0.93)

Table G.11: Univariate area-level logistic model with Construction as a covariate. Construction is not statistically significant.

---

Parameter	Estimate	95% credible interval (lower, upper)
Intercept	-2.40	(-2.78, -2.11)
SES	-0.07	(-0.75, 0.60)
$\sigma_v^2$	0.34	(0.03, 0.93 )

Table G.12: Univariate area-level logistic model with SES as a covariate. SES is not statistically significant.

## Appendix H: Area-Level Logistic- Model Building

Parameter	Estimate	95% credible interval (lower, upper)
Intercept	-3.43	(-4.25, -2.71)
Mother wheezed (prenatal)	0.01	(-0.01, 0.03)
Mother asthma (prenatal)	0.02	(0.00, 0.05)
$\sigma_v^2$	0.28	(0.03, 0.79 )
DIC	127.1	-

Table H.1: Area-level logistic model with two prenatal covariates, Mother wheezed and Mother asthma. Odds ratios were calculated by taking the exponent of the covariate parameter estimates.

Parameter	Estimate	95% credible interval (lower, upper)
Intercept	-3.42	(-4.21, -2.69)
Mother wheezed (prenatal)	0.01	(-0.01, 0.03 )
Mother asthma (prenatal)	0.03	(0.00, 0.05 )
Mother smoked (prenatal)	0.00	(-0.03, 0.03 )
$\sigma_v^2$	0.26	(0.03, 0.81 )
DIC	132.7	-

Table H.2: Area-level logistic model with three prenatal covariates, Mother wheezed, Mother asthma and Mother smoked.

## Appendix I: Area-Level ZIP- Univariate Analysis

Parameter	Estimate	95% credible interval (lower, upper)
Intercept	-0.53	(-1.36, 0.16 )
Mother smoked (prenatal)	-0.02	(-0.05, 0.01 )
$\gamma_0$	-555.00	(-1780.00, -93.00 )
$1 - \pi$	1.00	(1.00, 1.00 )
$\sigma_v^2$	1.92	(1.34, 2.72 )

Table I.1: Univariate area-level ZIP model with random effects and prenatal Mother smoked as a covariate. Mother smoked is not statistically significant.

Parameter	Estimate	95% credible interval (lower, upper)
Intercept	-0.69	(-1.91, 0.31 )
Mother wheezed (prenatal)	0.00	(-0.02, 0.02 )
$\gamma_0$	-801.00	(-2230.00, -30.20 )
$1 - \pi$	1.00	(1.00, 1.00 )
$\sigma_v^2$	1.97	(1.38, 2.80 )

Table I.2: Univariate area-level ZIP model with random effects and prenatal Mother wheezed as a covariate. Mother wheezed is not statistically significant.

Parameter	Estimate	95% credible interval (lower, upper)
Intercept	-1.22	(-2.38, -0.31 )
Mother asthma (prenatal)	0.02	(0.00, 0.05 )
$\gamma_0$	-796.00	(-2250, -29.90 )
$1 - \pi$	1.00	(1.00, 1.00 )
$\sigma_v^2$	1.91	(1.34, 2.72 )

Table I.3: Univariate area-level ZIP model with random effects and prenatal Mother asthma as a covariate. Mother asthma is statistically significant.

Parameter	Estimate	95% credible interval (lower, upper)
Intercept	-0.40	(-1.24, 0.30 )
Father smoked (prenatal)	-0.04	(-0.10, 0.00 )
$\gamma_0$	-795.00	(-2230, -25.60 )
$1 - \pi$	1.00	(1.00, 1.00 )
$\sigma_v^2$	1.91	(1.32, 2.71 )

Table I.4: Univariate area-level ZIP model with random effects and prenatal Father smoked as a covariate. Father smoked is not statistically significant.

Parameter	Estimate	95% credible interval (lower, upper)
Intercept	-0.70	(-1.88, 0.31 )
Father wheezed (prenatal)	0.00	(-0.03, 0.02 )
$\gamma_0$	-795.00	(-2230.00, -28.20 )
$1 - \pi$	1.00	(1.00, 1.00 )
$\sigma_v^2$	1.98	(1.38, 2.82 )

Table I.5: Univariate area-level ZIP model with random effects and prenatal Father wheezed as a covariate. Father wheezed is not statistically significant.

Parameter	Estimate	95% credible interval (lower, upper)
Intercept	-0.51	(-1.60, 0.39 )
Father asthma (prenatal)	-0.01	(-0.04, 0.02 )
$\gamma_0$	-489.00	(-1660.00, -32.20 )
$1 - \pi$	1.00	(1.00, 1.00 )
$\sigma_v^2$	1.96	(1.36, 2.79 )

Table I.6: Univariate area-level ZIP model with random effects and prenatal Father asthma as a covariate. Father asthma is not statistically significant.

Parameter	Estimate	95% credible interval (lower, upper)
Intercept	-1.06	(-2.25, -0.13 )
Highway	0.02	(-0.01, 0.05 )
$\gamma_0$	-988.00	(-2040.00, -44.70 )
$1 - \pi$	1.00	(1.00, 1.00 )
$\sigma_v^2$	1.96	(1.37, 2.77 )

Table I.7: Univariate area-level ZIP model with random effects and Highway as a covariate. Highway is not statistically significant.

---

Parameter	Estimate	95% credible interval (lower, upper)
Intercept	-0.59	( -1.62, 0.27 )
Water	-0.01	( -0.06, 0.03 )
$\gamma_0$	-755.00	( -2160.00, -22.00 )
$1 - \pi$	1.00	( 1.00, 1.00 )
$\sigma_v^2$	1.97	( 1.36, 2.84 )

Table I.8: Univariate area-level ZIP model with random effects and Water as a covariate. Water is not statistically significant.

Parameter	Estimate	95% credible interval (lower, upper)
Intercept	-0.65	( -1.52, 0.06 )
Factory	-0.03	( -0.11, 0.04 )
$\gamma_0$	-729.00	( -2090.00, -25.10 )
$1 - \pi$	1.00	( 1.00, 1.00 )
$\sigma_v^2$	1.97	( 1.38, 2.79 )

Table I.9: Univariate area-level ZIP model with random effects and Factory as a covariate. Factory is not statistically significant.

Parameter	Estimate	95% credible interval (lower, upper)
Intercept	0.07	( -0.69, 0.73 )
Farm	-0.04	( -0.06, -0.01 )
$\gamma_0$	-797.00	( -2260.00, -31.40 )
$1 - \pi$	1.00	( 1.00, 1.00 )
$\sigma_v^2$	1.69	( 1.17, 2.42 )

Table I.10: Univariate area-level ZIP model with random effects and Farm as a covariate. Farm is not statistically significant.

Parameter	Estimate	95% credible interval (lower, upper)
Intercept	-0.96	( -2.03, -0.11 )
Construction	0.02	( -0.02, 0.07 )
$\gamma_0$	-805.00	( -2170.00, -36.10 )
$1 - \pi$	1.00	( 1.00, 1.00 )
$\sigma_v^2$	1.97	( 1.39, 2.77 )

Table I.11: Univariate area-level ZIP model with random effects and Construction as a covariate. Construction is not statistically significant.

---

Parameter	Estimate	95% credible interval (lower, upper)
Intercept	-0.46	(-1.29, 0.29 )
SES	0.96	(-0.14, 2.12 )
$\gamma_0$	-793.00	(-2350.00, -5.14 )
$1 - \pi$	0.99	(0.99, 1.00 )
$\sigma_v^2$	1.84	(1.23, 2.62 )

Table I.12: Univariate area-level ZIP model with random effects and SES as a covariate. SES is not statistically significant.

## Appendix J: Area-Level ZIP Model- Covariates for Both Components

Parameter	Estimate	95% credible interval (lower, upper)
Intercept	1.10	(0.50, 1.63 )
Mother smoked (prenatal)	-0.02	(-0.06, 0.01 )
$\sigma_v^2$	1.13	(0.75, 1.67 )

Table J.1: Univariate area-level ZIP model with Mother smoked as a covariate for both the logistic and Poisson component of the model. Mother smoked is not statistically significant.

Parameter	Estimate	95% credible interval (lower, upper)
Intercept	1.34	(0.30, 2.34 )
Mother wheezed (prenatal)	-0.01	(-0.04, 0.01 )
$\sigma_v^2$	1.16	(0.76, 1.75 )

Table J.2: Univariate area-level ZIP model with Mother wheezed as a covariate for both the logistic and Poisson component of the model. Mother wheezed is not statistically significant.

Parameter	Estimate	95% credible interval (lower, upper)
Intercept	0.79	(-0.38, 1.73 )
Mother asthma (prenatal)	0.00	(-0.02, 0.03 )
$\sigma_v^2$	1.21	(0.79, 1.81 )

Table J.3: Univariate area-level ZIP model with Mother asthma as a covariate for both the logistic and Poisson component of the model. Mother asthma is not statistically significant.

---

Parameter	Estimate	95% credible interval (lower, upper)
Intercept	1.04	(0.29, 1.66 )
Father smoked	-0.02	(-0.09, 0.06 )
$\sigma_v^2$	1.19	(0.77, 1.76)

Table J.4: Univariate area-level ZIP model with Father smoked as a covariate for both the logistic and Poisson component of the model. Father smoked is not statistically significant.

Parameter	Estimate	95% credible interval (lower, upper)
Intercept	1.14	(-0.08, 2.24)
Father wheezed	-0.01	(-0.04, 0.02)
$\sigma_v^2$	1.21	(0.79, 1.80)

Table J.5: Univariate area-level ZIP model with Father wheezed as a covariate for both the logistic and Poisson component of the model. Father wheezed is not statistically significant.

Parameter	Estimate	95% credible interval (lower, upper)
Intercept	1.47	(0.61, 2.24 )
Father asthma	-0.03	(-0.07, 0.01 )
$\sigma_v^2$	1.13	(0.75, 1.68 )

Table J.6: Univariate area-level ZIP model with Father asthma as a covariate for both the logistic and Poisson component of the model. Father asthma is statistically significant.

Parameter	Estimate	95% credible interval (lower, upper)
Intercept	1.43	(0.60, 2.13)
Highway	-0.02	(-0.04, 0.01)
$\sigma_v^2$	1.10	(0.71, 1.65)

Table J.7: Univariate area-level ZIP model with Highway as a covariate for both the logistic and Poisson component of the model. Highway is not statistically significant.

---

Parameter	Estimate	95% credible interval (lower, upper)
Intercept	0.90	(0.06, 1.65 )
Water	0.00	(-0.05, 0.06 )
$\sigma_v^2$	1.21	(0.79, 1.82 )

Table J.8: Univariate area-level ZIP model with Water as a covariate for both the logistic and Poisson component of the model. Water is not statistically significant.

Parameter	Estimate	95% credible interval (lower, upper)
Intercept	0.98	(0.26, 1.58)
Factory	-0.01	(-0.12, 0.09 )
$\sigma_v^2$	1.20	(0.79, 1.79 )

Table J.9: Univariate area-level ZIP model with Factory as a covariate for both the logistic and Poisson component of the model. Factory is not statistically significant.

Parameter	Estimate	95% credible interval (lower, upper)
Intercept	1.16	(0.53, 1.70)
Farm	-0.03	(-0.06, 0.01)
$\sigma_v^2$	1.12	(0.74, 1.67 )

Table J.10: Univariate area-level ZIP model with Farm as a covariate for both the logistic and Poisson component of the model. Farm is statistically significant.

Parameter	Estimate	95% credible interval (lower, upper)
Intercept	1.28	(0.50, 1.97)
Construction	-0.02	(-0.06, 0.01)
$\sigma_v^2$	1.14	(0.74, 1.70 )

Table J.11: Univariate area-level ZIP model with Construction as a covariate for both the logistic and Poisson component of the model. Construction is not statistically significant.

---

Parameter	Estimate	95% credible interval (lower, upper)
Intercept	1.06	(0.51, 1.53 )
SES	1.08	(0.11, 2.05 )
$\sigma_v^2$	1.03	(0.65, 1.58 )

Table J.12: Univariate area-level ZIP model with SES as a covariate for both the logistic and Poisson component of the model. SES is statistically significant.

Parameter	Estimate	95% credible interval (lower, upper)
Intercept	0.86	(0.28, 1.36 )
SES (average before)	0.59	(0.03, 1.19 )
$\sigma_v^2$	1.05	(0.66, 1.61 )

Table J.13: Univariate area-level ZIP model with SES (average before) as a covariate for both the logistic and Poisson component of the model. In this model the average of the individual covariates that are used to make the SES index (Father/Mother Education and household income) is taken by area and then an area level SES index is calculated SES (average before) is statistically significant.

Parameter	Estimate	95% credible interval (lower, upper)
Intercept	-3.14	(-7.51, 2.94 )
Mother education	0.26	(-0.14, 0.54)
$\sigma_v^2$	1.07	(0.68, 1.64 )

Table J.14: Univariate area-level ZIP model with Mother education as a covariate for both the logistic and Poisson component of the model. In this model one of the covariates that are used to make the SES index is aggregated to the area-level by averaging within each area. Mother education is not statistically significant.

Parameter	Estimate	95% credible interval (lower, upper)
Intercept	-4.02	(-9.09, -0.03 )
Father education	0.33	(0.07, 0.66 )
$\sigma_v^2$	1.07	(0.69, 1.62 )

Table J.15: Univariate area-level ZIP model with Father education as a covariate for both the logistic and Poisson component of the model. In this model one of the covariates that are used to make the SES index is aggregated to the area level by averaging within each area. Father education is statistically significant.

---

Parameter	Estimate	95% credible interval (lower, upper)
Intercept	-2.74	(-6.89, 0.19 )
Income	0.54	(0.12, 1.12)
$\sigma_v^2$	1.05	(0.69, 1.58 )

Table J.16: Univariate area-level ZIP model with Income as a covariate for both the logistic and Poisson component of the model. In this model one of the covariates that are used to make the SES index is aggregated to the area level by averaging within each area. Father education is statistically significant.

Parameter	Estimate	95% credible interval (lower, upper)
Intercept	0.96	(0.38, 1.48 )
$\sigma_v^2$	1.16	(0.77, 1.72)

Table J.17: Area-level ZIP model with no covariate on both the logistic and Poisson component of the model.

## Appendix K: Model Building for Area ZIP- Co- variates for Both Components

Parameter	Estimate	95% credible interval(lower, upper)
Intercept	1.04	(0.42, 1.59 )
SES (average before)	0.50	(-0.10, 1.08)
Farm	-0.02	(-0.05, 0.01 )
$\sigma_v^2$	1.04	(0.66, 1.59 )
DIC	168.6	-

Table K.1: Area-level ZIP model with SES average before and Farm as a covariates for both the logistic and Poisson component of the model.

Parameter	Estimate	95% credible interval (lower, upper)
Intercept	0.52	(-0.47, 1.44 )
SES (average before)	0.64	(0.04, 1.23)
Farm	-0.02	(-0.05, 0.01 )
Mother asthma (prenatal)	0.02	(-0.01, 0.04 )
$\sigma_v^2$	1.03	(0.65, 1.59 )
DIC	169.6	-

Table K.2: Area-level ZIP model with three covariates, for both the logistic and Poisson component of the model.

---

Parameter	Estimate	95% credible interval (lower, upper)
Intercept	0.74	(-0.52, 1.92 )
SES (average before)	0.54	(-0.16, 1.23)
Farm	-0.02	(-0.06, 0.01 )
Mother asthma (prenatal)	0.01	(-0.02, 0.04 )
Mother smoked (prenatal)	-0.01	(-0.05, 0.03 )
$\sigma_v^2$	1.05	(0.66, 1.63 )
DIC	183.3	-

Table K.3: Area-level ZIP model with four covariates, added prenatal covariate Mother smoked to the previous model for both the logistic and Poisson component of the model.

Parameter	Estimate	95% credible interval (lower, upper)
Intercept	0.95	(-0.02, 1.90 )
Farm	-0.03	(-0.6, 0.01)
Mother asthma (prenatal)	0.01	(-0.02, 0.03 )
$\sigma_v^2$	1.17	(0.76, 1.75 )
DIC	164.7	-

Table K.4: Area-level ZIP model with two covariates Farm and Mother Asthma for both the logistic and Poisson component of the model.

Parameter	Estimate	95% credible interval (lower, upper)
Intercept	1.27	(0.18, 2.25)
Farm	-0.03	(-0.06, 0.00 )
Mother asthma (prenatal)	0.00	(-0.02, 0.03 )
Mother smoked (prenatal)	-0.02	(-0.06, 0.01 )
$\sigma_v^2$	1.12	(0.72, 1.71 )
DIC	152.8	-

Table K.5: Area-level ZIP model with three covariates: Farm, Mother Asthma and smoked for both the logistic and Poisson component of the model.

---

Parameter	Estimate	95% credible interval (lower, upper)
Intercept	1.34	( 0.69, 1.93 )
Farm	-0.03	(-0.06, 0.00)
Mother smoked (prenatal)	-0.03	(-0.06, 0.01)
$\sigma_v^2$	1.08	(0.70, 1.64 )
DIC	175.9	-

Table K.6: Area-level ZIP model with two covariates Farm and Mother smoked for both the logistic and Poisson component of the model.