The Capstone Committee for Ioannis Malagaris Certifies that this is the approved version of the following capstone:

Assessing the Effect of "Time of Birth" on Nasopharyngeal Microbial Load in Infants

Committee:

Kristofer Jennings, PhD, Supervisor

Heidi Spratt, PhD

Janak A. Patel, MD

Assessing the effect of “time of birth” on nasopharyngeal microbial load in infants

by

Ioannis Malagaris, MSc

Capstone

Presented to the Faculty of the Graduate School of

The University of Texas Medical Branch

in Partial Fulfillment

of the Requirements

for the Degree of

Master of Public Health

The University of Texas Medical Branch

July 2019

Dedication

For Fifaki and my beautiful caring wife, Effie.

Acknowledgements

I would like to express my sincere gratitude to the people who helped me to complete this project. Foremost, I would like to express my gratitude to my supervisor and committee chair, Dr. Kristofer Jennings, for the endless hours he invested mentoring me and reviewing the material I produced working on this project.

I would also like to thank my Capstone committee members, Dr. Heidi Spratt and Dr. Janak A. Patel, for improving this thesis with their insightful comments.

Furthermore, I extend my appreciation to Dr. Cara Pennel, Director of Public Health Graduate Program and Dr. Kristen M. Peek Director of Population Health Sciences Program, for their support and the solutions they provided to any problem presented during my studies.

Finally, I feel obliged to make special mention of my former PhD mentor, Dr. David N. Herndon who encouraged me to pursuit a degree in biostatistics and provided the required funding for my studies.

The data used herein were collected through an NICHD grant funded project: R01 DC005841-06A1 “Pathogenesis of virus-induced acute otitis media”, Principal Investigator, Tasnee Chonmaitree, MD.

Assessing the effect of “time of birth” on nasopharyngeal microbial load in infants

Publication No._____

Ioannis Malagaris, MPH

The University of Texas Medical Branch, 2019

Supervisor: Kristofer Jennings

Abstract

Acute otitis media (AOM) is one of the most common childhood infections and its pathogenesis involves complex interactions between bacteria and viruses. Bacteria and viruses contributing to the AOM are collectively known as otopathogens. The objective of the capstone is to assess the effect of “month of birth” (MOB) on the microbial load of the most abundant otopathogen, Moraxella Catarrhalis. This is a retrospective analysis of data collected between 2009 – 2014 as part of a longitudinal study to determine risk factors for AOM. Subjects were recruited near birth and followed up to 1 year of age. For measurement of nasopharyngeal microbial abundance, approximately seven specimens were taken per subject. The total number of patients and specimens in the dataset are

139 and 948 respectively. The outcome variable was the log-transformed relative abundance of Moraxella genera. Its relationship with MOB was modeled using generalized additive mixed effects models (GAMM) controlling for age, month of specimen collection and other covariates while blocking on subject to control for repeated measures. Model selection was based on Bayesian Information Criterion (BIC). MOB showed a statistically significant non-linear relationship with Moraxella microbial abundance ($p < 0.001$). Increasing age and birth order were positively associated with the outcome ($p < 0.001$ and $p = 0.03$ respectively). The effect of MOB displayed a cyclic seasonal nature. This finding suggests that the timing of birth affects the average Moraxella microbial abundance in the first year of life. Our data demonstrate that MOB can be used to identify high risk populations for AOM. Further investigation on the underlying mechanisms mediating this complex relationship may aid in broadening the clinical understanding of AOM disease process.

TABLE OF CONTENTS

Contents

List of Tables

List of Figures

List of Abbreviations

ACF	Autoregressive Correlation Function
AIC	Akaike Information Criterion
AOM	Acute Otitis Media
BIC	Bayesian Information Criterion
DNA	Deoxyribonucleic Acid
GAMM	Generalized Additive Mixed Effects Models
mgcv	Mixed GAM Computation Vehicle
MOB	Month of Birth
nlme	Linear and Nonlinear Mixed Effects Models
NP	Nasopharyngeal
PCR	Polymerase Chain Reaction
RNA	Ribonucleic Acid
URI	Upper Respiratory tract Infection

Chapter 1 Introduction

Acute otitis media (AOM) is one of the most common childhood infections, the leading cause of doctor's visits by children and the most common cause of surgery and antibiotics consumption of this population in most countries(Kathleen A. Daly et al., 2010). In the first year of life, 62% of infants contract AOM, while 70% of two-year old children have experienced the disease at least once(Schilder, Lok, & Rovers, 2004). Better understanding of associated risk factors for AOM is required in the effort of reducing public health burden.

AOM is a polymicrobial disease and its pathogenesis involves complex interactions between bacteria and viruses(Chonmaitree, Jennings, et al., 2017). The prevalence of otopathogen genera (*Haemophilus*, *Streptococcus*, and *Moraxella*) is positively associated with frequencies of upper respiratory tract infections (URI), and the otopahogens' presence seems to be a necessary condition for AOM during viral infection(Chonmaitree, Jennings, et al., 2017).

Several disease-dependent mechanisms have been proposed relating seasonality or month of birth with risk of disease(Boland, Shahn, Madigan, Hripcsak, & Tatonetti, 2015). Recent studies have linked timing of birth with several disorders(Boland et al., 2015) and overall lifespan(Doblhammer & Vaupel, 2001). It has been shown that otopathogens' microbiome load is not constant but varies throughout the year(Teo et al., 2015). This seasonal pattern may be attributed to several factors that also vary with season (temperature, humidity, outdoor exposure). Moreover, the type and the likelihood of exposure to certain respiratory

viruses throughout the year may partly determine otopathogens' microbiome load. Time of birth determines the environment that a newborn is exposed to during the first months of life and likely affects otopathogen colonization.

The aim of this study is to assess the effect of "month of birth" (MOB) on the microbiome load during the first year of age. This will be a retrospective analysis of data collected between 2009 – 2014 as part of a longitudinal study to assess the prevalence and risk factors for upper respiratory tract infection (URI) and AOM(Chonmaitree, Trujillo, Jennings, & Alvarez-fernandez, 2017). The nasopharyngeal specimens were collected from each subject in monthly (30 days) intervals for the first 6 months of life and at least once during the second half, or approximately at 9 months of age. In total, 948 nasopharyngeal specimens were collected from 139 study subjects (~ 7 specimens per subject). Given the nature of the available data, mixture outcomes with repeated observations per subject, this analysis has several challenges.

The effect of MOB on microbiome load cannot be assessed validly without considering several potential longitudinal confounders: in particular, age and calendar month the sample was controlled for. During the first year of life, factors such as the state of the immune system of the newborn and the degree of its exposure to external environment vary as function of age; therefore, age is probably associated with microbiome load. Furthermore, it is expected that seasonal, yearly cyclic trends of genera to be evident. It follows that the calendar month that a sample was taken should also be associated with microbiome load.

An important aspect of this discussion has to do with the fact that MOB is a function of age and time of the sample. For example, a sample taken in July, from a 6-month-old patient completely determines the MOB of the patient, namely, January. As a result, the true effect of MOB cannot be determined while ignoring age and calendar month of sample measurement and due to linear dependence, inclusion of all three variables in multiple regression analysis is not possible (due to nearly complete collinearity). However, in our dataset, due to repeated measurements per subject during the first year of life, a significant proportion of the sample space of possible combinations of the three variables is recorded, and proper statistical analysis can be applied to test our hypothesis.

The statistical analysis in this thesis incorporates random effects to account for the repeated measurements, adjust for any trends of microbiome load during study period (2009 – 2014), test for within subject correlation of residuals, and accounts for the appropriate correlation structure if necessary.

Chapter 2 Methods

DATA DESCRIPTION

This is a retrospective observational study and the aim of this section is to describe the nature of dataset used for the analysis. The data were collected between 2009 – 2014 as part of a longitudinal study to assess the prevalence and risk factors for URI and AOM(Chonmaitree, Trujillo, et al., 2017). Detailed documentation of the dataset is presented in the original clinical study of this cohort (Chonmaitree, Trujillo, et al., 2017). Subjects were infants recruited near birth and followed up to first diagnosis of AOM or up to 1 year of age. For measurement of nasopharyngeal microbial abundance, an average of seven specimens were taken per subject. Study personnel made home visits to collect Nasopharyngeal (NP) specimen swabs. The total number of patients and specimens in the dataset are 139 and 971 respectively(Chonmaitree, Jennings, et al., 2017). Descriptive statistics of subjects and the distribution of sample collection timing is presented on **Table** 1 (Chonmaitree, Jennings, et al., 2017).

Table 1: Characteristics of subjects and specimens[1].

	Total	Subjects with AOM	Subjects without AOM
Number of subjects	139 (100)	65 (47)[a]	74 (53)
Male	83 (60)[b]	36 (55)	47 (64)
Female	56 (41)	29 (45)	27 (36)
Race			
- White	119 (86)	56 (86)	63 (85)
- African American	18 (13)	9 (14)	9 (12)
- Asian	2 (1)	0	2 (3)
Ethnicity			
- Hispanic/ Latino	77 (56)	35 (54)	42 (57)
- NonHispanic/ Latino	62 (44)	30 (46)	32 (43)
Breastfeeding			
- Exclusive breastfeeding for 6 months	13 (9)[b]	5 (8)	8 (11)
- Exclusive breastfeeding for 3 months	7 (5)	5 (8)	2 (3)
- Exclusive formula feeding	62 (45)	32 (49)	30 (41)
- Mixed feeding	57 (41)	23 (35)	34 (46)
One or more siblings at home (% yes)	55	57	54
Daycare attendance (% yes)	33	22	43
Cigarette smoke exposure (% yes)	23	18	27
Number of Specimens	971[c]	432	539
Average number of specimens per subject	7	6.6	7.3
Age at samples collection[d]			
- 1 month	131	57	74
- 2 months	136	64	72
- 3 months	148	71	77
- 4 months	139	65	74
- 5 months	145	73	72
- 6 months	137	63	74
- 7–12 months	135	39	96
Specimens collected during URI[e]	223	119	104
Specimens collected during AOM	45	45	0
Specimens collected after antibiotic use			
- 7 days	37	30	7
- 14 days	43	36	7
- 1 month	71	57	14
- 2 months	98	72	26

a- row %

b-column % (% within the same category)

c- all specimens sequenced; 23 (2%) of these had <1000 sequence reads were excluded

d- nearest month, both monthly and URI samples

e- median day of URI at the time of sample collection = day 4, excluded samples collected during AOM diagnosis

The relative populations of the most abundant microbiota genera detected in the specimens are presented in

[1] Adapted from (Chonmaitree, Jennings, et al., 2017)

Phylum	Class	Order	Family	Genus	Samples from all subjects [a]
					(N = 948)
Actinobacteria	Actinobacteria	Corynebacteriales	Corynebacteriaceae	Corynebacterium	17.8%
Proteobacteria	Gammaproteobacteria	Pseudomonadales	Moraxellaceae	Moraxella	9.7%
Firmicutes	Bacilli	Lactobacillales	Carnobacteriaceae	Dolosigranulum	6.8%
Firmicutes	Bacilli	Bacillales	Staphylococcaceae	Staphylococcus	5.7%
Proteobacteria	Gammaproteobacteria	Pseudomonadales	Moraxellaceae	Acinetobacter	4.1%
Proteobacteria	Gammaproteobacteria	Pasteurellales	Pasteurellaceae	Haemophilus	3.7%
Proteobacteria	Gammaproteobacteria	Pseudomonadales	Pseudomonadaceae	Pseudomonas	3.5%
Firmicutes	Bacilli	Lactobacillales	Streptococcaceae	Streptococcus	3.5%
Actinobacteria	Actinobacteria	Bifidobacteriales	Bifidobacteriaceae	Bifidobacterium	2.9%
Proteobacteria	Gammaproteobacteria	Enterobacteriales	Enterobacteriaceae	Enterobacter	2.5%
Actinobacteria	Actinobacteria	Micrococcales	Micrococcaceae	Micrococcus	1.6%
Proteobacteria	Gammaproteobacteria	Chromatiales	Ectothiorhodospiraceae	Arhodomonas	1.5%
Bacteroidetes	Bacteroidia	Bacteroidales	Bacteroidaceae	Bacteroides	1.3%
Firmicutes	Clostridia	Clostridiales	Ruminococcaceae	Incertae Sedis	1.0%
Proteobacteria	Betaproteobacteria	Burkholderiales	Burkholderiaceae	Ralstonia	0.9%
Bacteroidetes	Flavobacteriia	Flavobacteriales	Flavobacteriaceae	Myroides	0.8%
Proteobacteria	Gammaproteobacteria	Enterobacteriales	Enterobacteriaceae	Pantoea	0.6%
Proteobacteria	Gammaproteobacteria	Enterobacteriales	Enterobacteriaceae	Yersinia	0.6%
Firmicutes	Clostridia	Clostridiales	Clostridiaceae 1	Clostridium sensu stricto 1	0.6%
Proteobacteria	Alphaproteobacteria	Sphingomonadales	Sphingomonadaceae	Sphingomonas	0.6%
Proteobacteria	Alphaproteobacteria	Sphingomonadales	Sphingomonadaceae	Sphingobium	0.6%

Error! Reference source not found. Pathogens such as bacteria and viruses contributing to the AOM are collectively known as otopathogens. The response variable, *Moraxella Catarrhalis,* is most abundant species among otopathogens with relative abundance of 9.7%.

Table 2: Relative abundance of microbiota in specimens[2]

Phylum	Class	Order	Family	Genus	Samples from all subjects [a]
					(N = 948)
Actinobacteria	Actinobacteria	Corynebacteriales	Corynebacteriaceae	Corynebacterium	17.8%
Proteobacteria	Gammaproteobacteria	Pseudomonadales	Moraxellaceae	Moraxella	9.7%
Firmicutes	Bacilli	Lactobacillales	Carnobacteriaceae	Dolosigranulum	6.8%
Firmicutes	Bacilli	Bacillales	Staphylococcaceae	Staphylococcus	5.7%
Proteobacteria	Gammaproteobacteria	Pseudomonadales	Moraxellaceae	Acinetobacter	4.1%
Proteobacteria	Gammaproteobacteria	Pasteurellales	Pasteurellaceae	Haemophilus	3.7%
Proteobacteria	Gammaproteobacteria	Pseudomonadales	Pseudomonadaceae	Pseudomonas	3.5%
Firmicutes	Bacilli	Lactobacillales	Streptococcaceae	Streptococcus	3.5%
Actinobacteria	Actinobacteria	Bifidobacteriales	Bifidobacteriaceae	Bifidobacterium	2.9%
Proteobacteria	Gammaproteobacteria	Enterobacteriales	Enterobacteriaceae	Enterobacter	2.5%
Actinobacteria	Actinobacteria	Micrococcales	Micrococcaceae	Micrococcus	1.6%
Proteobacteria	Gammaproteobacteria	Chromatiales	Ectothiorhodospiraceae	Arhodomonas	1.5%
Bacteroidetes	Bacteroidia	Bacteroidales	Bacteroidaceae	Bacteroides	1.3%
Firmicutes	Clostridia	Clostridiales	Ruminococcaceae	Incertae Sedis	1.0%
Proteobacteria	Betaproteobacteria	Burkholderiales	Burkholderiaceae	Ralstonia	0.9%
Bacteroidetes	Flavobacteriia	Flavobacteriales	Flavobacteriaceae	Myroides	0.8%
Proteobacteria	Gammaproteobacteria	Enterobacteriales	Enterobacteriaceae	Pantoea	0.6%
Proteobacteria	Gammaproteobacteria	Enterobacteriales	Enterobacteriaceae	Yersinia	0.6%
Firmicutes	Clostridia	Clostridiales	Clostridiaceae 1	Clostridium sensu stricto 1	0.6%
Proteobacteria	Alphaproteobacteria	Sphingomonadales	Sphingomonadaceae	Sphingomonas	0.6%
Proteobacteria	Alphaproteobacteria	Sphingomonadales	Sphingomonadaceae	Sphingobium	0.6%

SPECIMEN ANALYSIS

DNA extraction/ amplification and sequencing, and sequence analysis are described in (Chonmaitree, Jennings, et al., 2017). Briefly, PowerMicrobiome DNA/RNA Isolation kit was used for DNA extraction from the Nasopharyngeal specimens. PCR was used for DNA amplification, and the MiSeq platform was used for DNA sequencing. Analysis of subsequences was performed in order to

2 Adapted from (Chonmaitree, Jennings, et al., 2017)

identify the type of bacteria in the sample, using CLC Genomics Workbench 8.0.1 Microbial Genomics Module.

STATISTICAL ANALYSIS PLAN

The final sample include in the current analysis is a subset of the dataset described in the previous section. To increase signal-to-noise ratio only nasopharyngeal specimens having more than 1000 bacterial counts were included in the analysis. Characteristics of the final sample are presented in Tables 3 and 4 in the Results Chapter.

The dependent variable in this analysis was the log-transformed relative abundance of Moraxella genus. To this end, the relative abundance of Moraxella genus in each sample was computed by dividing Moraxella counts by total counts per sample. To avoid undetermined values (like LOG_e0) the relative abundance was computed by adding 1 to the absolute Moraxella counts before division by the total sample counts. The same procedure was followed in the original clinical study of this cohort (Chonmaitree, Jennings, et al., 2017) The outcome was also modeled with logistic regression and the results were similar. However, binomial models occasionally suffered from instability (model convergence failure) and produced unreliable results. These issues were resolved by modeling the log-transformed relative abundance of Moraxella genus. All statistical analyses were performed using R(R Core Team, 2018).

Due to repeated measurement within subject, the effect of MOB on microbiome load was assessed using linear mixed effects models and generalized additive mixed models (to assess potential non-linear relationships). Models will include random intercept for subject and the possibility of inclusion of random effects for slope were explored. Linear mixed effects models were examined first using the “nlme” package(Pinheiro, Bates, DebRoy, Sarkar, & R Core Team, 2018) in R. In the first model, the only fixed effect is the MOB. The two main suspected confounding variables age and sample-calendar-month were added sequentially, and the changes in the estimated effect of MOB and its precision were assessed. The results are reported on **Table** 6. The original clinical study of this cohort (Chonmaitree, Trujillo, et al., 2017) reports the effect of several environmental factors on upper respiratory tract infections (URI) in infants up to 6 months old. The identified factors were number of siblings in the home, exclusive breastfeeding for at least six months, length of breastfeeding, day care attendance and being born after February 2010, To obtain greater precision of the effect of MOB on the relative abundance of Moraxella genus several covariates were added to obtain the full model. The additional covariates were birth weight, gestational age, type of feeding, birth order, length of exclusive breast feeding, time to formula feeding and cigarette smoke exposure. Calendar year of sample measurement were also included in the analysis to account for any differences in microbiome load due to year (study period 2009 – 2014). The results are shown on **Table 8**. Stepwise model selection using Akaike information criterion (AIC) was used to obtain the final linear model. The results are reported on table. AIC was preferred over more

strict criteria for two reasons. First, the objective at this stage of the analysis was not to build a predictive model but to obtain higher precision of MOB effect. Second, the effects of the variables obtained from the linear model was also assessed in the next steps of the analysis where non-linear relationships were examined.

Evidence of temporal serial correlation was assessed empirically by calculating the within subject residual correlation using the Autoregressive Correlation (ACF) function. Improvement in model fit by inclusion of correlation structure was assessed with the Bayesian information criterion (BIC).

To assess whether a non-linear relationship exists between the outcome and MOB, generalized additive mixed models (GAMM) were employed using the "mgcv" package(Wood, 2019) in R. Smoothing spline functions were applied on the main predictor, MOB, the two main covariates, age and sample-calendar-month and the transformed variables were used in the analysis. To allow for greater resolution, before the transformation the three variables were expressed in day units with range from 1 to 365. The time origin for MOB and sample-calendar month will be January 1st and date of birth for age. The results are reported on **Table 9**. The non-linear association between MOB and the outcome was further investigated by attempting to decompose the aggregate relationship into distinct elements of seasonal nature which may be amenable to clinical interpretation. To this end, several sinusoidal functions- varying in period -having as phase argument the variable MOB were fitted to the outcome, while including in the model all the

statistically significant covariates identified in the previous model. The sinusoidal functions were of the form:

$$\mathrm{a} \cdot \sin\left(2\pi f \cdot \frac{t}{365}\right) + b \cdot \cos(2\pi f \cdot \frac{t}{365})$$

where “f” is the frequency of the seasonal component (1 over period), “t” is the time of birth in days and “a” and “b” are the coefficients determined by the model. The results are reported on **Table 10**.

Chapter 3 Results

DESCRIPTIVE STATISTICS

Data were collected from a total of 139 subjects. Subjects' characteristics are shown on **Table 3**. Subjects were predominantly male (60%) and white (86%) and of Hispanic origin (55%). During the study period, 77% of subject lived in a smoke-free environment. Formula exclusive feeding was more prevalent (45%) than exclusive breast feeding (14%) or mixed feeding (41%).

Table 3: Subjects' characteristics

Number of Patients	139
Male: Female	83:56
Race	
White	119
African American	18
Asian	2
Hispanic: Non-Hispanic	77:62
Breast Feeding	
Exclusive for 6 months	19
Exclusive for 3 months	7
Exclusive Formula	62
Mixed feeding	57
Cigarette smoke exposure (% yes)	23
Birth Weight (kg)	3.3 (0.5)
Birth Order (relative to siblings)	
1	62
2	41
3	21
4	10
5	4
7	1
Gestational Age (months)	
36	6
37	22
38	34
39	38
40	31
41	8

Out of 139 subjects, 62 (45%) had no siblings at the study period while 15 (10%) had a within family birth order of four or more. Weight at birth was on average 3.3±0.5 kg and followed approximately a bell-shaped distribution (**Figure** 1).

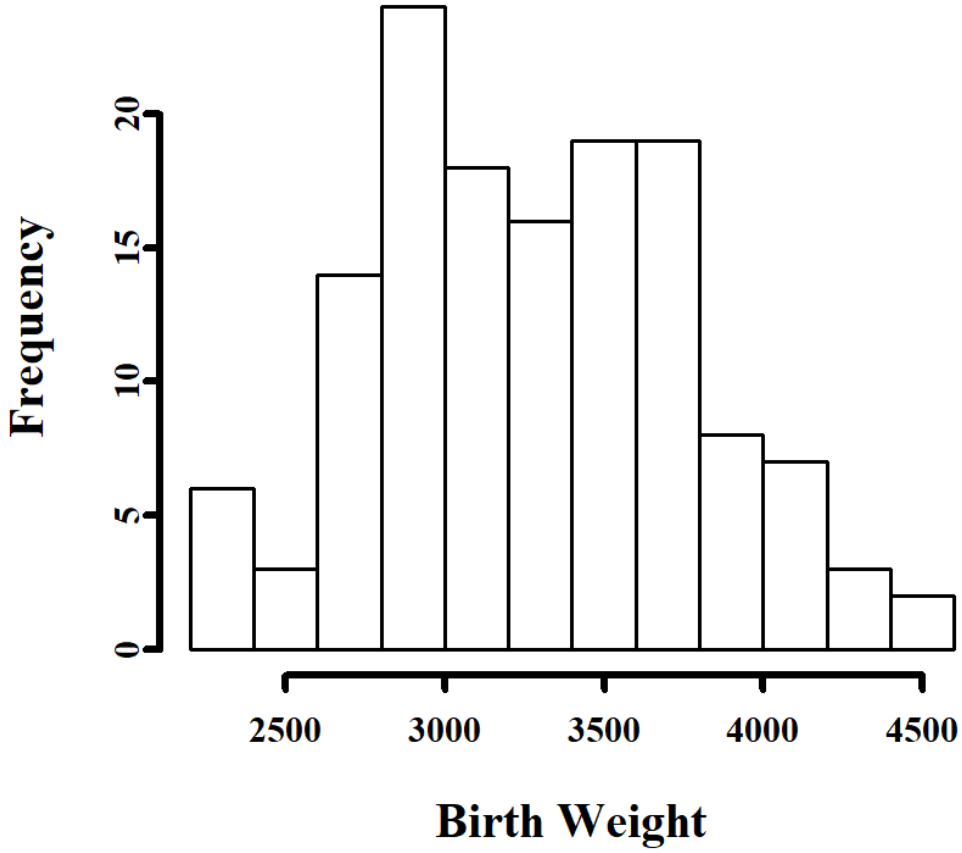

Figure 1: Subjects' weight distribution in grams at birth (N = 139)

The gestational age also followed bell-like distribution with central tendency at the 39th week (**Figure** 2).

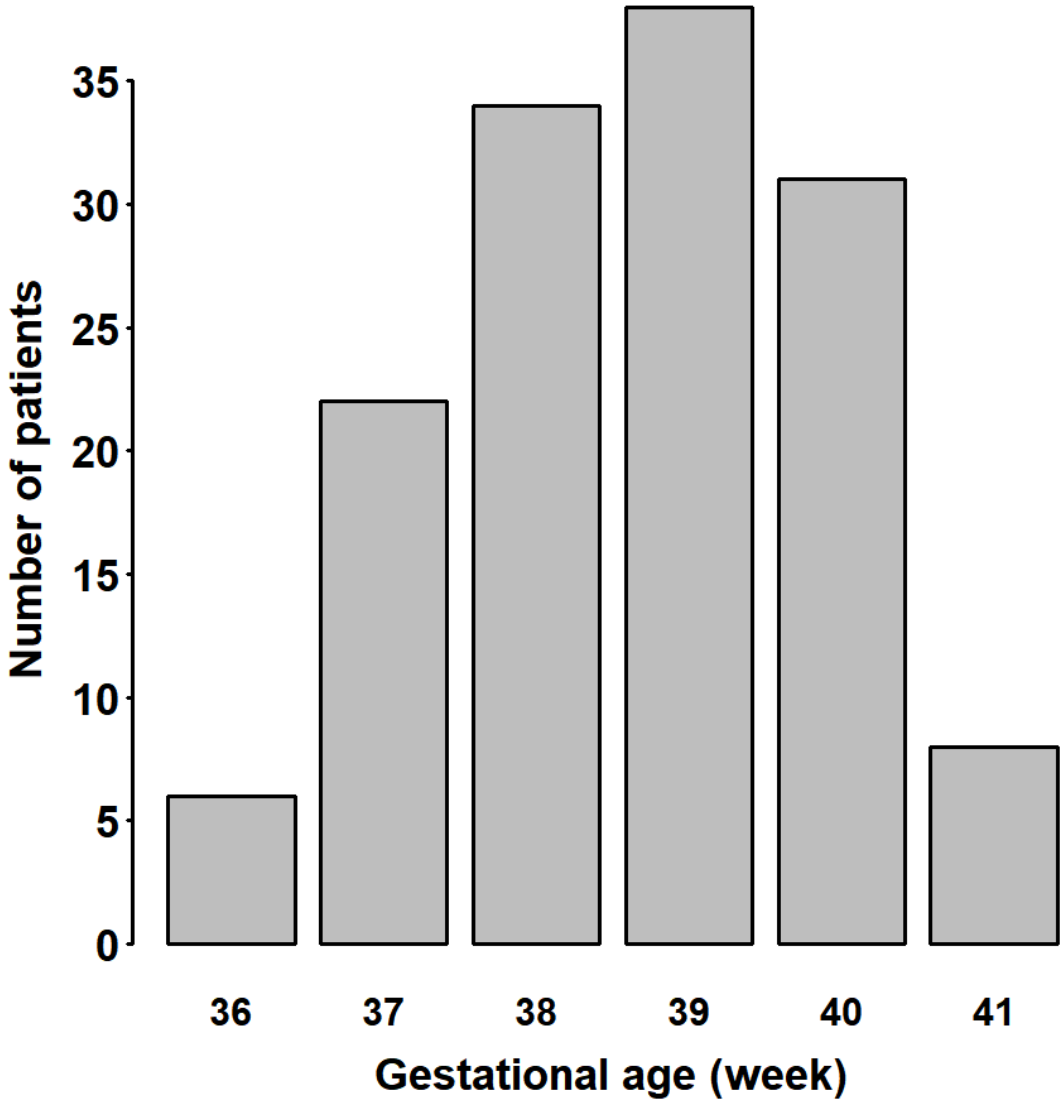

Figure 2: Subjects' gestational age distribution

N = 139

Out of 139 subjects, a total of 971 specimens were collected between August 2008 and January 2014. Specimens which yielded less than 1000 readings in sequencing analysis were excluded because of sample quality concerns. The final number of specimens included in this analysis was 947. Specimen collection characteristics are displayed on **Table** 4

Table 4: Specimen collection characteristics

Number of specimens	947
Date range of collection	2009 - 2014
Age (months) at collection	
1	126
2	133
3	144
4	137
5	141
6	137
7 - 12	129
Calendar Year of collection	
2009	49
2010	122
2011	36
2012	450
2013	288
2014	2
Season of collection (quarterly)	
1st	291
2nd	243
3nd	188
4th	225
Within patient number of specimens	
4	7
5	11
6	35
7	51
8	22
9	12
13	1

The bulk of specimens (86%) were collected from subjects aged 6 months or less. The first 6 months of age were almost uniform equally sampled (**Figure** 3). The second half of the first life of year is represented by 129 specimens (14%) of which the majority was obtained from 9 months old subjects (n = 87).

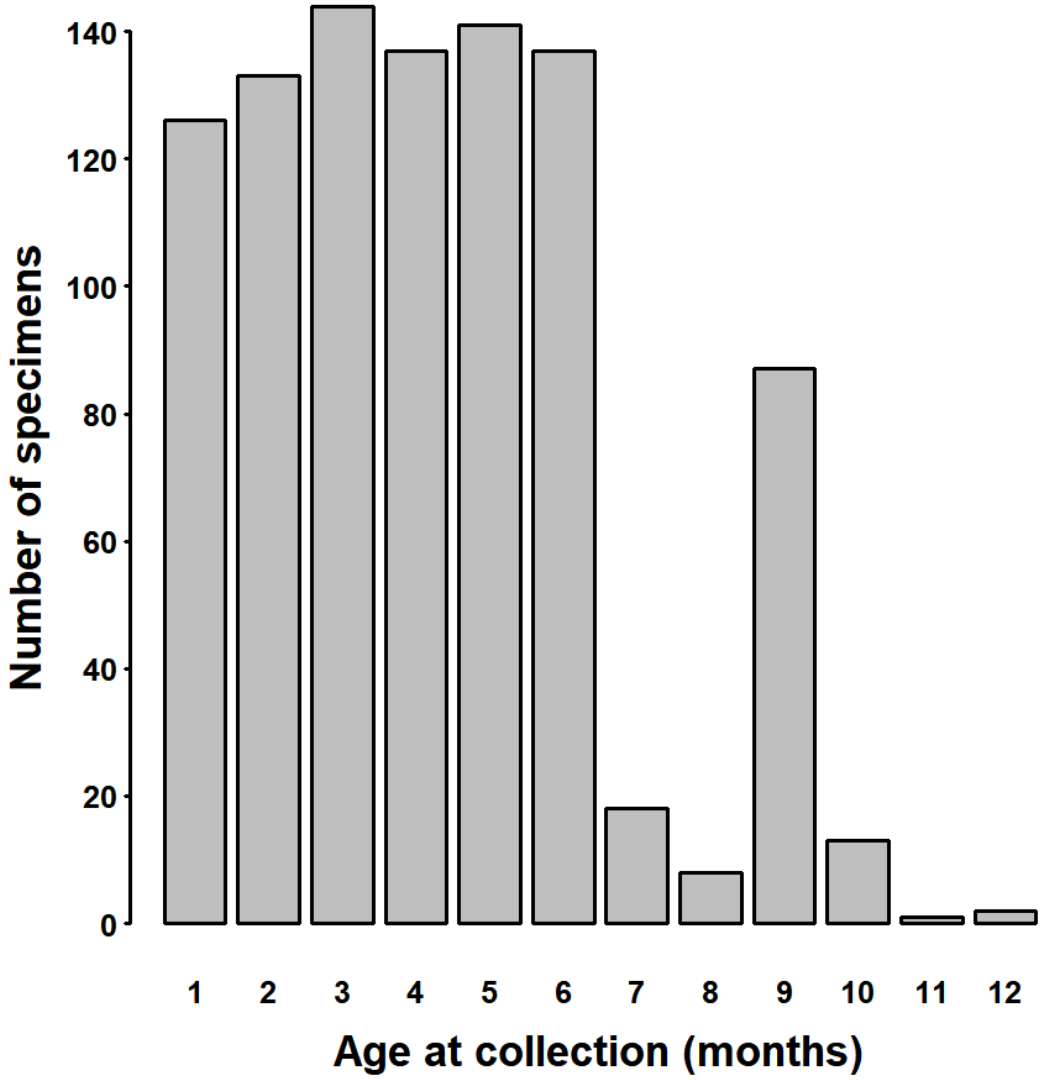

Figure 3: Number of specimens collected per age group
N = 947

The distribution of specimen sampling among seasons is displayed on **Table 4** (quarters of a year), and **Figure 4** (months). All months were almost uniformly sampled, therefore there is no risk of bias due to uneven sampling.

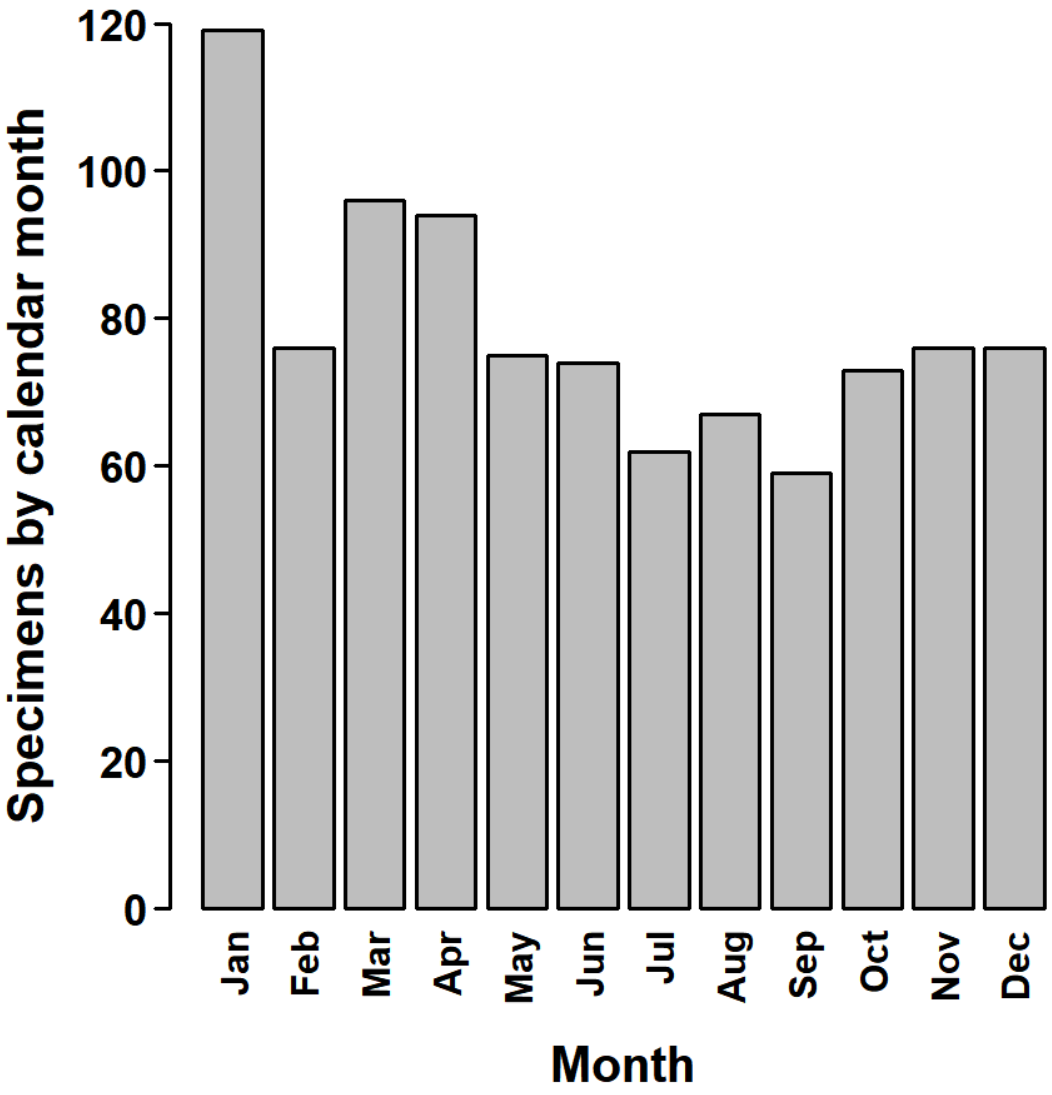

Figure 4: Total number of specimens collected per calendar month
N = 947

The specimens were collected between August 2009 and January 2014. The majority of collection took place at calendar years 2012 and 2013 (**Figure** 5).

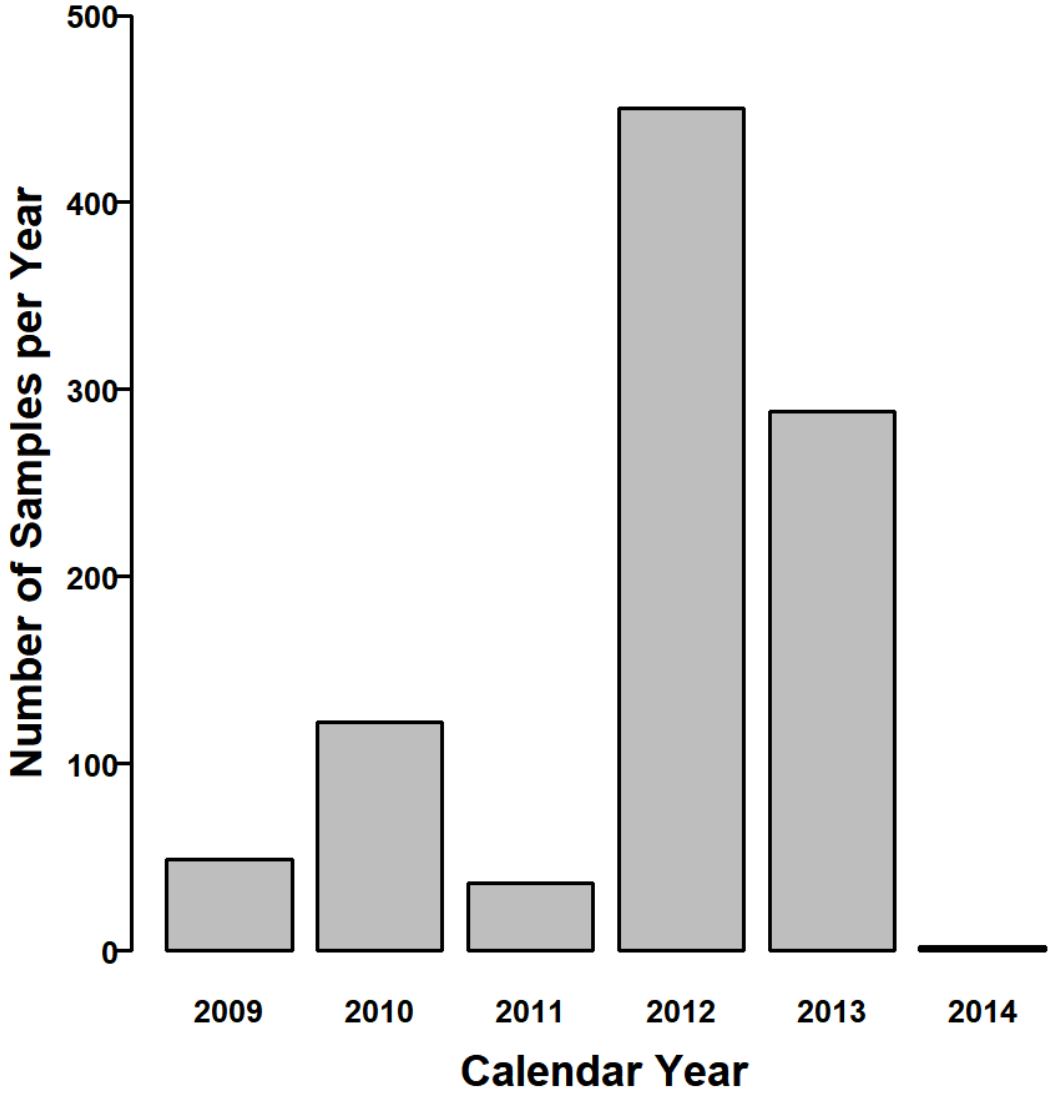

Figure 5: Specimen collection per calendar year
N = 947

According to study design(Chonmaitree, Jennings, et al., 2017), specimens were to be collected from each subject in monthly (30 days) intervals for the first 6 months of life and ideally at least once during the second half, or approximately at 9 months of age. The median number of specimens per subject was 7 (n = 51), with most subjects contributing 6 to 8 specimens. The groups of patients -based on number of specimens per subject- and their frequency are shown on **Figure** 6.

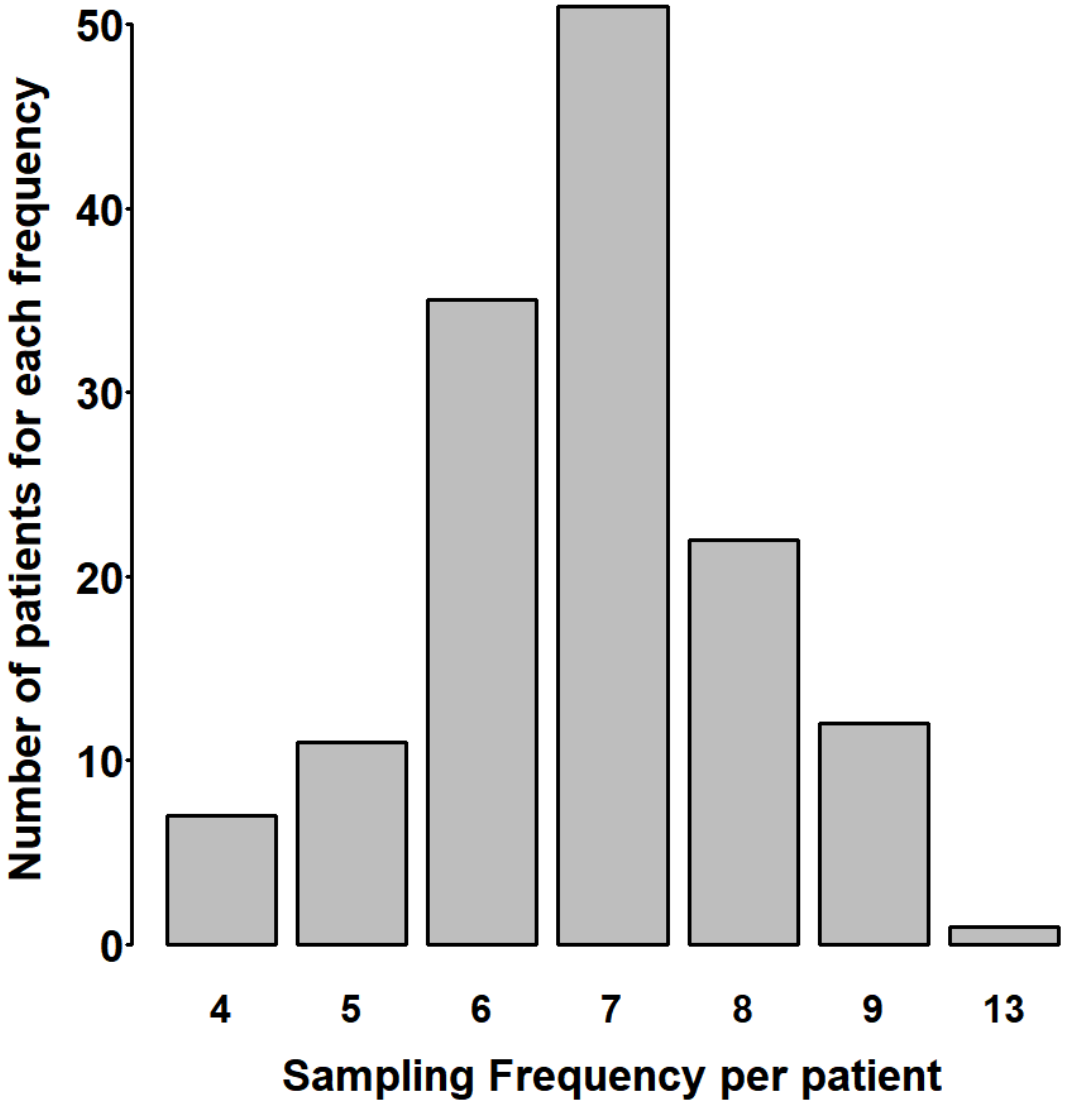

Figure 6: Specimen collection characteristics

N= 139

The main variable of interest of this analysis is the MOB and its sufficient representation in the sample is important. Moreover, to discern the effect of MOB given the potentially confounding effect of 1) subject's age and 2) calendar month at specimen collection, a significant proportion of the sample space of possible combinations of the three variables should be represented in the dataset. Given the sample size of 139, each MOB would be ideally represented by 11-12 subjects.

The distribution of MOB in the dataset is shown on Error! Reference source not found..

Table 5: Number of births by calendar month from 2009 to 2014

Month	Count
Jan	8
Feb	12
Mar	13
Apr	12
May	2
Jun	7
Jul	10
Aug	15
Sep	19
Oct	11
Nov	19
Dec	11

N = 139

Except for May, all MOB are sufficiently represented in the data. The two subjects born in May contributed together 14 specimens covering an appreciated proportion of the sample space. Finally, instead of MOB, the day of the year at birth (1 to 365) was used in the analysis, therefore the uneven sampling is not expected to meaningfully affect the results.

DATA ANALYSIS

The effect of MOB on the microbial abundance of Moraxella genus (logarithm of relative abundance), was initially modeled using mixed effects linear regression while controlling on subset level due to repeated observations within subjects. In the first model, the only fixed effect is the MOB. The two main suspected confounding variables age and sample-calendar-month were added sequentially, and the changes in the estimated effect of MOB and its precision were assessed. According to the initial models, there is a statistically significant relationship of MOB with the outcome and the result remained significant after accounting for the potential confounding effect of the variables age and the calendar month of specimen collection. The results are summarized on **Table 6**.

Table 6: Assessing potential confounding

Terms	Coefficient	SE	t-value	p-value
MOB only	0.10	0.04	2.7	0.008
MOB + age	0.11	0.04	2.7	0.007
MOB + age + sampling month	0.10	0.04	2.7	0.009

Effect of MOB assessed by sequentially adding age and month of specimen collection to the initial linear mixed model.

In addition to MOB, age was also statistically significant variable. The positive association with the outcome indicates higher relative abundance of Moraxella with increasing age during the first year of life. On the other hand, the variable month of specimen collection was associated with the outcome when both MOB and age are included in the model. The results are shown on **Table 7**.

Table 7: Linear model including both potential confounders

Terms	Coefficient	SE	t-value	p-value
MOB	0.1	0.04	2.7	0.009
Age	0.21	0.03	6.6	< 0.001
Month of sampling	-0.02	0.02	1.1	0.3

Linear Mixed effects model including the variables MOB, age and calendar month of specimen collection.

To obtain higher precision of the effect of MOB on the outcome additional variables were also included in the analysis. The covariates included in the full model were birth weight, gestational age, type of feeding, birth order, length of exclusive breast feeding, time to formula feeding and cigarette smoke exposure and calendar year of specimen collection. An interaction term between MOB and age was also included. From the full model, the final linear model was obtained using stepwise model selection using AIC criterion. The variable that remained in the final linear model were MOB, age, birth order, length of exclusive breast feeding and calendar year of specimen collection. Model summary is shown on **Table 8**.

Table 8: Final linear model

Terms	Coefficient	SE	t-value	p-value
MOB	0.1	0.04	2.7	0.009
Age	0.23	0.03	6.9	< 0.001
Birth order	0.28	0.12	2.4	0.017
Exclusive breast-feeding length	-0.08	0.04	-1.8	0.07
Specimen collection year	-0.2	0.12	-1.6	0.1

Final linear mixed effects model obtained by applying stepwise model selection.

Plots of model residuals against MOB and covariates indicated a potentially non-linear relationship of MOB and calendar year at specimen collection with the outcome. The plot of the residuals of the final linear model against the MOB is

shown in **Figure 7**. In Figure 8 we observe that the mean deviation from the model is different for each month of birth which is an indication that the linear model is not adequate to describe the functional relationship of the outcome with the variable MOB.

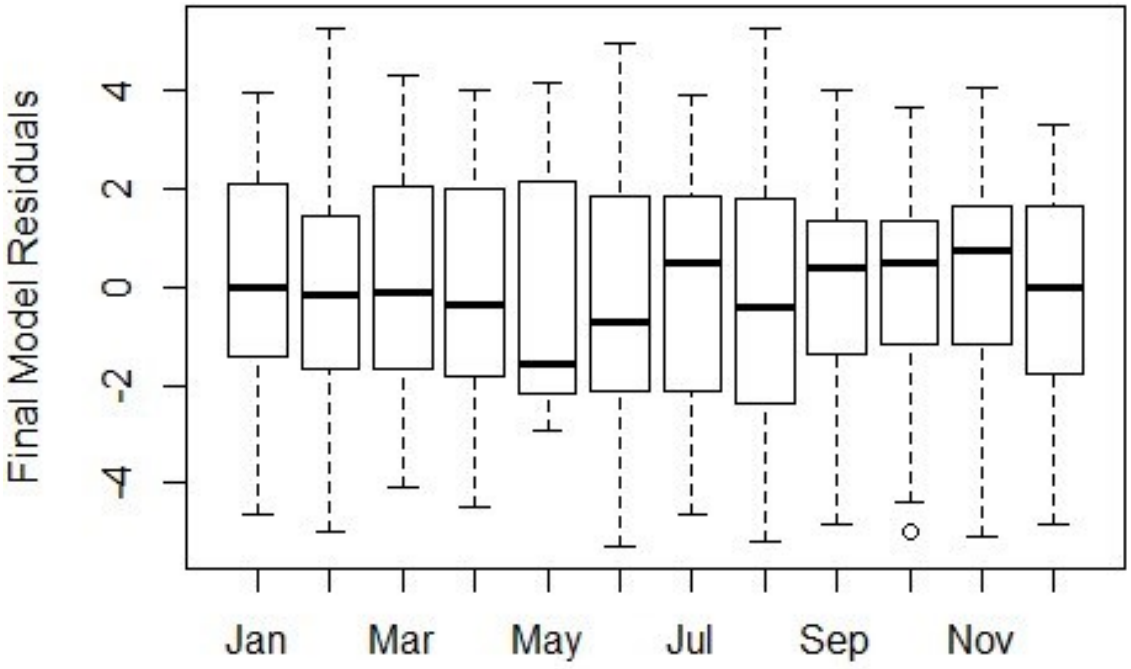

Figure 7: Residuals of final linear mixed effects model against MOB

The non-linear relationships were examined using Generalized Additive mixed models(Wood, 2019) (GAMM). The variables considered at this stage of the analysis were the ones of the final linear model; namely, MOB, age, birth order, length of exclusive breast feeding and calendar year of specimen collection. Due to special interest in the variable month of specimen collection and the interaction of age with MOB, the effect of these terms was also considered even though they were found not statistically significant according to the linear model.

More specifically, to examine non-linear relationships with the outcome smoothing spline functions were applied on the continuous variables 1) MOB, 2) age, calendar 3) month and 4) year of specimen collection and the transformed variables were used in the analysis. The non-linear interaction of age with MOB was modeled as a tensor product smooth function of the two variables. BIC was used as model selection criterion. Details of the GAMM model are presented on **Table 9**. statistically significant positive linear relationships ($p < 0.05$) were detected between the outcome and the variables 1) age and 2) birth order. Moreover, there was a statistically significant non-linear effect of the variables 1) calendar year and 2) MOB. The month at specimen collection was not statistically significant when either age or MOB or both were in the model. Similarly, as was the case in the final linear model, the interaction of age with MOB was not statistically significant.

Table 9: Summary of final GAMM

Linear terms	**Coefficient**	**SE**	**t-value**	**p-value**
Birth Order	0.26	0.11	2.3	0.02
Age	0.29	0.04	7.0	< 0.001
Non-Linear terms			F-value	p-value
Month of Birth (MOB)			4.9	< 0.001
Calendar Year			4.7	0.001

Summary of final Generalized Additive Mixed Model.

The relationship of the outcome with the year of specimen collection is displayed on **Figure 8**. The statistically significant association may be due to multiple factors that may change within a five-year period as well as the expected cycles of the relative abundance of microbiota in the environment.

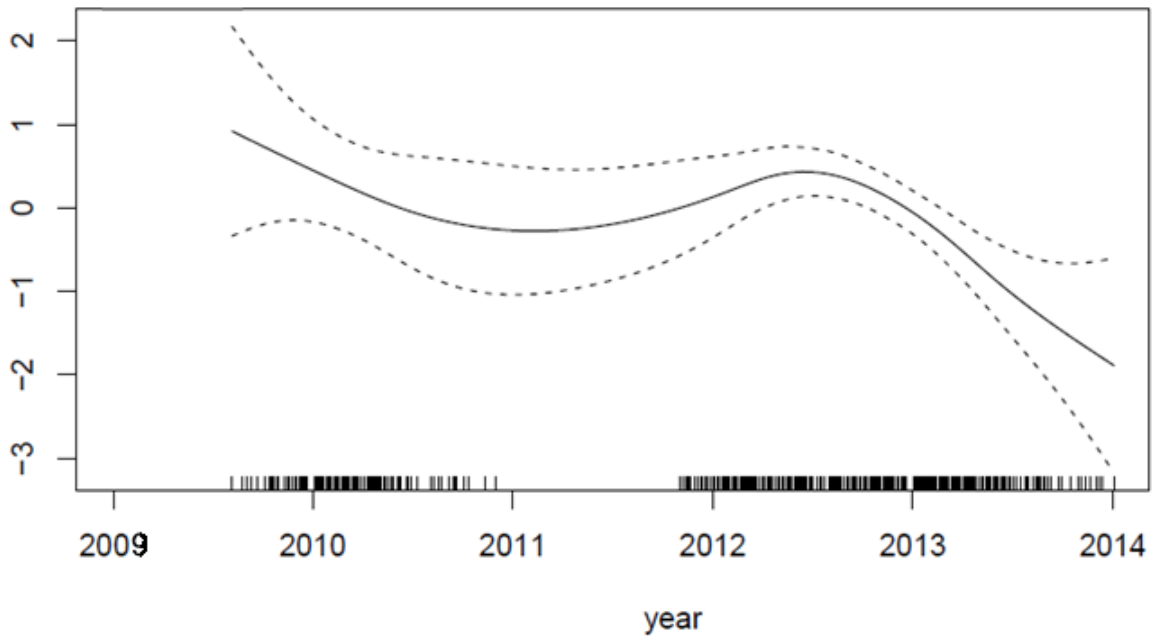

Figure 8: Non-Linear relationship of Moraxella relative abundance with year of specimen collection

The dashed lines indicate the standard error of the predicted mean. The vertical axis is the standardized log-relative abundance or Moraxella microbial abundance with the average mean across the whole time period (horizontal axis) corresponding to zero.

The statistically significant non-linear association between MOB and the outcome which was detected by the GAM model was further investigated by attempting to decompose the aggregate relationship into distinct elements of seasonal nature which may be amenable to clinical interpretation. To this end, several sinusoidal functions- varying in period -having as phase argument the variable MOB were fitted to the outcome, while including in the model all the statistically significant covariates identified in the previous model. Three statistically significant sinusoidal functions of MOB were detected. The results are summarized on **Table 10**.

Table 10: Summary of final model

Linear terms	**Coefficient**	**SE**	**t-value**	**p-value**
Birth Order	0.23	0.11	2.1	0.03
Age	0.31	0.04	8.3	< 0.001
Sinusoid - 12 months cycle	0.88	0.19	4.5	< 0.001
Sinusoid - 6 months cycle	0.5	0.18	2.8	0.005
Sinusoid - 4 months cycle	0.47	0.17	2.7	0.006
Non-Linear terms				**p-value**
Calendar Year				0.001

The variation of the Moraxella microbial abundance due to the distinct seasonal elements of MOB, as well as the aggregate effect, is depicted on **Figure 9** and **Figure 10**. The first element has 12 months period with peak intensity in end of November, the second has 6 months period with peak intensities in April and October and the third has 4 months period with peak intensities in March, July and November. The composite effect displays global maximum and minimum intensities at November and May respectively and a relative maximum in March.

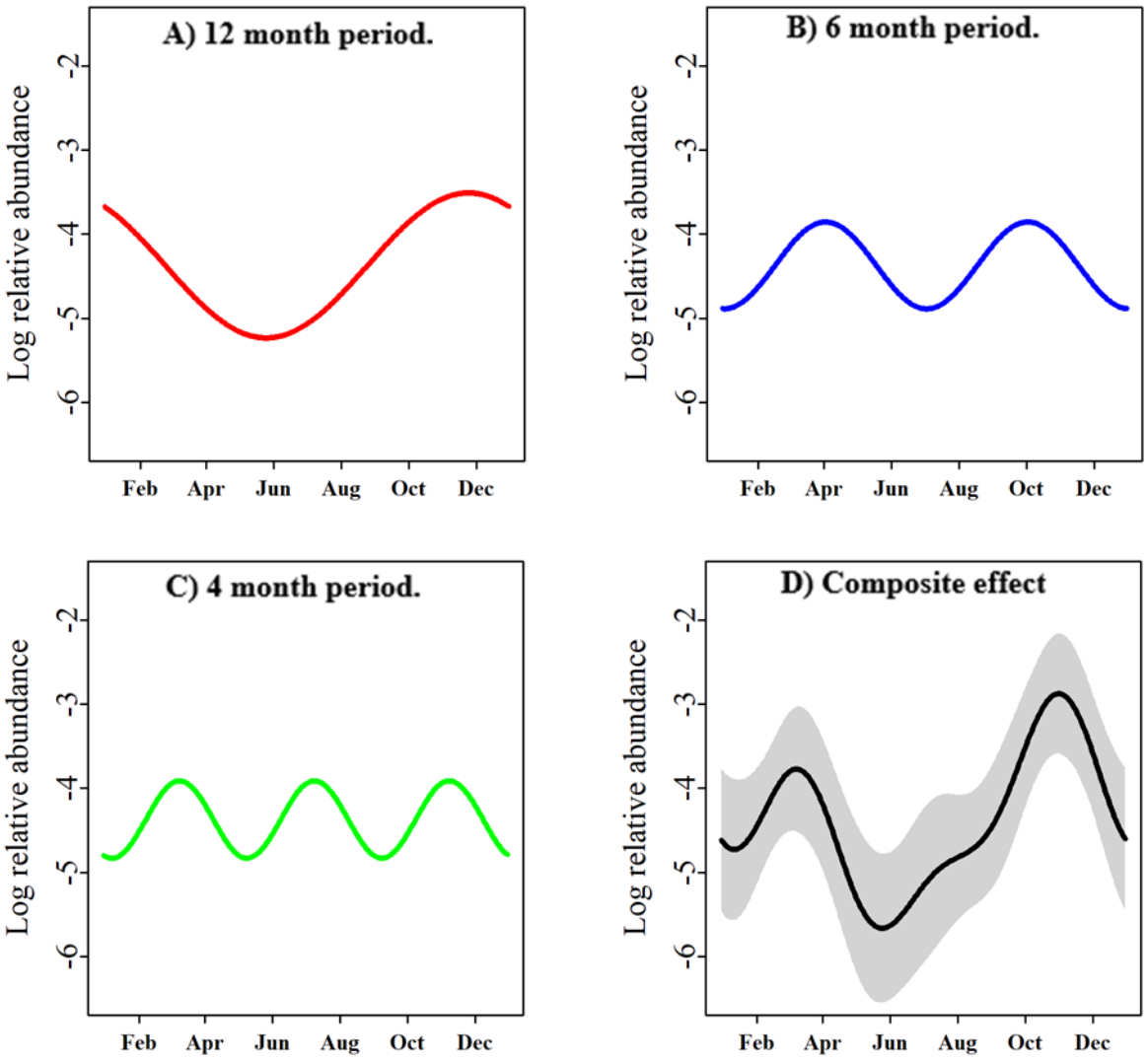

Figure 9: Variation of Moraxella microbial abundance due to distinct seasonal elements of MOB

A) 12-month period. B) 6-month period. C) 4-month period. D) Composite effect. The shaded region indicates the 95% confidence interval of the predicted mean across a calendar year.

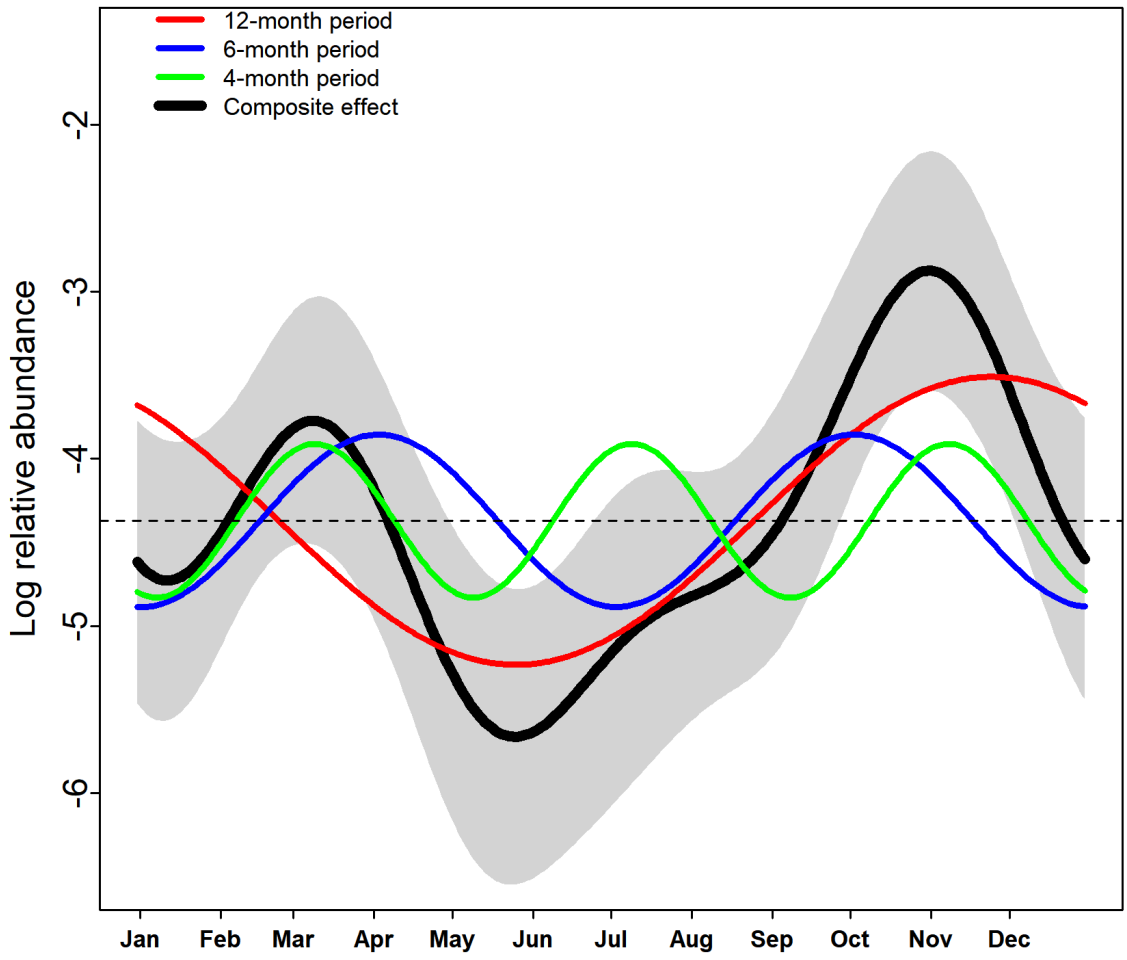

Figure 10: Aggregate effect of MOB on Moraxella microbial abundance

The aggregate effect is illustrated by the bold black line and the shaded region corresponds to the 95% confidents intervals estimated by the final GAMM model. The superimposed colored curves correspond to the distinct seasonal elements shown on Figure 9.

Chapter 4 Discussion

In this study we showed that there is a complex association between the relative abundance of Moraxella genera with MOB in infants during the first year of life. Time of birth determines the environment that a newborn is exposed to during the first months of life and likely affects otopathogen colonization. Moreover, the type and likelihood of exposure to certain respiratory viruses throughout the year may partly determine otopathogens' microbiome load. Better understanding of associated risk factors for AOM is required in the effort of reducing its public health burden.

Prior studies have shown that URI and AOM are mainly associated with daycare attendance(Ladomenou, Kafatos, Tselentis, & Galanakis, 2010; Lok, Anteunis, Meesters, Chenault, & Haggard, 2012), having siblings(K A Daly et al., 1999; Ladomenou et al., 2010; Lok et al., 2012) and (absence of) exclusive breastfeeding for at least 6 months(Ladomenou et al., 2010). In accordance to these studies, the original clinical study of this cohort(Chonmaitree, Trujillo, et al., 2017) reports the effect of several environmental factors on URI in infants up to 6 months old. The identified factors were number of siblings in the home, breastfeeding, day care attendance and being born after February 2010 (which is probably associated with the change of vaccination practice of using 13-valent protein-conjugate pneumococcal vaccine instead of 7-valent). The association of MOB with Moraxella genera microbial abundance, the most abundant otopathogen(Chonmaitree, Jennings, et al., 2017), that was shown in the original

study, indicates that the timing of birth may be used to identify high risk population for AOM. To our knowledge ours is the first study to assess the association between seasonality in birth and abundance of Moraxella, one of the most predominant pathogens in infants with AOM.

The main finding of this study is that there is a statistically significant non-linear association of the MOB with the relative abundance of Moraxella genera (**Table 9**) and this association remained significant when the effect of subjects' age and calendar month of specimen collection were accounted for (data not shown). The effect of MOB displayed a cyclic seasonal nature (**Table 10**, **Figure 9**). Specifically, specimens collected during the first year of life from subjects born in November had on average the highest relative abundance of Moraxella while those collected from subjects born in May had the lowest (**Figure 10**). Moreover, another local maximum was detected for subjects born in March. These results are consistent with the study of Daly et al. that showed association between recurring otitis media in infants and season of birth, albeit not accounting for specific pathogens; in that study infants born during fall were found to be 2.6 times more likely to present consecutive episodes of AOM compared to infants born in spring(K A Daly et al., 1999).

Among the secondary results of this analysis are the statistically significant associations of subject's age, birth order and calendar year of specimen collection with Moraxella microbial abundance. Within the range of 0 to 12 months, increasing age is linearly associated with (higher) log-transformed relative abundance of Moraxella microbial abundance. Older subjects have more time to

interact with the environment relative to younger subjects and therefore greater chance to encounter endogenous sources of infection. The positive linear association of birth order with higher Moraxella microbial abundance has also a reasonable casual path. Subjects living in the same environment with older siblings are more likely to carry higher microbial abundance and diverse profile of microbiota due to contamination by their older siblings. The non-linear relationship of Moraxella microbial abundance with calendar year of specimen collection shows an overall downward trend which is in accordance with the reduced risk for subjects born after February 2010 reported in (Chonmaitree, Trujillo, et al., 2017). This relationship is depicted on **Figure 8**. It is likely that the plateau shown for the time interval around the year 2011 is an artefact of the lower recruitment rate at this year.

There are certain limitations to this study; because the original cohort study focused on AOM, many subjects completed the study upon the first incidence of AOM or at 6 months of age (**Figure 3**). The smaller sample size for 6-12 months of age may have affected the analysis, specifically the interaction between age and MOB. While in this study we adjusted for known factors related to otopathogen microbial abundance, there are other potential mechanisms for explaining seasonality that we could not account for. Such factors would be environmental pollution levels, temperature or climate and seasonal variation in viral infection loads. Finally, while it is unknown whether the environmental exposure during the first months of birth may affect an infant's long-term risk for AOM infection, results from this study cannot be generalized beyond the first year of life.

Acute otitis media is one of the leading causes of pediatric doctor visits and antibiotic consumption and poses a significant public health burden. We have shown that month of birth is associated with microbial lode of Moraxella Catarrhalis, the most predominant bacterial infection in infant AOM cases. Early recognition of infants at higher risk of AOM can assist clinicians in decision making and providing prompt treatment to avoid complications.

Bibliography

Boland, M. R., Shahn, Z., Madigan, D., Hripcsak, G., & Tatonetti, N. P. (2015). Birth month affects lifetime disease risk: A phenome-wide method. *Journal of the American Medical Informatics Association*, *22*(5), 1042–1053. https://doi.org/10.1093/jamia/ocv046

Chonmaitree, T., Jennings, K., Golovko, G., Khanipov, K., Pimenova, M., Patel, J. A., … Fofanov, Y. (2017). Nasopharyngeal microbiota in infants and changes during viral upper respiratory tract infection and acute otitis media. *PLoS ONE*, *12*(7), 1–15. https://doi.org/10.1371/journal.pone.0180630

Chonmaitree, T., Trujillo, R., Jennings, K., & Alvarez-fernandez, P. (2017). Acute Otitis Media and Other Complications of Viral Respiratory Infection, *137*(April 2016).

Daly, K A, Brown, J. E., Lindgren, B. R., Meland, M. H., Le, C. T., & Giebink, G. S. (1999). Epidemiology of otitis media onset by six months of age. *Pediatrics*, *103*(6 Pt 1), 1158–1166. https://doi.org/10.1542/peds.103.6.1158

Daly, Kathleen A., Hoffman, H. J., Kvaerner, K. J., Kvestad, E., Casselbrant, M. L., Homoe, P., & Rovers, M. M. (2010). Epidemiology, natural history, and risk factors: Panel report from the Ninth International Research Conference on Otitis Media. *International Journal of Pediatric Otorhinolaryngology*, *74*(3), 231–240. https://doi.org/10.1016/j.ijporl.2009.09.006

Doblhammer, G., & Vaupel, J. W. (2001). Lifespan depends on month of birth. *Proceedings of the National Academy of Sciences*, *98*(5), 2934–2939. https://doi.org/10.1073/pnas.041431898

Ladomenou, F., Kafatos, A., Tselentis, Y., & Galanakis, E. (2010). Predisposing factors for acute otitis media in infancy. *The Journal of Infection*, *61*(1), 49–53. https://doi.org/10.1016/j.jinf.2010.03.034

Lok, W., Anteunis, L. J. C., Meesters, C., Chenault, M. N., & Haggard, M. P. (2012). Risk factors for failing the hearing screen due to otitis media in Dutch infants. *European Archives of Oto-Rhino-Laryngology : Official Journal of the European Federation of Oto-Rhino-Laryngological Societies (EUFOS) : Affiliated with the German Society for Oto-Rhino-Laryngology - Head and Neck Surgery*, *269*(12), 2485–2496. https://doi.org/10.1007/s00405-011-1896-0

Pinheiro, J., Bates, D., DebRoy, S., Sarkar, D., & R Core Team. (2018). {nlme}: Linear and Nonlinear Mixed Effects Models. Retrieved from https://cran.r-project.org/package=nlme

R Core Team. (2018). R: A Language and Environment for Statistical Computing. Vienna, Austria. Retrieved from https://www.r-project.org/

Schilder, A. G. M., Lok, W., & Rovers, M. M. (2004). International perspectives on management of acute otitis media: A qualitative review. *International Journal of Pediatric Otorhinolaryngology*, *68*(1), 29–36. https://doi.org/10.1016/j.ijporl.2003.09.002

Teo, S. M., Mok, D., Pham, K., Kusel, M., Serralha, M., Troy, N., … Inouye, M. (2015). The infant nasopharyngeal microbiome impacts severity of lower respiratory infection and risk of asthma development. *Cell Host and Microbe*, *17*(5), 704–715. https://doi.org/10.1016/j.chom.2015.03.008

Wood, S. (2019). mgcv: Mixed GAM Computation Vehicle with Automatic Smoothness Estimation. Retrieved from https://cran.r-project.org/package=mgcv

Vita

Ioannis Malagaris was born to Georgios and Aggeliki Malagaris on 28 November 1983 in Athens, Greece. He graduated from Harokopio University of Athens with a Bachelor of Science in Nutrition and Dietetics in 2007 and a Master of Science in Sports Nutrition in 2013. In August 2016, he enrolled to Master in Public Health with a concentration in Biostatistics program at the Department of Preventive Medicine and Community Health at the University of Texas Medical Branch in Galveston. He currently resides in Galveston, Texas.

This dissertation was typed by Ioannis Malagaris