Clinical phenotypes associated with impaired response to treatment in children with acute asthma

Workplace Project Portfolio

Master of Biostatistics

University of Sydney

Noviani Minaee (311259723)

TABLE OF CONTENTS

Preface

Student's role

I was responsible for all analyses and writing up the report. My supervisors Dr Zhang and Dr Laing came up with the project. The aim of this project is to try to find cluster memberships based on clinical phenotypes in children with severe asthma and determine whether we could use it to predict their response to treatment. They thought that as someone who has no preconceived idea on asthma, I'd be ideal to do this project. The idea is for me to find clusters that have never been found before. To start with, I performed a preliminary exploratory data analysis on the clinical phenotypic data, questionnaire data and blood sample levels data to get better ideas on the data itself and to determine which variables I should use. Prof Nick de Klerk was kind enough to come on board to supervise me towards the middle of my project and offered a different perspective in seeing the data. After further consultations and more in-depth investigation on asthma clinical phenotypes, we decided to use multivariable linear regression to select which variables I would use for my cluster trials. This way I would also have a better understanding of the relationship between each covariate and the outcome variable – which was the response to treatment. Both univariate and multivariable linear regressions were conducted. I started the selection of covariates process univariately and then all covariates that were significant at 20% level were considered as potential covariates for the inclusion into the multivariable model. The variables used in the final multivariable model were deemed the best fitted and then used in the cluster analysis to build the cluster memberships. The final cluster memberships used in this project were constructed after several trials and tests, both within the clusters themselves and between the clusters and the outcome of interest. The clusters were also tested against genotypes of several important asthma candidate genes and virus data to test for differing susceptibility to

viral infection. Lastly, I performed multivariable Poisson regression on the available hospital presentation data to see if the recurrence of hospitalization varied between phenotypes and was influenced by any particular respiratory risk factors. All the data used in this project was derived from results from MAVRIC study. My supervisors and I intend to submit aspects of this report for publication.

Reflections on learning

Clinically, undertaking this project was a very challenging task for me. I had no clinical background whatsoever, especially in paediatric asthma. I had to learn, and am still learning, a lot about asthma and its impact on children. I also had to learn about a lot of allergy, genetics, viruses, blood samples and plasma levels and all the accompanying biomedical terminology to get a well-rounded idea about paediatric asthma and what MAVRIC study are all about.

Statistically, this project also imposed lots of challenges. Cluster analysis is a new concept for me and was not covered in BCA units. One of my statistical supervisors, Dr Zhang, came up with this idea. He advised me to use the k-means clustering method for the purpose of my analysis. The reasoning was because k-means method is quite simple to implement and has been successfully utilized in medical and applied statistics (Haldar et al., 2008). As with the clinical subject, I had to learn and familiarize myself with the clustering method. This process took months, especially when I started conducting my cluster trials. At the end, I ended up combining multivariable regression and cluster analysis in the variable selection process and to have better understanding of the dynamics of the relationship between the covariates and the outcome of interest both individually and combined. Prof De Klerk further suggested for

me to investigate the hospitalization recurrence after the participants were recruited in the study. As the hospital visits data were count data, I used negative binomial regression model for the analysis. In doing the overall statistical analysis, I drew a lot from skills and knowledge I gained from BCA units in particular Linear Modelling (LMR) and Categorical Data Analysis (CDA).

Another major thing I learnt along the way is about the data itself. MAVRIC is a large study that has been going on for more than a decade. The original questionnaire that they used has evolved accordingly to reflect this. As a result, familiarization with the dataset has proven to be a difficult task. There are easily more than a thousand variables in the overall dataset. The dataset is constructed from various individual databases: questionnaire, acute data, short-term follow up data, long term follow up data, skin prick test, FBC, lung function test, levels, PNA (virus), PCAAS and genotype. Although I would not be using or needing all of these data I was still required to be familiar with them well enough to choose the most appropriate variables and to interpret their values and effects correctly. The whole process took me just over 1.5 years.

Teamwork

This project involved close collaborations with Dr Brad Zhang, Prof Nick de Klerk, Dr Ingrid Laing, who are my statistical and clinical supervisor respectively. Dr Zhang and I had many discussions and conducted many clustering trials together throughout the course of the project. Prof Nick de Klerk provided an additional and invaluable support for the statistical challenges I faced, in the discussion of the analysis and the interpretation of the results – especially when the clustering analysis alone did not yield sufficient information to base our

conclusions on. We met frequently to discuss about what other options we could do with the data we had. He suggested the multivariable linear regression and negative binomial regressions on top of the cluster analysis to answer some pertinent research questions. Dr Laing and I mostly discussed about the clinical side of the project. She helped me to view the results from a wider context and how it all fits the bigger picture. She helped me to apply statistics to a clinical setting and for it to be understood by a wider audience. Prof Peter Le Souef, the head of School of Paediatrics and Child Health, was there to oversee the whole process. He was always supportive when I expressed the difficulty I had with the data at times at the team's weekly lab meeting. I also closely collaborated with Dr Joelene Bizzintino, who is in charge of the overall MAVRIC database. Her assistance was invaluable.

Ethical considerations

MAVRIC study obtained consent form from study participants and ethics approval (1761/EP) for recruitment and subsequent follow-ups from the Ethics Committee at Princess Margaret Hospital (Perth, WA, Australia).

Project Summary

Project title

Clinical phenotypes associated with impaired response to treatment in children with acute asthma

Location and dates

This project was completed at the School of Paediatrics and Child Health at the University of Western Australia between January 2014 and September 2015.

Context

This project is part of a larger study titled “Mechanisms of Acute Viral Respiratory Infection in Children (MAVRIC)”. It is a prospective study of the role of viruses and immunogenetic risk profiles for viral infections. My project's aim is to identify distinct phenotypic groups using cluster analysis and to explore the association between those identified clusters both with asthma candidate genes and with a susceptibility to human rhinovirus. The acute asthma cohort from the MAVRIC study provides a unique opportunity to investigate this research question.

Student contribution

- Investigation of asthma clinical phenotypes and potential risk factors in children
- Exploratory data analysis and preparation of data for analysis
- Fitting of linear regression models to determine the relationships between asthma clinical phenotypes and potential risk factors in children

- The construction of cluster memberships to see if there were any groups emerging within these clinical phenotypes and potential risk factors
- Conducting the cluster analysis to see if the cluster memberships were associated with the response to treatment, and a selection of asthma-related genotypes.
- Fitting of negative binomial regression models to determine the association between cluster memberships, asthma clinical phenotypes and potential risk factors, with the recurrence of hospital presentations after the children were recruited into the MAVRIC study.

Statistical issues

- Evolving questionnaire

  MAVRIC study began more than 10 years and since then the questionnaire used has evolved. This has imposed constant data cleaning and analysis challenges. Many questions have become obsolete and were duly replaced. This in turn has caused responses to these questions becoming patchy.

- Missing data

  A direct effect from an evolving questionnaire has been missing data. There had been questions that participants were asked differently depending when they were recruited. This has led to a considerable amount of missing data.

  Another issue was the fact that not all children were tested for respiratory viruses and the asthma-related genes were only typed for small number of cases. This caused a major missing data issue.

- Multiple handling of data entry

  Further inconsistencies in the data can be attributed to the data entry process, which was carried out by different people. Differing interpretations have produced multiple

problems. For example, answers to the question “How many time did your child experience asthma exacerbation in the past 3 months?” vary from “twice” to “2”, “two”, “two times” and “?”. This inconsistency has made data cleaning a time-consuming process.

- Lack of data dictionary at the beginning

  A considerable challenge was the absence of a data dictionary for the project that would have provided a description of variables and the theoretical foundations for the questions. My clinical supervisor's help in interpreting the project was invaluable in slowly coming to grips with the detailed nature of the study.

- Selection of variables for data analysis

  MAVRIC has vast amount of variables in the dataset. It took some familiarization with the study itself, the topics of paediatric asthma, immunology, and genetics, to try narrowing it down to a more asthma-specific focus, so that appropriate potential variables could be considered for analysis purposes.

Acknowledgements

I would like to express my deep gratitude to the Asthma Genetics Group of the UWA School of Paediatrics and Child Health for providing me with an opportunity to gain working experience, supplying me with a project for my workplace project portfolio, and for allowing me to use their MAVRIC cohort for my project.

Last but not least, I would like to thank all my supervisors for their endless support throughout this project.

Student declaration

I declare, that, with assistance provided by my supervisors, all work referred to in this report has been my own and has not been previously submitted for academic credit.

Noviani Minaee

10th November 2015

Supervisor declaration

Novia has worked very hard collecting and collating the various data sources in this long-running and fairly complex project. She was fully engaged in the work from her first involvement and has learned and mastered the various statistical analysis methods that have been necessary in answering the different study questions. She had to spend a great deal of time cleaning and editing all the data as well as examining all the anomalies in data coding and collection that have arisen during the study's evolution. For much of the time she worked independently, only requiring guidance while completing the final stages of the work. She is now in a good position to be able to apply the methods and principles learned in other areas and in other jobs, and I think this report represents a valuable piece of research.

Professor Nick de Klerk

10th November 2015

Glossary

Atopy	A predisposition towards developing certain allergic hypersensitivity.
Eosinophils	White blood cells that control mechanisms associated with allergy and asthma. Increased level of eosinophils in peripheral blood and in airway secretions are characteristic feature of asthma.
Neutrophils	A type of white blood cells that formed an essential part of the innate immune system. Increased level of neutrophils in induced sputum is usually a sign of asthma.
Salbutamol	A short-acting β_2-agonist used for the relief of bronchospasm in conditions such as asthma and chronic obstructive pulmonary disease.

List of Abbreviation

Adv	Adenovirus
ED	Emergency Department
HRV	Human Rhinovirus
HRV-C	Human Rhinovirus (strain) C
IgE	Immunoglobulin-E
InfV	Influenzae (strain) A
Myco	Mycoplasma Pneumoniae Virus
PIV	Parainfluenzae Virus
PNA	Peptide Nucleid Acid
RSV	Respiratory Syncytial Virus
SPT	Skin Prick Test
VRI	Viral respiratory infection

Project report

Clinical phenotypes of acute asthma associated with impaired response to treatment in children with acute asthma

Introduction

Asthma comprises inflammation of the airways, characterised by reversible airways obstruction, airway hyper-responsiveness (AHR) and bronchial spasm (Dougherty & Fahy, 2009). Asthma, particularly childhood asthma, is a National Health Priority costing Australia >$606 million annually. First symptoms of asthma usually occur during the first few years of childhood. Of those, a large proportion of children become symptom free by the time they reach school age and the rest continue to develop persistent asthma throughout childhood (Martinez & Vercelli, 2013).

The aetiology of asthma is complex as it is multifactorial, with contributions from both genetic heritability and environmental stimuli. The interaction between genetics and environment makes asthma a heterogeneous syndrome with overlapping individual phenotypes (Borish & Jeffrey, 2008). This heterogeneity of asthma makes the task of treating asthma quite problematic, as the specific mechanism of asthma that leads to the airway dysfunction has still not been properly identified (Drazen, 2012). As a result, the response to asthma medication varies considerably from patient to patient depending on the level of sensitivity of the different asthma phenotypes (Borish & Jeffrey, 2008).

The causes of acute asthma exacerbations are numerous. **E**nvironmental stimulus such as virus and allergens can cause inflammation of airways with resultant loss of control of

asthma, which results in an exacerbation (Lenney, 2009; Martin et al., 2006). Genetic susceptibility can also make children more prone to developing asthma (Taussig, 2002). The likelihood of these precipitants to cause exacerbation of this heterogeneous disease varies among children depending on their asthma phenotype (Dougherty & Fahy, 2009).

Viral respiratory tract infections, in particular those caused by human rhinovirus (HRV), are believed to be the most common cause of acute asthma exacerbations (Bizzintino et al., 2011; Jackson et al., 2010). It is very likely because HRVs can affect the lower airway and are very common infection in infant and young children (Cox & Le Souef, 2014; Papadopoulos et al., 2000). Other viruses such as respiratory syncytial virus (RSV) can also cause acute exacerbations, with a higher prevalence found on children with a family history of asthma (Bizzintino et al., 2011; Papadopoulos et al., 2011).

Genetic predisposition plays an important role in the development of childhood asthma (Hodges, 1996; Sibbald et al., 1980). Twin studies from 1970s to 1990s found evidence of strong genetic component in the factors underlying allergic and Immunoglobulin-E-(IgE)-mediated disease such as asthma (Edfors-Lubs, 1971; Duffy et al., 1990). This leads to a greater co-existence of the disease in monozygotic compared to dizygotic twins. Tucson's 1989 longitudinal respiratory study of in children found that those with a family history of asthma were more likely to develop persistent asthma compared to those without a family history of asthma (Taussig, 2002). It is thought that these genetic predispositions cause children to develop hypersensitivity to specific allergens that make them susceptible to developing allergy and IgE-mediated disease.

Martin et al. (2006) found that CD14 and CC16 played an important role in the immune system and airway inflammation. The increased levels of both protein levels illustrated their role during an acute episode. Their study also found that children with CD14 C-159T and CC16 A38G genotypes are more likely to suffer a more severe level of acute asthma exacerbation compared to those with other genotypes.

Immunoglobulin E (IgE) is a type of antibody known to trigger inflammation and allergic reactions (Platts-Mills, 2001). Total serum IgE and allergen-specific IgE are two measurements that are commonly used to establish a child's allergic status. A high level of total IgE concentrations usually is sufficient to establish a presence of allergy; nonetheless, a low level of total IgE does not preclude the presence of allergen-specific IgE (Sinclair & Peters, 2004). Thus, the allergen-specific IgE is often requested to accompany total IgE as well. Furthermore, allergen-specific IgE measurement also provides additional information on the level of allergy to a specific allergen or group of allergens that is not available from total IgE.

Another most widely used method to determine allergy and atopy status in children is via skin-prick testing (SPT). SPT results are commonly used in conjunction with total serum IgE and allergen-specific IgE as together they provide better chance of allergy detection (Sinclair & Peters, 2004). Total serum IgE, allergen-specific IgE, SPT results and atopy status are among the most common phenotypes used as markers in asthma studies (Martinez & Vercelli, 2013).

Few studies have tried to examine the different clinical phenotypes of asthma in both adults and children (Haldar et al., 2008; Kelley et al., 2005; Moore et al., 2009). Haldar et al. (2008)

and Moore et al. (2009) both applied the principle of cluster analysis to identify distinct asthma phenotypic groups in adults. Kelley et al. (2005) used multivariable regression to determine if risk factors and measures of severity varied between children with different asthma phenotypes.

To date, no studies had tried to clarify asthma phenotype groups in children with severe asthma. The Mechanisms of Acute Viral Respiratory Illness in Children (MAVRIC) study provides a unique opportunity to investigate this research question using the combination of the use of cluster analysis and multivariable regression techniques.

This project aims to predict the treatment response by identifying a specific group of children with acute wheezing and asthma who do not respond the treatment well and are likely to relapse and re-present to hospital. The recurrence of hospital presentations was further examined to see whether it is associated with specific asthma phenotypes and potential risk factors.

Methods

Study participants

Study participants were children recruited from the Mechanisms of Acute Viral Respiratory Infection in Children (MAVRIC) Study that began in 2001. The MAVRIC study recruited children with acute viral respiratory infection that led to hospital admission, between the ages 0 to 18, with and without a cold at the time of recruitment from the Emergency Department (ED) at Princess Margaret Hospital for Children (PMH). As MAVRIC is an ongoing study and continues to recruit participants, a cut-off was made at case number 702 for the purpose of analysis (n = 702).

Potential covariates

There were six types of data used in this study:

1. The questionnaire data. This consists of the child's history of allergy and asthma exacerbation, hospital presentation, family background and history of allergy and asthma.
2. Skin prick test data. This consists of the results from skin prick test from 11 allergens. The complete list of the 11 allergens is provided in Table 1.
3. Full blood count data. This consists of peripheral blood differential count during acute asthma exacerbation.
4. Serum levels for total IgE and allergen-specific IgEs during acute asthma exacerbation.
5. PNA (Peptide Nucleid Acid) data. This consists of information from virus typing results.
6. Genotypes data. This consists of information from genotyping results.

Table 1: List of 11 allergens used in skin prick test

Allergens
1. Mixed ryegrass
2. Alternaria tenuis (fungus)
3. Dermatophagoides farinae (American house dust mite)
4. Whole cow's milk
5. Dermatiphagoides pteronyssinus (European house dust mite)
6. Aspergillus fumigatus (mould or filamentous fungi)
7. Egg white
8. Mixed grasses
9. Cat pelt
10. Dog hair and dander
11. Cockroach – American and German

One of the limitations of the study was that not all children got tested for viruses and not all the genes were typed, which resulted in a significant number of missing values. So for the purpose of the statistical analysis, only variables collected for 500 or more children were selected. For the gene-specific variables, only those that were genotyped in more than 180 children were selected. The lists of potential variables are listed below:

Table 2: List of potential continuous variables

Continuous variables
1. Age
2. Neutrophils level
3. Eosinophils level
4. Severity of acute asthma exacerbation

Table 3: List of potential categorical variables

Categorical variables
1. Gender
2. Viral infection to respiratory-related virus present during acute exacerbation
3. Human rhinovirus (HRV) present during acute exacerbation
4. Human rhinovirus C (HRV-C) present during acute exacerbation
5. Adenovirus (AdV) present during acute exacerbation
6. Respiratory syncytial virus (RSV) present during acute exacerbation
7. Influenzae A (InfV) virus present during acute exacerbation
8. Parainfluenzae virus (PIV) type 1, 2 or 3 present during acute exacerbation
9. Mycoplasma pneumoniae (Myco) virus present during acute exacerbation
10. Aero-allergy status (encompasses both aero-atopy and allergy)
11. Child's asthma diagnosis status
12. Mother past smoking status
13. Mother current smoking status
14. Household member's past smoking status
15. Household member's current smoking status
16. Child usually have a cough
17. Child ever wheezed
18. Family history of asthma
19. Previous hospitalization due to asthma

Table 4: List of potential gene-specific variables

CD14_159CT	GSTP1114	TLR3_6300CT	TLR_746AG
CC16_38AG	IL8_781CT	TLR3_8441AT	X5LO_1708GA
CysLT1_927TC	IL8_251AT	TLR7_17961AT	ST2_26999AG
GSTM1	IL13_1112CT	TLR7_4452CT	LTC4S_444AC
GSTP1105	IL13_130GA	TLR8_558CT	LTC4S_1072GA

Outcome variables

Short-term outcome

- Length of hospitalization (hours).

  This represents how long each patient was kept in the hospital starting from when they presented themselves to ED due to acute exacerbation and finishing when they were deemed fit enough to be discharged.

- Dose of β_2-agonist administered within the first 24 hours of asthma exacerbation.

  This represents the emitted dose of inhaled β_2-agonist needed by patients to relieve their exacerbation within 24 hours. One dose equals to six puffs whereby each puff contains 100mcg of salbutamol (β_2-agonist).

Long-term outcome (***after*** recruitment)

There were two aspects of long-term outcome measured in this study: (1) the total number of inpatient admissions; and (2) the total number of ED visits resulting from respiratory illness, that occurred after the children were recruited into the MAVRIC study. The inpatient admissions and ED visits were chosen because they were the most relevant measures of acute asthma exacerbations.

Data preparation

1. Binary coding was done for most of the categorical variables.
2. A preliminary exploratory data analysis was performed on both continuous and categorical variables, including the short-term outcome variables. The results indicated that age, neutrophils level, eosinophils level, length of hospitalization and β_2-agonist dose were skewed. Hence they were all log-transformed. Natural log (ln) was used for the transformation. A more detailed result from the preliminary exploratory data analysis is provided in Appendix 1.
3. The severity score. The score represents the severity of acute exacerbation. Severity scoring results were standardized using z-scores because the scoring mechanism was different for children under 2 and those 2 and over. Standardizing the results allowed the scores to be amalgamated and thus enable the level of severity to be compared across different ages. For children 2 and over, the score was determined by MAVRIC investigator using a modified National Institutes of Health Scores (Reddel et al., 2009). It applied clinical parameters that were corrected to baseline (Bizzintino et al., 2011). The scoring had three categories:
    - Mild: score of 0-2
    - Moderate: score of 3-6
    - Severe: score of 7-10

    For children under 2, different clinical scoring system adapted from Bentur et al. (1992) was applied. There were four variables the children were tested on: heart rate, respiratory rate, the degree of accessory muscle use and wheezing. The sum of the individual score for these four variables made up the total clinical score for each child. More details of the clinical scoring system are provided in Table 5.

Table 5. The clinical scoring system for acute asthma exacerbation for children < 2

Score	Heart Rate	Respiratory Rate	Wheezing	Accessory Muscle Use
0	≤120	≤30	None and well	None
1	121 – 140	31 - 45	End-expiratory only	Mild
2	141 – 160	46 - 60	Entire expiration + inspiration with stethoscope only	Moderate (including tracheosternal)
3	>160	>60	Loud wheezing audible without stethoscope (or silent chest in the presence of tachypnea)	Severe with nasal flare

4. Aero-atopy status: This indicates whether someone is atopic or not, based on their skin prick test result; and allergen-specific IgE levels (house dust mites and cats) during an acute exacerbation. As the focus is on airflow obstruction and airway inflammation, only skin prick test results from inhaled allergens were included here.

Statistical analysis

There were three components of statistical methodology used in this study: Multivariable linear regression analysis, cluster analysis and multivariable Poisson/negative binomial regression analysis. All analysis was conducted in both Stata version 14 and SPSS version 22.

1. Multivariable linear regression

Multivariable linear regression was conducted to determine whether response to treatment varied between phenotypes and whether potential risk factors played a part in it. The response to treatment was measured by the length of hospitalization and the

dose of β_2-agonist required within 24 hours of acute exacerbation. This method was chosen mainly because (1) the outcome (the response to treatment) was a continuous variable; (2) the covariates (the phenotypes and the potential risk factors) were a mixture of continuous and categorical variables; (3) there were potential confounders that need to be taken into account when ascertaining the effect of the covariates of interest. Basically this study would like to create a model that explains the variation in the response to treatment conditional on the values of phenotypes and risk factors.

Multivariable linear regression was also used to determine if genotypes played a role in how a child responds to a treatment. Selection of variables for the purpose of multivariable linear regressions analysis was based on the univariate linear regression analysis results. Univariate linear regression was conducted for each of the predictor variable against each of the outcome variable. All variables that were significant at the 20 percent level were deemed as candidates for inclusion in the multivariable analysis. However, for the alpha level was set at 0.05 for the final multivariable model. Model checking for the final multivariable model is provided in the Appendix 2.

2. Cluster analysis

Cluster analysis was conducted as an exploratory investigation of the collected questionnaire data, skin prick test data, full blood count data, peptide nucleic acid data and levels data to see if there was any groups emerging within these clinical phenotypes and potential risk factors; and how these clusters associated with the length of hospitalization, dose of β_2-agonist required within 24 hours of exacerbation, and asthma-related genotypes such as CD14 159CT and CC16 A38G that have been

shown to react differently depending on the severity level of acute asthma in children (Martin et al., 2006). Linear regression was used to identify variables to be included in the cluster analysis with variables that were significant at the 20 percent level were deemed as candidates for inclusion.

Kaufman and Rousseeuw (1990) defined cluster analysis as "the art of finding groups in data". There are several methods of cluster analysis, whereby the difference between them mostly lie in how they measure the "distance" which is the similarity or the dissimilarity between observations (Everitt et al., 2011).

A k-means cluster analysis method was chosen over another method because it best suited MAVRIC's medium-sized data, and because it does not require complex calculation of the "distance", as it uses Euclidean distance in its calculation by default. K-means clustering works by simply partitioning the available observations (n) into k clusters and then grouping them into a cluster with the nearest means. The only drawback with k-means method is that the number of clusters or "k" must be decided prior to the analysis. Mirroring the approach taken by Haldar et al. (2008) on their cluster analysis study, after the emergence of clusters, the between-clusters comparison of covariates of interest was performed using one-way analysis of variance (ANOVA) for continuous variables and a chi-square test (χ^2) for categorical variables. Univariate ANOVA with the cluster membership as a covariate was further performed to verify the significance of the cluster as a predictor for any observed differences in the outcome.

3. Multivariable Poisson/negative binomial regressions

For the last part of the analysis, the study wished to explore the effects of the enrolment admission characteristics on the rate of further ED visits and hospital admissions. Poisson and Negative Binomial regressions were chosen because the rate of occurrence of ED visits and inpatient admissions were count variables.

Poisson regressions were first applied to the data and were then compared with negative binomial models to look for over-dispersion or variability in the admission/visit rates that was greater than would be expected under a Poisson distribution. Negative binomial regression fit the model in a manner similar to Poisson regression except it allows extra variation in the count data than Poisson which would accommodate any potential case of over-dispersion. The result of the comparison between the two models indicated significant over-dispersion and so negative binomial regression was conducted throughout to determine whether recurrence of inpatient and ED visits varied between phenotypes and whether potential risk factors played a part in it.

The hospital presentation data collected from all public hospital in WA combined with MAVRIC data was used for this purpose. Cluster memberships resulting from the cluster analysis were also included as explanatory variables to determine whether the amount of inpatient and ED visits varied between clusters.

The selection of variables for the purpose of multivariable negative binomial regressions was based on the univariate negative binomial regression analysis results. Univariate analysis was conducted for each of the predictor variable against each of

the outcome variable. All variables that were significant at the 20 percent level were deemed as candidates for inclusion in the multivariable analysis. Variables were included in final models if the p-value for removal was less than or equal to 0.05 or they were probable or known confounders.

The potential covariates remained the same as the ones used for the multivariable linear regressions analysis with the addition of cluster membership information. However the outcome of interest was now set to the number of inpatient and ED visits ***after*** the recruitment to MAVRIC study. Length of hospitalization and dose of β_2-agonist required within first 24 hours of acute asthma exacerbation were included as potential determinants in addition to the aforementioned phenotypes. The analysis was adjusted for age at recruitment. The duration of time at risk was counted from the date of recruitment into the MAVRIC study until the date of final observation of the WA hospital admission data (which was the same for everybody - March 31, 2014).

Fractional polynomial models were applied to the continuous variable so as to find the best fit for the covariates of interest when they were entered to the linear predictor in the negative binomial regression model (Sauerbrei et al, 2007). Only children aged 12 months and older were included in the analysis for two reasons: (1) For most of these infants, the recorded trip to hospital was their first one; (2) Most of these infants were not given the reliever that was given to the older ones – i.e. they were not given salbutamol (β_2-agonist). Therefore the existence of their data had the potential to skew the covariates distribution for the rest of the study cohort and thus was excluded. Model checking for the final multivariable model (including test for over-dispersion) is provided in the Appendix 2.

Results

Summary of study participants' characteristics

Overall there were 700 cases observed in this study, however due to missing observations, cases that were not tested for virus and lost hospital notes, the majority of variables had less than 700 cases in them.

The average age was 5 years old. The majority of the cases were male (60%), asthmatic (65%), and were atopic and/or allergic (62%). 81% of them were positive for viral infection to respiratory-related virus, and of those tested positive, 69% of them were positive for human-rhinovirus. 86% of the cases had bouts of wheezing in the past 12 months. 53% had previously been hospitalized due to asthma and 82% had family history of asthma. More details on the study participants' characteristics are provided in Table 6 and Table 7.

Table 6. Summary of categorical data

Variable [*n* (%)]	***n***		
Gender	694	Male [418 (60)]	Female [276 (40)]
Positive for respiratory-related virus	628	Yes [509 (81)]	No [119 (19)]
HRV positive	607	Yes [421 (69)]	No [186 (31)]
HRV-C positive	604	Yes [235 (39)]	No [369 (61)]
Adenovirus	701	Yes [10 (1)]	No [691 (99)]
Respiratory virus	502	Yes [74 (15)]	No [428 (85)]
Influenzae A	500	Yes [5 (1)]	No [495 (99)]
Mycoplasma pneumoniae	616	Yes [3 (1)]	No [613 (99)]
Aero-atopy and allergy status	701	Yes [437 (62)]	No [264 (38)]
Asthma diagnosis status	688	Yes [444 (65)]	No [244 (35)]
Mother's past smoking status	668	Yes [248 (37)]	No [420 (63)]
Mother's current smoking status	668	Yes [118 (18)]	No [550 (82)]
Household member's past smoking status	542	Yes [236 (44)]	No [306 (56)]
Household member's current smoking status	641	Yes [168 (26)]	No [473 (74)]
Presence of cough	664	Yes [255 (38)]	No [409 (62)]
History of wheezing	662	Yes [571 (86)]	No [91 (14)]
Family history of asthma	666	Yes [546 (82)]	No [120 (18)]
Previous hospitalization due to asthma	664	Yes [349 (53)]	No [315 (47)]

Table 7. Summary of continuous data

Variable [median (IQR)]	***n***	
Age, years	700	4 (2 to 7)
Neutrophil level, x 10^9 g/L	512	8.2 (5.17 to 10.9)
Eosinophil level, x 10^9 g/L	512	0.05 (0 to 0.24)
Length of hospitalization, hours	680	28 (16.5 to 52)
Inhaled dose of β2-agonist in 24 hours, mcg	541	13 (9 to 18)
Severity of acute exacerbation, z-score	613	-0.11 (-0.53 to 0.74)

Short-term outcomes

Both univariate and multivariable linear regression analysis were used in this study. Univariate linear regression analysis was conducted to determine the relationship of each potential covariate with each of the outcome variable. Any result from univariate analysis that was significant at 20% level was included in the multivariable analysis.

Length of hospitalization (hours)

Age (p = 0.160), female (p = 0.002), positive for RSV, Adv and PIV (p = 0.006, 0.004 and 0.035 respectively), mother ever smoke (p = 0.003), mother smoking now (p < 0.001), household member used to smoke (p = 0.05), mother smoking now (p = 0.002), coughing (p = 0.036), previous hospitalization due to asthma (p < 0.001), asthmatic (p = 0.005) and the severity of acute exacerbation (p < 0.001) were significant in influencing the patients' length of hospitalization.

Female, virus infection by respiratory-related virus, maternal current smoking status, household-member's smoking past, previous hospitalization due to asthma and the severity of acute exacerbation were found to be significant in predicting patients' length of hospitalization.

Dose of β_2-agonist within the first 24 hours of acute asthma exacerbation

Age ($p < 0.001$), neutrophil level ($p = 0.003$), positive for respiratory-related virus ($p = 0.04$), positive for HRV ($p = 0.008$), positive for HRV-C ($p = 0.189$), positive for RSV ($p = 0.003$), positive for aero-atopy and allergy ($p < 0.001$), mother ever smoked ($p = 0.041$), mothers smoking now ($p = 0.093$), wheezing ($p = 0.001$), previous hospitalization due to asthma ($p < 0.001$), asthmatic ($p < 0.001$), family history of asthma ($p = 0.148$), and the severity of acute exacerbation ($p < 0.001$) were found to be significant in determining the patients' dose of β_2-agonist required to relieve their acute exacerbation.

Other variables did invariably inject some influence to the length of hospitalization and the reliever dose, however, of a much lesser importance. The detailed results of the univariate regression analysis are provided in Table 8 and 9.

Table 8. Univariate regression analysis result for length of hospitalisation (hours)

Variable	β	*p*	95% CI	
Age (ln), years	-0.050	0.160	-0.119	0.020
Neutrophil level (ln), x 10^9 g/L	-0.015	0.830	-0.150	0.121
Eosinophil level (ln), x 10^9 g/L	0.037	0.328	-0.038	0.113
Gender (Female)	0.219	0.002	0.082	0.357
Positive for respiratory-related virus	0.039	0.683	-0.147	0.224
Positive for HRV	0.029	0.716	-0.130	0.188
Positive for HRV-C	-0.044	0.566	-0.194	0.106
Positive for AdV	-0.783	0.006	-1.344	-0.221
Positive for RSV	0.326	0.004	0.104	0.548
Positive for InfV	-0.164	0.689	-0.965	0.640
Positive for PIV	-0.518	0.035	-0.999	-0.037
Positive for Myco	-0.520	0.327	-1.562	0.521
Positive for aero-atopy and allergy	-0.043	0.545	-0.184	0.097
Mother ever smoked	0.220	0.003	0.077	0.362
Mother smoking now	0.385	<0.001	0.206	0.565
Household member ever smoked	0.158	0.050	0.000	0.317
Household member smoking now	0.256	0.002	0.097	0.415
Coughing	-0.153	0.036	-0.296	-0.010
Wheezing	-0.106	0.302	-0.306	0.095
Previous hospitalization due to asthma	0.248	<0.001	0.110	0.386
Asthmatic	0.201	0.005	0.060	0.342
Family history of asthma	-0.031	0.739	-0.212	0.150
Severity of acute exacerbation	0.141	<0.001	0.072	0.210

Table 9. Univariate regression analysis result for β_2-agonist dose (600 mcg of salbutamol*) consumed in 24 hours

Variable	β	*p*	95% CI	
Age (ln), years	0.196	<0.001	0.108	0.284
Neutrophil level (ln), x 10^9 g/L	0.199	0.003	0.070	0.329
Eosinophil level (ln), x 10^9 g/L	0.033	0.345	-0.036	0.102
Gender (Female)	-0.081	0.205	-0.207	0.045
Positive for respiratory-related virus	0.179	0.040	0.009	0.349
Positive for HRV	0.209	0.008	0.055	0.364
Positive for HRV-C	0.094	0.189	-0.047	0.235
Positive for AdV	-0.260	0.346	-0.803	0.282
Positive for RSV	-0.371	0.003	-0.616	-0.126
Positive for InfV	-0.268	0.483	-1.020	0.483
Positive for PIV	0.170	0.586	-0.444	0.784
Positive for Myco	0.021	0.977	-1.421	1.464
Positive for aero-atopy and allergy	0.245	<0.001	0.114	0.376
Mother ever smoked	0.137	0.041	0.006	0.269
Mother smoking now	0.142	0.093	-0.024	0.307
Household member ever smoked	0.011	0.895	-0.149	0.171
Household member smoking now	0.039	0.606	-0.110	0.189
Coughing	0.033	0.620	-0.097	0.162
Wheezing	0.372	0.001	0.162	0.583
Previous hospitalization due to asthma	0.229	<0.001	0.102	0.356
Asthmatic	0.310	<0.001	0.177	0.442
Family history of asthma	0.129	0.148	-0.046	0.303
Severity of acute exacerbation	0.179	<0.001	0.115	0.243

*1 dose equals to 6 puffs of β_2-agonist, whereby each puff contains 100 mcg of salbutamol. Hence 1 dose = 600 mcg of salbutamol.

Viral infection to respiratory-related virus, aero-atopy and allergy status, asthma diagnosis status and child's severity score were significant in predicting the amount of inhaled β_2-agonist required to relieve acute exacerbation. The significant level for status of viral infection was actually higher than 0.05, however, it had been included in the final model because viral infection is deemed clinically important and might play a vital part in how a child responds to acute exacerbation. Complete results of multivariable analysis are provided in Table 10 and 11.

Table 10. Multivariable linear regression analysis result for the length of hospitalisation (hours) (n = 426)

Length of hospitalisation	**β**	***p***	**95% CI**	
Gender (Female)	0.173	0.036	0.011	0.334
Positive for RSV	0.453	<0.001	0.226	0.681
Mother smoking now	0.286	0.001	0.151	0.560
Previous hospitalisation due to asthma	0.286	0.001	0.121	0.451
Severity of acute exacerbation	0.150	<0.001	0.069	0.231

Table 11. Multivariable linear regression analysis result for β_2-agonist dose (600 mcg of salbutamol) within 24 hours (n = 428)

β_2-agonist dose consumed within 24 hours	**β**	***p***	**95% CI**	
Positive for any respiratory virus	0.159	0.067[+]	-0.011	0.329
Positive for aero-atopy and allergy	0.159	0.042	0.006	0.312
Asthmatic	0.310	<0.001	0.129	0.438
Severity of acute exacerbation	0.186	<0.001	0.116	0.255

[+]Significance level > 0.05 but has been included as it is deemed as clinically important.

The multivariable linear regression results suggested that those who were female, positive for RSV, smoking mother, been hospitalized due to asthma and had more severe exacerbation had on average between 0.15 to 0.45 more hours in hospital than their counterpart. Further, those who were positive for any respiratory virus, asthmatic, were aero-atopic/allergic and suffered a more severe exacerbation required on average 0.16 to 0.31 higher dose of β_2-agonist within 24 hours to relieve their exacerbation compared to the rest.

These final multivariable models were then fitted to determine if cases with specific genotype took longer in hospital and required higher dose of β_2-agonist in a space of 24 hours. Out of 20 genotypes tested, only one gene – GSTP1105 – that marginally reached 0.05 significance level ($p = 0.062$), in predicting the length of hospitalization. None of them was significant in predicting the dose of salbutamol consumed within 24 hours. However, this may be due to the small number of cases that got genotyped.

Clustering of admission characteristics

The results from univariate linear regression analyses were employed to select variables for the purpose of cluster analysis. The variables that were significant at the 20 percent level were deemed as potential candidates for inclusion in the cluster analysis. These variables are listed in Table 12 below.

Table 12: List of candidate variables for cluster analysis purpose

Candidate variables
1. Age
2. Gender
3. Neutrophil level
4. Viral infection to respiratory-related virus present during acute exacerbation
5. Human rhinovirus (HRV) present during acute exacerbation
6. Human rhinovirus C (HRV-C) present during acute exacerbation
7. Adenovirus (AdV) present during acute exacerbation
8. Respiratory syncytial virus (RSV) present during acute exacerbation
9. Parainfluenzae virus (PIV) type 1, 2 or 3 present during acute exacerbation
10. Aero-allergy status (encompasses both aero-atopy and allergy)
11. Child's asthma diagnosis status
12. Mother past smoking status
13. Mother current smoking status
14. Household member's past smoking status
15. Household member's current smoking status
16. Child usually have a cough
17. Child ever wheezed
18. Family history of asthma
19. Previous hospitalization due to asthma
20. Severity of acute asthma exacerbation

Based on MAVRIC's moderate size data, it was decided that k = 2 and 3 were good number of clusters to try on. One-way ANOVA and Chi-square test were conducted to see if the measures of covariates varied between phenotypes. One-way ANOVA was also conducted to see if the average length of hospitalization and β_2-agonist dose consumed within 24 hours significantly varied between the clusters. Further details of the trials are provided below.

1. First trial, k = 3. The results are presented in Table 12a and 12b.
2. Second trial, k = 2. The results are presented in Table 13a and 13b.

Table 12a. Result for cluster analysis with $k = 3$

Variable	Cluster 1 $n = 43$	2 $n = 94$	3 $n = 93$	F	χ^2	p
Gender	Female	Male	Male		2.37	0.305
Age (years)	1.03	6.11	5.25	46.29	-	<0.001
Neutrophil level	4.51	7.82	9.83	29.34	-	<0.001
Positive for respiratory-related virus infection	Yes	Yes	Yes	-	9.67	0.008
Positive for HRV	Yes	Yes	Yes	-	10.31	0.006
Positive for HRV-C	No	No	Yes	-	14.23	0.001
Positive for AdV	No	No	No	-	5.01	0.082
Positive for RSV	Yes	No	No	-	97.92	<0.001
Positive for PIV	No	No	No	-	0.83	0.659
Severity of acute exacerbation	-0.25	-0.74	0.88	131.89	-	<0.001
Mother ever smoked	Yes	No	No	-	8.06	0.018
Mother smoking now	No	No	No	-	0.45	0.799
Household member ever smoked	No	Yes	No	-	1.55	0.460
Household member smoking now	No	No	No	-	0.32	0.852
Cough	No	No	No	-	4.74	0.094
Wheeze	Yes	Yes	Yes	-	24.68	<0.001
Asthma diagnosis	No	Yes	Yes	-	82.05	<0.001
Previous hospitalization due to asthma	No	Yes	Yes	-	53.39	<0.001
Family history of asthma	Yes	Yes	Yes	-	2.75	0.253
Aero-allergy status	No	Yes	Yes	-	77.72	<0.001

Table 12b. One-way ANOVA result for clustering trial 1, *k* = 3

	Cluster	*n*	Mean	*p*
Length of hospitalization	1	43	50.25	0.065
	2	93	35.10	
	3	93	44.68	
β_2-agonist dose in 24 hours	1	23	10.48	0.005
	2	73	12.47	
	3	83	17.63	

Table 13a. Result for cluster analysis involving all possible covariates with $k = 2$

Variable	Cluster 1 $n = 41$	Cluster 2 $n = 189$	F	χ^2	p
Gender	Male	Male	-	0.52	0.470
Age (years)	0.98	5.65	84.05	-	<0.001
Neutrophil level	4.41	8.79	42.97	-	<0.001
Positive for respiratory-related virus infection	Yes	Yes	-	6.64	0.010
Positive for HRV	No	Yes	-	11.79	0.001
Positive for HRV-C	No	No	-	14.16	<0.001
Positive for AdV	No	No	-	4.95	0.026
Positive for RSV	Yes	No	-	96.96	<0.001
Positive for PIV	No	No	-	0.29	0.589
Severity of acute exacerbation	-0.30	0.08	4.76	-	0.030
Mother ever smoked	Yes	No	-	7.89	0.005
Mother smoking now	No	No	-	0.01	0.911
Household member ever smoked	Yes	No	-	0.49	0.483
Household member smoking now	No	No	-	0.22	0.642
Cough	No	No	-	2.02	0.155
Wheeze	Yes	Yes	-	24.72	<0.001
Asthma diagnosis	No	Yes	-	78.66	<0.001
Previous hospitalization due to asthma	No	Yes	-	50.77	<0.001
Family history of asthma	Yes	Yes	-	3.32	0.069
Aero-allergy status	No	Yes	-	72.45	<0.001

Table 13b. One-way ANOVA result for clustering trial 1, $k = 2$

	Cluster	n	Mean	p
Length of hospitalization	1	41	50.45	0.113
	2	188	39.96	
β_2-agonist dose in 24 hours	1	21	9.57	0.040
	2	158	15.27	

It was found that three groups *(k* = 3) yielded a better result compared to two groups *(k* = 2). The sample size got imbalanced when *k* = 2 and thus *k* = 3 was the one chosen.

The result showed that there were significantly distinct clusters memberships emerging within the identified clinical phenotypes and risk factors among the cohort. The main characteristics that differentiated the clusters were age, neutrophils level, positive for HRV, positive for RSV, wheezing, aero-atopy/allergy, previous hospitalisation due to asthma, asthma diagnosis, mother's current smoking status and family history of asthma. More details of the cluster memberships' main characteristics are provided in Table 14.

Table 14. Main characteristics of the cluster memberships

Main characteristics	Cluster 1	Cluster 2	Cluster 3
Age (years) [median (IQR)]	1 (0.2 to 1.5)	5.7 (3.1 to 8.8)	4.4 (3.1 to 7.3)
Neutrophil level (x 10^9 g/L) [median (IQR)]	3.8 (2.6 to 6.1)	8.3 (5 to 10.4)	9.4 (6.6 to 12.4)
Severity of acute exacerbation (z-score) [median (IQR)]	0.3 (-0.7 to -0.8)	-0.5 (-1 to -0.5)	0.7 (0.3 to 1.2)
Positive for HRV	51%	73%	77%
Positive for HRV-C	19%	46%	53%
Positive for RSV	65%	9%	2%
Wheezing	63%	95%	87%
Aero-atopic / allergic	5%	75%	79%
Has been previously hospitalised due to asthma	2%	69%	54%
Asthmatic	5%	84%	69%
Mother currently smokes	23%	19%	23%
Has a family history of asthma	72%	82%	84%

Last part of the cluster analysis process was to conduct a chi-square test to determine how the clusters associated with asthma-related genotypes CD14-159CT and CC16-38AG. However, the sample size was considered too small and there was no valid case for both genotypes in Cluster 1. The analysis was then terminated here and did not proceed any further.

Long-term outcomes

Summary of study participants' hospital presentations

Overall the subset of data had 533 cases observed, however due to missing observations, and/or lost hospital notes, the majority of variables had less than 533 cases in them. The average number of total number of hospital presentations post-recruitment was around 5 visits. The average age at first hospital presentation for respiratory purpose was 33 months. More details on the study participants' hospital presentations are provided in Table 15.

Table 15. Summary of study participants' hospital presentations

Variable	***n***	**median (min max)**
Age at recruitment, months	532	46 (0 to 213)
Time at risk, months	506	55 (16 to 145)
Total inpatient respiratory visits after recruitment	520	1 (0 to 17)
Total ED respiratory visits after recruitment	520	1 (0 to 17)
Total number of hospital presentation after recruitment	519	2 (0 to 147)

Both univariate and multivariable negative binomial regression analysis were used in this study. Univariate Negative Binomial regression analysis was conducted to determine the relationship of each potential covariate, including the cluster memberships, with the total number of inpatient hospital admissions and total number of ED visits. Any result from univariate analysis that was significant at 20% level was included in the multivariable analysis.

Inpatient admissions

It was found that eosinophil level (p = 0.005), positive for HRV-C (p = 0.021), positive for RSV (p = 0.009), positive for InfV (p = 0.127), positive for aero-atopy/allergy (p = 0.180), household member past smoking status (p = 0.153), household member current smoking status (p = 0.157), presence of wheeze (p = 0.001), previous hospitalization due to asthma (p = 0.008), length of hospitalization during acute exacerbation (p = 0.035), and the amount of reliever administered within 24 hours (p = 0.005) during acute exacerbation played a significant part in determining the number of total hospital presentation due to respiratory illness after recruitment. Each of this analysis was adjusted for age at recruitment to the study. The time at risk was set as the length of time the hospital presentation due to respiratory illness was observed for each child. The complete results are provided in Table 16.

Table 16. Univariate negative binomial regression analysis result for total number of inpatient visits after recruitment due to respiratory illness, adjusting for age at recruitment

Variable [median (min max)]	RR	*p*	95% CI	
Neutrophil level (ln), x 10^9 g/L	1.08	0.478	0.88	1.33
Eosinophil level (ln), x 10^9 g/L	1.15	0.005	1.04	1.28
Gender (Female)	1.12	0.256	0.92	1.35
Positive for respiratory-related virus infection	0.91	0.509	0.68	1.21
Positive for HRV	1.03	0.793	0.81	1.31
Positive for HRV-C	1.27	0.021	1.04	1.56
Positive for AdV	0.67	0.334	0.30	1.51
Positive for RSV	0.59	0.009	0.40	0.88
Positive for InfV	0.41	0.127	0.13	1.28
Positive for PIV	0.79	0.502	0.39	1.58
Positive for Myco	1.68	0.561	0.29	9.60
Positive for Aero-atopy and allergy	0.86	0.180	0.69	1.07
Mother ever smoked	1.05	0.618	0.86	1.28
Mother smoking now	0.94	0.606	0.73	1.20
Household member ever smoked	0.85	0.153	0.68	1.06
Household member smoking now	0.85	0.157	0.68	1.06
Cough	1.10	0.324	0.91	1.34
Wheeze	1.80	0.001	1.27	2.57
Previous hospitalization due to asthma	1.34	0.008	1.08	1.66
Asthmatic	0.94	0.688	0.71	1.25
Family history of asthma	0.92	0.561	0.70	1.21
Severity of acute exacerbation	1.02	0.678	0.92	1.13
Length of hospitalization during acute exacerbation	1.00	0.035	1.00	1.01
β_2-agonist dose within 24 hours of acute exacerbation	1.01	0.139	1.00	1.02

Additional analysis was also conducted to examine whether the number of inpatient visits varied between clusters. The result showed that, after adjusting for age at recruitment, there was no significant different at 20% level in the total number of inpatient hospital representations among clusters (Table 17).

Table 17. Univariate negative binomial regression analysis result for total number of inpatient visits after recruitment due to respiratory illness, using cluster memberships as covariates and adjusting for age at recruitment

Cluster memberships	**RR**	***p***	**95% CI**	
Cluster 1	(Reference group)			
Cluster 2	1.23	0.472	0.70	2.16
Cluster 3	1.15	0.629	0.66	2.00

The result from univariate analysis was used to build the best-fitted model for the purpose of multivariable analysis. All covariates that were identified as significant at 20% level at univariate level were treated as potential candidates for multivariable analysis. When fitted as a multivariable model, some of the covariates failed to keep their significance level and thus were removed.

Eosinophil level (p = 0.002), presence of wheeze (p = 0.008), positive for aero-atopy and allergy status (p = 0.007) and household-member past smoking status (p = 0.001) were found to be highly significant in influencing the recurrence of a patient's visit as an inpatient in the hospital. More detailed results are provided in Table 18.

Table 18. Multivariable Negative Binomial regression analysis result for total number of inpatient visits after recruitment due to respiratory illness, adjusting for age at recruitment (n = 261)

Variable	RR	*p*	95% CI	
Eosinophil level (ln), x 10^9 g/L	1.20	0.002	1.07	1.34
Positive for aero-atopy and allergy	0.68	0.007	0.52	0.90
Household member ever smoked	0.66	0.001	0.52	0.84
Wheeze	1.81	0.008	1.17	2.82
Previous hospitalization due to asthma	1.37	0.021	1.05	1.78

Those who with an elevated level of eosinophil, wheeze and had been hospitalized before due to asthma had between 1.2 to 1.8 times more inpatient visits compared to those with a normal level of eosinophil, did not wheeze and never been hospitalized before. Those who were aero-atopic and allergic and used to be exposed to secondhand smoking had 0.7 times less inpatient visits than the rest.

ED visits

It was found that eosinophil level (p = 0.107), female (p = 0.122), maternal past smoking status (p = 0.003), maternal current smoking status (p = 0.113), household member past smoking status (p = 0.002), household member current smoking status (p = 0.069), asthma diagnosis status (p = 0.032), and amount of reliever administered within 24 hours during acute exacerbation (p < 0.001) played a significant part in determining the number of total ED visits due to respiratory illness after recruitment. Each of this analysis was adjusted for age at recruitment to the study. The time at risk was set as the length of time the hospital presentation due to respiratory illness was observed for each child. The complete results are provided in Table 19.

Table 19. Univariate Negative Binomial regression analysis result for total number of ED visits after recruitment due to respiratory illness, adjusting for age at recruitment

Variable	RR	*p*	95% CI	
Neutrophil level (ln), x 10^9 g/L	0.89	0.534	0.62	1.28
Eosinophil level (ln), x 10^9 g/L	1.16	0.107	0.97	1.40
Gender (Female)	1.29	0.122	0.93	1.78
Positive for respiratory-related virus infection	0.91	0.681	0.58	1.43
Positive for HRV	1.04	0.858	0.71	1.52
Positive for HRV-C	1.22	0.260	0.87	1.71
Positive for AdV	1.82	0.317	0.56	5.90
Positive for RSV	0.80	0.453	0.45	1.43
Positive for InfV	0.34	0.266	0.05	2.28
Positive for PIV	1.10	0.861	0.37	3.26
Positive for Myco	2.87	0.401	0.25	33.70
Positive for Aero-atopy and allergy	0.86	0.411	0.61	1.23
Mother ever smoked	0.61	0.003	0.44	0.84
Mother smoking now	0.71	0.113	0.47	1.08
Household member ever smoked	0.57	0.002	0.40	0.82
Household member smoking now	0.71	0.069	0.49	1.03
Cough	0.98	0.911	0.71	1.35
Wheeze	1.40	0.229	0.81	2.40
Previous hospitalization due to asthma	1.02	0.894	0.71	1.47
Asthmatic	0.61	0.032	0.39	0.96
Family history of asthma	1.00	0.993	0.64	1.56
Severity of acute exacerbation	0.93	0.369	0.79	1.09
Length of hospitalization	1.00	0.912	0.99	1.00
β_2-agonist dose within 24 hours of acute exacerbation	0.95	<0.001	0.93	0.97

Additional analysis was also conducted to examine whether the number of ED visits varied between clusters. The result showed that, after adjusting for age at recruitment, there was no significant different at 20% level in the total number of ED visits among clusters (Table 20).

Table 20. Univariate Negative Binomial regression analysis result for total number of ED visits after recruitment due to respiratory illness, using cluster memberships as covariates and adjusting for age at recruitment

Cluster memberships	RR	*p*	95% CI	
Cluster 1	(Reference group)			
Cluster 2	1.22	0.704	0.44	3.36
Cluster 3	0.93	0.883	0.35	2.48

Similarly to the inpatient visits, the result from univariate analysis was used to build the best-fitted model for the purpose of multivariable analysis for ED visits. All covariates that were identified as significant at 20% level at univariate level were treated as potential candidates for multivariable analysis. When fitted as a multivariable model, some of the covariates failed to keep their significance level and thus were removed.

Household-member's smoking past and amount of reliever administered within 24 hours following an acute exacerbation were found to be highly significant in influencing the recurrence of a patient's ED visits ($p = 0.028$ and $p < 0.001$, respectively). Those who were used to be exposed to secondhand smoking and required more amount of β_2-agonist had 0.64 to 0.95 less visit to ED than the rest. More detailed results are provided in Table 21.

Table 21. Multivariable Negative Binomial regression analysis result for total number of ED visits after recruitment due to respiratory illness, adjusting for age at recruitment (n = 254)

Variable	RR	*p*	95% CI	
Household member ever smoked	0.64	0.028	0.43	0.95
β_2-agonist dose within 24 hours of acute exacerbation	0.95	<0.001	0.93	0.98

Discussion

Asthma is a multi-faceted clinical syndrome that is characterized by a high degree of phenotypes variability and severity as a result of genetic heritability and environmental stimuli (Borish & Jeffrey, 2008; Martinez & Vercelli, 2013). This heterogeneity of asthma makes the response to asthma medication varies considerably from patient to patient (Borish & Jeffrey, 2008). To date, it is unclear why some children respond differently to the same treatment. This study was designed to determine whether the lack of response to treatment was underpinned by a group of children with a particular acute wheezing phenotype.

The primary hypothesis of this study is that there is a specific group of acute wheezing phenotypes that cause children with these phenotypes not to respond well to the prescribed asthma treatment and as a result are likely to relapse and re-present to hospital. The results presented in the previous section provide support for this hypothesis and their broader meaning and context will be discussed in this section.

Response to treatment

The length of hospital stay and the amount of inhaled β_2-agonist within the first 24 hours of acute asthma exacerbation were used as indicators of how well a child is responding to treatment. The median length of stay for children in this study was 28 hours with the 25% to 75% of the stay ranged from 16.5 to 52 hours. The median dose of β_2-agonist within 24 hours of acute exacerbation was 13 emitted doses with the 25% to 75% of the dose ranged from 9 to 18 emitted doses.

Length of hospital stay

It was found that gender, respiratory syncytial virus, maternal smoking, previous hospitalization due to asthma and the severity of acute asthma exacerbation did significantly influence the length of hospital stay. In other words, it is expected that females with any of these attributes or risk factors would spend longer time in hospital compared to those without by taking about 17% longer hours to discharge.

The association between female gender and length of hospital stay could be attributed to the differences between genders in processing the treatment and tolerating the disease (Dell et al., 2001). Dell et al. (2001) used bivariate models to identify the determinants of short hospital stay among children who got admitted due to acute asthma exacerbations. They showed, consistently with the finding from this study, that younger age, male gender and mild severity of disease to have shorter stay in hospital (*Median* = 18 hours, *Range* = 6-24 hours) compared to children with the opposite attributes (*Median* = 47 hours, *Range* = 26-137 hours). Shanley et al. (2015) conducted multivariable logistic models to identify factors associated with hospitalization length of stay due to childhood asthma. They reported similar results with children who were older, females, obese and had complex chronic condition did have greater odds of a longer hospital stay with an increased odds ratio between 1.1 to 1.3 times. In contrast, Schatz et al. (2006) in their study of sex-related differences among hospitalized asthma patients found that asthma was more prevalent in male children than in female children but no gender differences in terms of asthma severity and the response to treatment. More research is needed to examine the effect of gender on asthma mechanism more closely.

In terms of virus infection, in support to the finding from this study, Guibas et al. (2012) and Papadopoulos et al. (2011) found that respiratory-related virus such as rhinovirus, respiratory syncytial virus, adenovirus and parainfluenza virus, to be associated with asthma and other wheezing phenotypes such as bronchiolitis especially in infancy and early childhood. These viruses attack the respiratory pathway and impair bronchial functionality which could precipitate severe acute exacerbations in young children (Guibas et al., 2012) that in turn would lead in turn could lead to a longer hospital stay. It is quite natural for children who had more severe exacerbation to spend longer time in hospital as they would have had worse airway inflammation that made it more difficult to reverse (Dell et al., 2001; Dougherty & Fahy, 2009). Due to the acute and episodic nature of asthma and the difficulty associated with a long-term prevention of asthma exacerbation, it is also likely that these children to have had hospital admission at least once in the past 12 months due to their susceptibility to acute asthma exacerbation (Kenyon et al., 2014).

It is not surprising that maternal smoking was found to be one of the indicators for the length of hospital stay. Tobacco smoke is a major asthma exacerbation-inducing factor both as first hand smoke in adults and as second hand smoke (SHS) in children (Guibas et al., 2012). UK Royal College of Physicians back in 2010 conducted a study to investigate the impact of passive smoking in children and found that the risk of children developing asthma doubled when they were exposed to passive smoking. Further study of exposure to SHS by Dong et al. (2011) found that passive smoking increases the presence of asthma symptoms in children younger than 2 years of age.

Dose of inhaled β_2-agonist within the first 24 hours of acute asthma exacerbation

It was found that the amount of β_2-agonist within the first 24 hours following an acute asthma exacerbation was significantly influenced by viral respiratory infection (VRI), aero-atopy and/or allergy status, asthma diagnosis status and the severity of acute asthma exacerbation. This suggests that children with acute asthma and symptoms of viral respiratory infection, who were aero-atopic and/or allergic and suffered a severe exacerbation, responded less effectively to β_2-agonist and thus required a higher dose.

The association between respiratory viruses and asthma exacerbation has been well documented in the literature. Respiratory viruses have the tendency to infect the lower airway which increases the airway inflammation and obstructs the airflow (Papadopoulos et al., 2011; Singh & Busse, 2006). This infection can both trigger and worsen asthma symptoms.

Rueter et al. (2012) conducted an asthma study to examine the influence of VRI on treatment response in acute asthma cohort. Similarly to this study, they found that the response to inhaled β_2-agonist in children with VRI symptoms within acute asthma cohort depends on whether the VRI clinical symptoms present or not; with those with VRI received a significantly higher inhaled β_2-agonists dose after 6 hours ($p = 0.010$), 12 hours ($p = 0.002$), and 24 hours ($p = 0.005$) compared to those without.

Murray et al. (2006) investigated the relationship between virus infection and allergen exposure and suggested that a synergistic interaction may exist between allergens and viruses. In other words, the combination of exposure to allergens and VRI symptoms would likely to trigger asthma exacerbation in children. This supports the finding from this study

that those children who were aero-atopic, allergic, asthmatic and with VRI symptoms tend to get more severe exacerbation and needing higher dose of inhaled β_2-agonist.

The severity of an acute asthma exacerbation affects the children's ability to respond to the inhaled β_2-agonists – in particular those administered in the first 24 hours of exacerbation. Children who had more severe exacerbation usually had worse airway inflammation that made it more difficult to reverse; this usually result in a higher dose of β_2-agonists administration (Dell et al., 2001; Guibas et al., 2012).

Cluster memberships

Cluster analysis was performed because it suits the multifaceted and multifactorial nature of asthma. It enabled the exploration of the clinical phenotypes of asthma, in particular acute asthma, and the risk factors associated with it. There was no a priori assumption regarding the cluster memberships and its constructs. A *k*-means clustering method was chosen as it best suited the study's medium-sized data, and because it is by far the simplest method that works by simply partitioning the available observations (*n*) into *k* clusters and then grouping them into a cluster with the nearest means. In doing so, *k*-means maximizes the segregation between clusters, thereby, maximizing the possibility of highly distinct clusters to emerge (Everitt et al., 2011; Haldar et al., 2008).

Few studies had used cluster analysis in the investigation of asthma phenotypes in adults; however, this study was the first in applying cluster analysis to identify asthma phenotypes in children – in particular within acute asthma cohort. Haldar et al. (2008) was the first to explore the application of *k*-means cluster analysis, to identify distinct phenotyping groups. They performed the clustering algorithm in three independent adult asthma populations

which comprised of patients with mild-moderate asthma and those with predominantly refractory asthma. They came up with two clusters that were characterized by symptoms expression such as early-onset atopic, obesity; and eosinophilic airway inflammation. Moore et al. (2009) applied an unsupervised hierarchical cluster analysis to adult population with persistent asthma. They came up with five clusters that were characterized by age, gender, atopy status, sputum induction (eosinophils and neutrophils), pulmonary function as determinant of severity, and medication use (corticosteroids and β-agonists). Schatz et al. (2014) applied a *post hoc* hierarchical clustering method with the Ward minimum variance method to children and adults with difficult-to-treat asthma population. They came up with five clusters within each age stratum; which were characterized by gender, atopy status, nonwhite race, passive smoke exposure and aspirin sensitivity. The results from the abovementioned studies as well as the current study demonstrated the possibility of utilizing a multivariable mathematical method to classify clinical phenotypes of asthma.

This study performed a cluster analysis in within children with acute asthma population. The analysis yielded three distinctive clusters memberships that were characterized by age, neutrophils level, positive for VRI, in particular HRV, HRV-C, RSV and adenovirus; the severity of acute exacerbation, maternal smoking history, wheezing, asthma diagnosis status, previous hospitalization due to asthma and aero-atopy/allergy status.

The children in cluster 1 were of the youngest age category, suffered an exacerbation of a medium level of severity, had the lowest neutrophil level and were positive for VRI, in particular HRV and RSV, suffered wheezing episodes and were non-atopic. The majority was never been hospitalized before and had never been diagnosed with asthma before. This was

mostly because this exacerbation was their first exacerbation due to their young age. The majority of them had mother who used to smoke and had family history of asthma.

The children in cluster 2 were of the oldest age category. Their neutrophil level was of a medium level and the severity of their exacerbation was the least severe level. They were mostly tested positive for viral infection especially HRV and most had wheezing episodes. In contrast with children in cluster 1, the majority of children in cluster 2 were atopic. As these children were older, most of them had previously diagnosed with asthma and been hospitalized due to asthma. The majority of them had a family member who used to smoke and had family history of asthma.

The children in cluster 3 were of the middle age category. Their neutrophil level was of the highest level and their exacerbation was of the most severe level. They were mostly tested positive for viral infection especially HRV-C and most had wheezing episodes. Similar to cluster 2, the majority of children in cluster 3 were also atopic, had previously diagnosed with asthma and been hospitalized due to asthma. In regards to family background, the majority of them had family history of asthma and no smoking history of maternal or other household member.

The clusters were significantly associated with the amount of reliever required within 24 hours of exacerbation, with the more severe cluster required higher amount of reliever. There was only a marginal difference between the three clusters in terms of the length of hospital presentation with the youngest children spent the longest time in the hospital followed by the oldest children and then by the children of mid-category age group. The results from the

aforementioned studies as well as this current study demonstrated the possibility of utilizing a multivariable mathematical method to classify clinical phenotypes of asthma.

Long-term outcome

Last part of this study was focused on determining whether the asthma phenotypes identified earlier via cluster analysis did influence the recurrence of hospital admission and ED visits after the children being recruited into the study. Slightly different phenotypes were found to play a major role in classifying which group of children that was likely to relapse and re-present to hospital as ED patients or admitted as inpatients.

Inpatient admissions

Children who had an elevated eosinophil level increased their risk of relapse and representing themselves back to hospital as inpatients by 20%. The risk increased to by about 80% for those who had wheezing episodes within the past 12 months, which was almost double the risk of those who did not wheeze. Prior hospital admission due to asthma also did significantly increase the likelihood of children to have recurring inpatient visits by about 37%. However, it was also found that the risk was reduced by about 22% to 24% for children who were aero-atopic/allergic and came from a household whose member/s used to smoke.

Elevated eosinophil level is a sign of a more serious airway inflammation, which leads to a more severe asthma (Louis et al., 2000). Thus a child with more severe asthma has a higher chance to relapse and hospital readmission (Chung et al., 2014). Wheezing is usually associated with a recurrent airway obstruction that is largely due to respiratory syncytial virus infection and can indicate low levels of lung function (Martinez, 2002; Taussig, 2002).

Simply put, wheezing is an indicator that viral infection is present and since viral infection is a major determinant of severe acute asthma exacerbation, it has an increased risk of hospital readmission as shown in this study (Wark et al., 2013). Prior hospital admission is also an important marker of whether a child is likely to be readmitted to hospital following an acute exacerbation. Several studies have found that previous asthma admission did increase the risk of readmission to about 2 to 3 fold (Auger et al., 2015; Chung et al., 2014; Li et al., 2012; Kenyon et al., 2014). A Canadian study back in 1996 (Schaubel et al., 1996) even found that the probability of hospital readmission increases with the number of previous asthma admission. It has been suggested that the reason for this is because previous asthma admission is likely to indicate higher asthma severity, poorer standard of asthma in primary care and lower access to asthma care, poor knowledge of asthma care, which would potentially lead to subsequent asthma admission (Auger et al., 2015; Chung et al., 2014).

The finding that the risk was reduced for children who were aero-atopic/allergic is in contrast with the literature. Atopy has been known as an important phenotype that makes one is prone to asthma and more susceptible to a more severe and difficult-to-treat asthma (Schatz et al., 2013). It is also a major risk factor for asthma exacerbation and interacts with viral infection that could lead to an increased risk of hospital admission and readmission in particular in HRV-C infected subjects (Cox et al., 2013). It is unclear why and how the risk was reduced by about 22% to 24% for children who were aero-atopic/allergic and came from a household whose member/s used to smoke – especially as tobacco exposure is also been named as an indirect trigger of asthma exacerbation (Chung et al., 2014; Schatz et al., 2013). Nevertheless, Auger et al. (2015) reported non-significant association between exposure to cigarette smoke and rate of hospital readmission. So there might be an important underlying reason that needs to be established by future study before any definitive conclusion can be made. It is also

possible that parents and caregivers did not report the full extent of secondhand exposure to tobacco that their children have had. As most studies rely solely on information provided by parents and caregivers, most children are not tested for exposure to secondhand smoke – i.e. no measurement on their serum and saliva for cotinine levels are made; so if there is any discrepancy between the actual exposure and self-reported ones, it goes undetected and eventually leads to a biased result (Chung et al., 2014).

ED visits

Household member's smoking history and the dose of β_2-agonist received within the first 24 hours were found to significantly reduce the likelihood of children to have recurring visit to ED by 15% and 32% respectively. It is unclear how the fact that children who lived with a smoker would have a lower risk of visiting ED due to respiratory illness compared to those who never lived with a smoker. There is a possibility that these children had gone straight to inpatient instead of visiting ED when they had exacerbation. Alternatively, similar to the inpatient visits, there is a possibility of underreporting by parents and caregivers in terms of secondhand exposure to tobacco their children have had which might have biased this result.

On the other hand, it does make perfect sense for children who received large dose of β_2-agonist within the first 24 hours after they had an acute exacerbation to have less recurring visit to ED as β_2-agonist serves to relieve the airway inflammation and improve the pulmonary function and hence less likely to relapse (Dell et al., 2001; Guibas et al., 2012; Topal et al., 2014).

Conclusion

In conclusion, this study has demonstrated the use of cluster analysis and multivariable regression technique to a clinical setting and showed how these techniques offer an alternative approach in identifying distinct asthma phenotypic groups by using the disparity in the short-term and long-term response to treatment as predictors.

Strength and limitations of the study

A major strength of this study is that this is the first study that tries to clarify asthma phenotype groups in children with severe asthma by using the data from MAVRIC study. The fact that MAVRIC study was undertaken at a single center (at Princess Margaret Hospital) also ensured the consistency of the clinical practice and medicine administration used in this study. Princess Margaret Hospital for Children is the only tertiary children hospital in WA so the study is confident that MAVRIC has captured large proportion of children that suffered acute asthma exacerbation especially of high severity ones. The large scale of MAVRIC allowed a large sample size for this study which provided adequate means for multivariable analysis. MAVRIC also collected a vast amount of information on virus, immune system, genes and immune cells were collected which enabled this study to apply a considerable measure of statistical control of some potential confounders.

A major limitation of this study is this study only focused on the severe asthma cohort. This limits the asthma phenotypes involved in the analysis and hence limits the generalization of the interpretation of the result as the predominant phenotypes found in this study might not necessarily reflect the phenotypes in the wider population.

Another limitation of the study was that virus testing was not done for all the children and the genes were typed for only a small proportion of children. This had caused some inconsistencies with the result and a significant number of missing values.

Future work in this area might benefit from a broader-based study cohort with a more comprehensive set of data to cover more asthma phenotypes and risk factors so as to reflect the population more accurately.

References

Auger, K. A., Kahn, R. S., Davis, M. M., & Simmons, J. M. (2015). Pediatric asthma readmission: asthma knowledge is not enough? *The Journal of Pediatrics*, *166*(1), 101-108.

Bentur, L., Canny, G. J., Shields, M. D., Kerem, E., Reisman, J. J., Fakhoury, K., Pedder, L., Levison, H., & Schuh, S. (1992). Controlled trial of nebulized albuterol in children younger than 2 years of age with acute asthma. *Pediatrics*, *89*(1), 133-137.

Bizzintino, J., Lee, W. M., Laing, I. A., Vang, F., Pappas, T., Zhang, G., Martin, A. C., Khoo, S. K., Cox, D. W., Geelhoed, G. C., McMinn, P. C., Goldblatt, J., Gern, J. E., & Le Souëf, P. N. (2011). Association between human rhinovirus C and severity of acute asthma in children. *European Respiratory Journal*, *37*(5), 1037-1042.

Borish, L., & Culp, J. A. (2008). Asthma: A syndrome composed of heterogeneous diseases. *Annals of Allergy, Asthma & Immunology*, *101*(1), 1-9.

Chung, H. S., Hathaway, D. K., & Lew, D. B. (2015). Risk factors associated with hospital readmission in pediatric asthma. *Journal of Pediatric Nursing*, *30*(2), 364-384.

Cox, D. W., Bizzintino, J., Ferrari, G., Khoo, S. K., Zhang, G., Whelan, S., Lee, W. M., Bochkov, Y. A., Geelhoed, G. C., Goldblatt, J., Gern, J. E., Laing, I. A. & Le Souëf, P. N. (2013). Human rhinovirus species C infection in young children with acute wheeze is

associated with increased acute respiratory hospital admissions. *American Journal of Respiratory and Critical Care Medicine*, *188*(11), 1358-1364.

Cox, D. W., & Le Souëf, P. N. (2014). Rhinovirus and the developing lung. *Paediatric Respiratory Reviews*, *15*(3), 268-274.

Dell, S. D., Parkin, P. C., & Macarthur, C. (2001). Childhood asthma admissions: determinants of short stay. *Pediatric Allergy and Immunology*, *12*(6), 327-330.

Dong, G. H., Ren, W. H., Wang, D., Yang, Z. H., Zhang, P. F., Zhao, Y. D., & He, Q. C. (2011). Exposure to secondhand tobacco smoke enhances respiratory symptoms and responses to animals in 8,819 children in kindergarten: results from 25 districts in northeast China. *Respiration*, *81*(3), 179-185.

Dougherty, R. H., & Fahy, J. V. (2009). Acute exacerbations of asthma: Epidemiology, biology and the exacerbation-prone phenotype. *Clinical & Experimental Allergy*, *39*(2), 193-202.

Drazen, J. M. (2012). Asthma: The paradox of heterogeneity. *Journal of Allergy and Clinical Immunology*, *129*(5), 1200-1201.

Duffy, D. L., Martin, N. G., Battistutta, D., Hopper, J. L., & Mathews, J. D. (1990). Genetics of asthma and hay fever in Australian twins. *American Review of Respiratory Disease*, *142*(6), 1351-1358.

Edfors-Lubs, M. L. (1971). Allergy in 7000 twin pairs. *Allergy*, *26*(4), 249-285.

Everitt, B. S., Landau, S., Leese, M. & Stahl, D. (2011). *Cluster Analysis, 5th Edition.* Chichester, UK: John Wiley & Sons, Ltd.

Guibas, G. V., Makris, M., & Papadopoulos, N. G. (2012). Acute asthma exacerbations in childhood: Risk factors, prevention and treatment, *Expert Review of Respiratory Medicine, 6*(6), 629-638.

Haldar, P., Pavord, I. D., Shaw, D. E., Berry, M. A., Thomas, M., Brightling, C. E., Wardlaw, A. J., & Green, R. H. (2008). Cluster analysis and clinical asthma phenotypes. *American Journal of Respiratory and Critical Care Medicine, 178*(3), 218-224.

Hodges, I. (1996). The aetiology of childhood asthma. *Current Paediatrics*, *6*(4), 247-251.

Jackson, D. J., Sykes, A., Mallia, P., & Johnston, S. L. (2011). Asthma exacerbations: Origin, effect, and prevention. *Journal of Allergy and Clinical Immunology*, *128*(6), 1165-1174.

Kaufman, L. R., & Rousseeuw, P.J. (1990). *Finding groups in data: An introduction to cluster analysis.* Hoboken NJ: John Wiley & Sons Inc.

Kelley, C. F., Mannino, D. M., Homa, D. M., Savage-Brown, A., & Holguin, F. (2005). Asthma phenotypes, risk factors, and measures of severity in a national sample of US children. *Pediatrics*, *115*(3), 726-731.

Kenyon, C. C., Melvin, P. R., Chiang, V. W., Elliott, M. N., Schuster, M. A., & Berry, J. G. (2014). Rehospitalization for childhood asthma: Timing, variation, and opportunities for intervention. *The Journal of Pediatrics*, *164*(2), 300-305.

Lenney, W. (2009). The aetiology of childhood asthma. *Paediatrics and Child Health*, *19*(6), 257-260.

Li, P., To, T., & Guttmann, A. (2012). Follow-up care after an emergency department visit for asthma and subsequent healthcare utilization in a universal-access healthcare system. *The Journal of Pediatrics*, *161*(2), 208-213.

Louis, R., Lau, L. C., Bron, A. O., Roldaan, A. C., Radermecker, M., & Djukanovic, R. (2000). The relationship between airways inflammation and asthma severity. *American Journal of Respiratory and Critical Care Medicine*, *161*(1), 9-16.

Martin, A. C., Laing, I. A., Khoo, S. K., Zhang, G., Rueter, K., Teoh, L., Taheri, S., Hayden, C. M., Geelhoed, G. C., Goldblatt, J. & LeSouëf, P. N. (2006). Acute asthma in children: relationships among CD14 and CC16 genotypes, plasma levels, and severity. *American Journal of Respiratory and Critical Care Medicine*, *173*(6), 617-622.

Martinez, F. D. & Vercelli, D. (2013). Asthma. *Lancet, 382*, 1360-1372.

Moore, W. C., Meyers, D. A., Wenzel, S. E., Teague, W. G., Li, H., Li, X., … Bleecker, E. R. (2010). Identification of asthma phenotypes using cluster analysis in the Severe Asthma Research Program. *American Journal of Respiratory and Critical Care Medicine*, *181*(4), 315-323.

Murray, C. S., Poletti, G., Kebadze, T., Morris, J., Woodcock, A., Johnston, S. L., & Custovic, A. (2006). Study of modifiable risk factors for asthma exacerbations: Virus infection and allergen exposure increase the risk of asthma hospital admissions in children. *Thorax, 61*(5), 376-382.

Papadopoulos, N. G., Bates, P. J., Bardin, P. G., Papi, A., Leir, S. H., Fraenkel, D. J., Meyer, J., Lackie, P. M., Sanderson, G., Holgate, S. T., & Johnston, S. L. (2000). Rhinoviruses infect the lower airways. *Journal of Infectious Diseases, 181*(6), 1875-1884.

Papadopoulos, N. G., Christodoulou, I., Rohde, G., Agache, I., Almqvist, C., Bruno, A., Bonini, S., … Zuberbier, T. (2011). Viruses and bacteria in acute asthma exacerbations – A GA2LEN-DARE* systematic review. *Allergy*, *66*(4), 458-468.

Platts-Mills, T. A. (2001). The role of immunoglobulin E in allergy and asthma. *American Journal of Respiratory and Critical Care Medicine, 164*, S1-S5.

Reddel, H. K., Taylor, D. R., Bateman, E. D., Boulet, L. P., Boushey, H. A., Busse, W. W., ... & Wenzel, S. E. (2009). An official American Thoracic Society/European Respiratory Society statement: Asthma control and exacerbations. Standardizing

endpoints for clinical asthma trials and clinical practice. *American Journal of Respiratory and Critical Care Medicine*, *180*(1), 59-99.

Rueter, K., Bizzintino, J., Martin, A. C., Zhang, G., Hayden, C. M., Geelhoed, G. C., Goldblatt, J., Laing, I. A., & Le Souëf, P. N. (2012). Symptomatic viral infection is associated with impaired response to treatment in children with acute asthma. *The Journal of Pediatrics*, *160*(1), 82-87.

Sauerbrei, W., Royston, P., & Binder, H. (2007). Selection of important variables and determination of functional form for continuous predictors in multivariable model building. *Statistics in Medicine*, *26*(30), 5512-5528.

Schatz, M., Clark, S., & Camargo, C. A. (2006). Sex differences in the presentation and course of asthma hospitalizations. *CHEST Journal*, *129*(1), 50-55.

Schatz, M., Hsu, J. W. Y., Zeiger, R. S., Chen, W., Dorenbaum, A., Chipps, B. E., & Haselkorn, T. (2014). Phenotypes determined by cluster analysis in severe or difficult-to-treat asthma. *Journal of Allergy and Clinical Immunology*, *133*(6), 1549-1556.

Schaubel, D., Johansen, H., Mao, Y., Dutta, M., & Manfreda, J. (1996). Risk of preschool asthma: incidence, hospitalization, recurrence, and readmission probability. *Journal of Asthma*, *33*(2), 97-103.

Shanley, L. A., Lin, H., & Flores, G. (2014). Factors associated with length of stay for pediatric asthma hospitalizations. *Journal of Asthma*, *52*(5), 471-477.

Sibbald, B., Horn, M. E., Brain, E. A., & Gregg, I. (1980). Genetic factors in childhood asthma. *Thorax, 35*(9), 671-674.

Sinclair, D., & Peters, S. A. (2004). The predictive value of total serum IgE for a positive allergen specific IgE result. *Journal of Clinical Pathology, 57*(9), 956-959.

Singh, A. M., & Busse, W. W. (2006). Asthma exacerbations, 2: Aetiology. *Thorax, 61*(9), 809-816.

Taussig, L. M., Wright, A. L., Holberg, C. J., Halonen, M., Morgan, W. J., & Martinez, F. D. (2003). Tucson children's respiratory study: 1980 to present. *Journal of Allergy and Clinical Immunology, 111*(4), 661-675.

Topal, E., Gücenmez, O. A., Harmancı, K., Arga, M., Derinoz, O., & Turktas, I. (2014). Potential predictors of relapse after treatment of asthma exacerbations in children. *Annals of Allergy, Asthma & Immunology, 112*(4), 361-364.

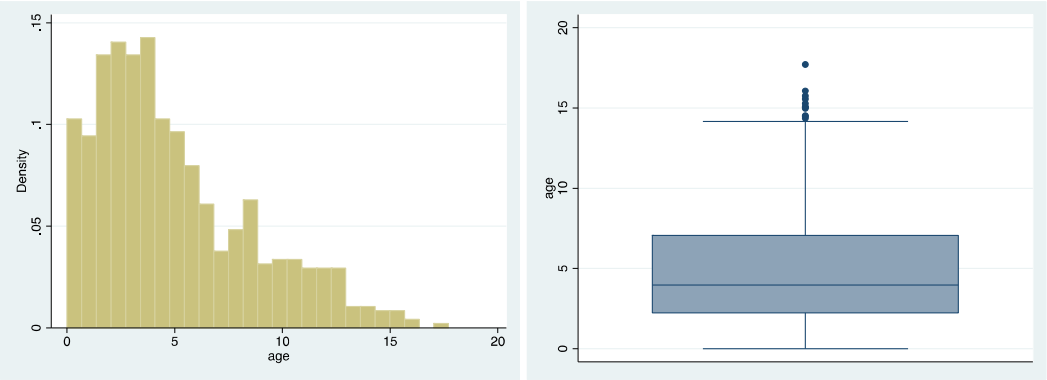
Wark, P. A., Tooze, M., Powell, H., & Parsons, K. (2013). Viral and bacterial infection in acute asthma and chronic obstructive pulmonary disease increases the risk of readmission. *Respirology, 18*(6), 996-1002.

Appendix 1

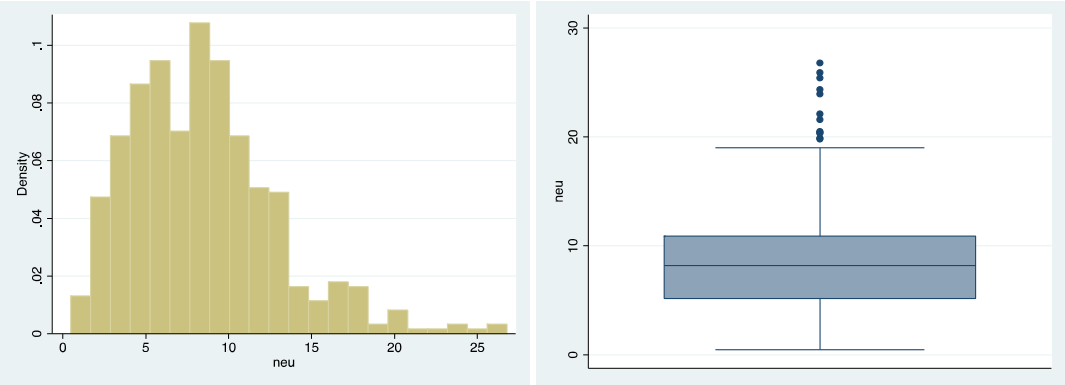
Preliminary exploratory data analysis

1. Checking the distribution of continuous variables of interest.

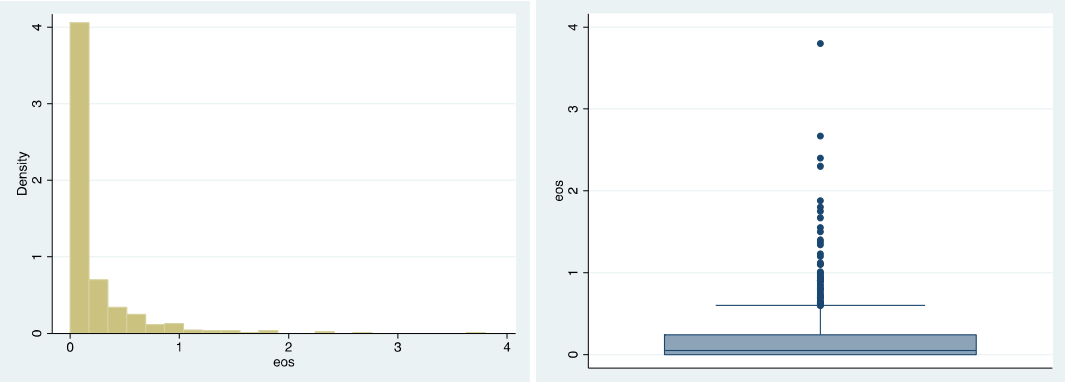
Age



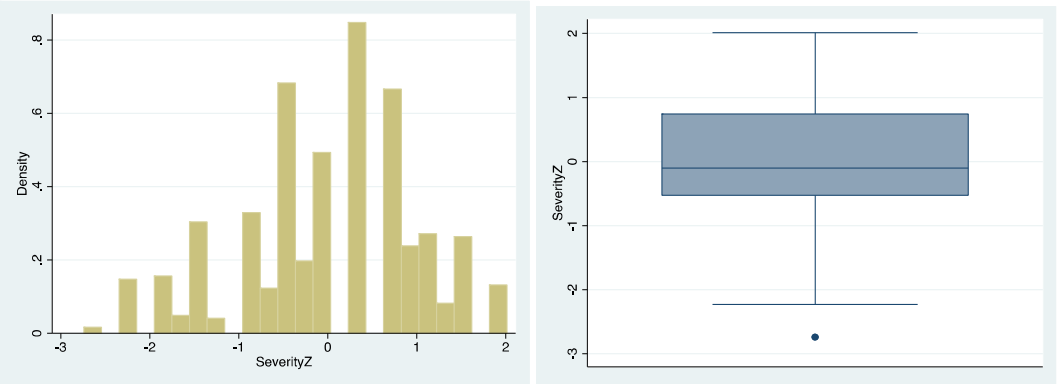
Neutrophil level



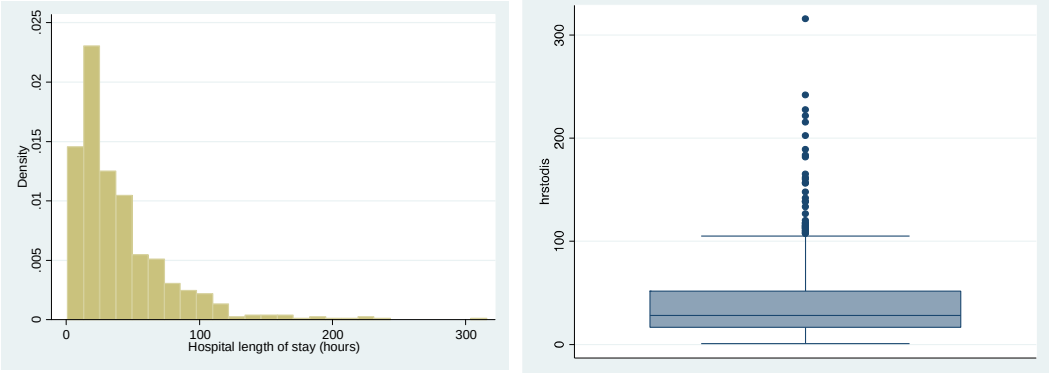
Eosinophil level



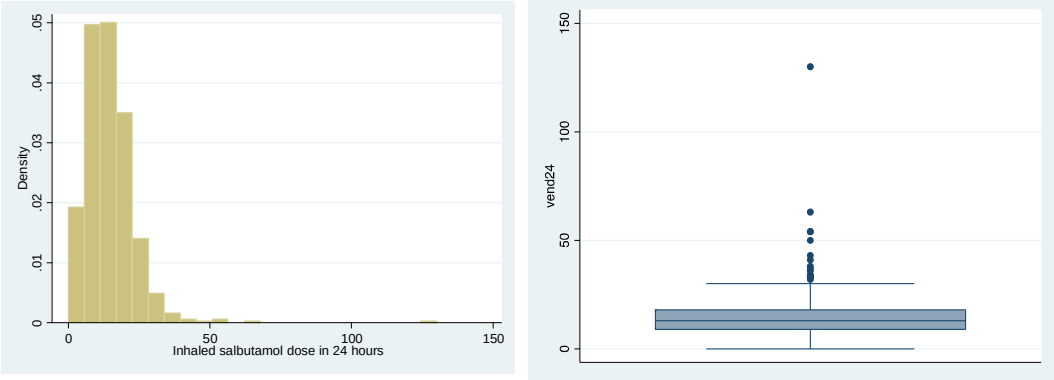
Severity of acute exacerbation (Z-Scored)



Hospital length of stay



Dose of inhaled β_2-agonist within the first 24 hours of acute asthma exacerbation

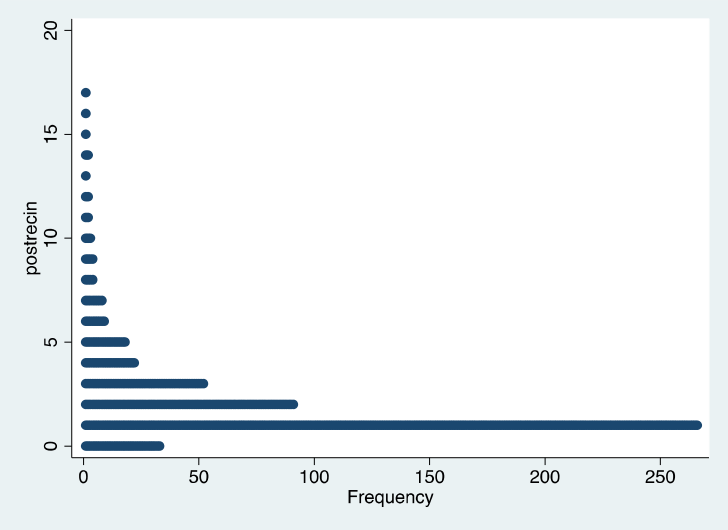


2. Checking the distribution of categorical variables

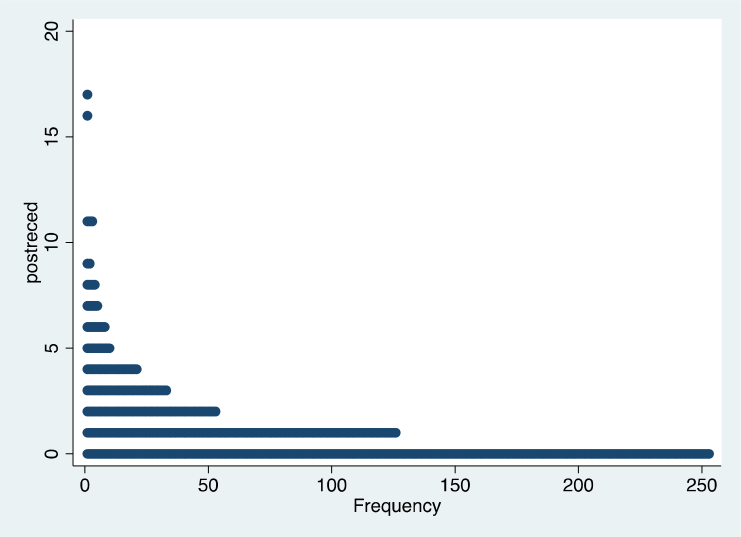
The summary of categorical data is provided in Table 6 of the main report.

3. Checking the distribution of count variables

Inpatient admissions



ED visits



4. Transformation of selected continuous variables

Age, neutrophil level, eosinophil level, length of hospitalization and the dose of inhaled β_2-agonist within the first 24 hours of acute asthma exacerbation were all (natural) log-transformed due to them were being highly skewed. The scatterplots illustrating the comparisons between the unlogged variables and logged variables are provided below (Figure 1a – 3b).

5. Extreme outliers

Patient #223 had an unusually long stay at the hospital.

Patient #226 had an unusually high dose of inhaled β_2-agonist within the first 24 hours of acute asthma exacerbation and quite a length stay at the hospital.

Patient #484 had a higher than average (of his/her study peers) eosinophils level, but it is still within the normal range (0.0 – 6.0).

ID	Age	Neutrophils level	Eosinophils level	Hospital LOS (hrs)	β_2-agonist within the first 24 hours
223	8.51	1.93	1.23	316.05	Missing
226	8.94	6.83	0	114.38	130
484	3.24	13.6	3.8	28.67	17

These outliers had been checked and were confirmed as true cases and not a data entry error and thus were included in the analysis.

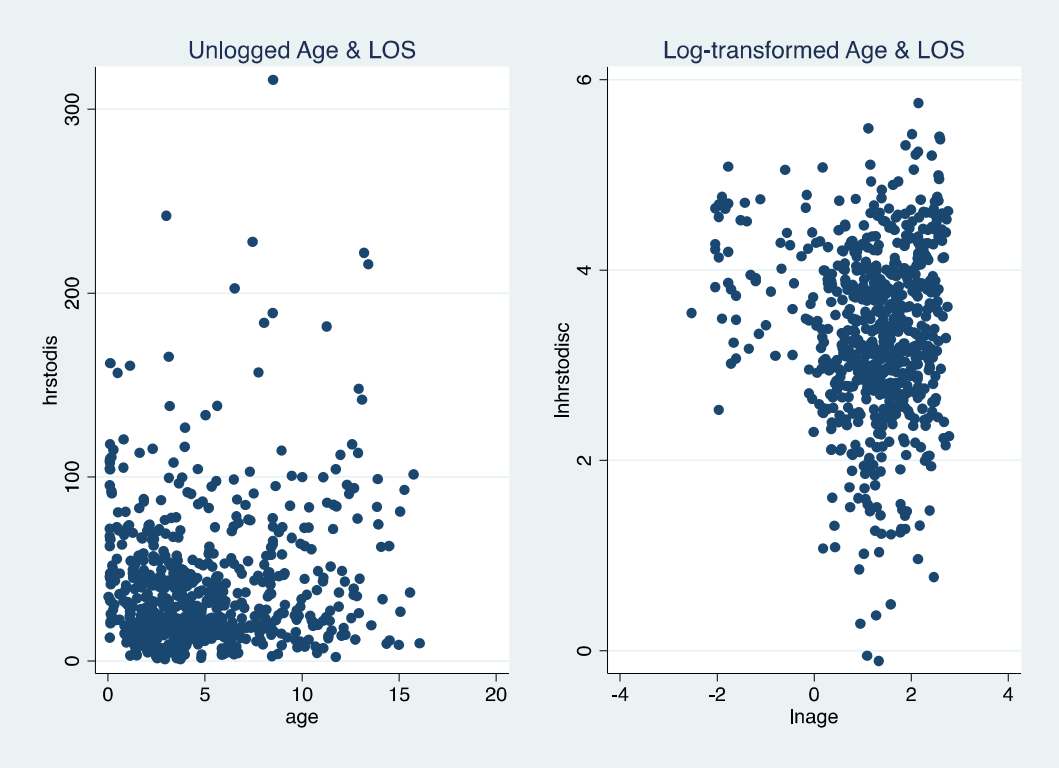


Figure 1a. Age vs. Length of hospitalization stay

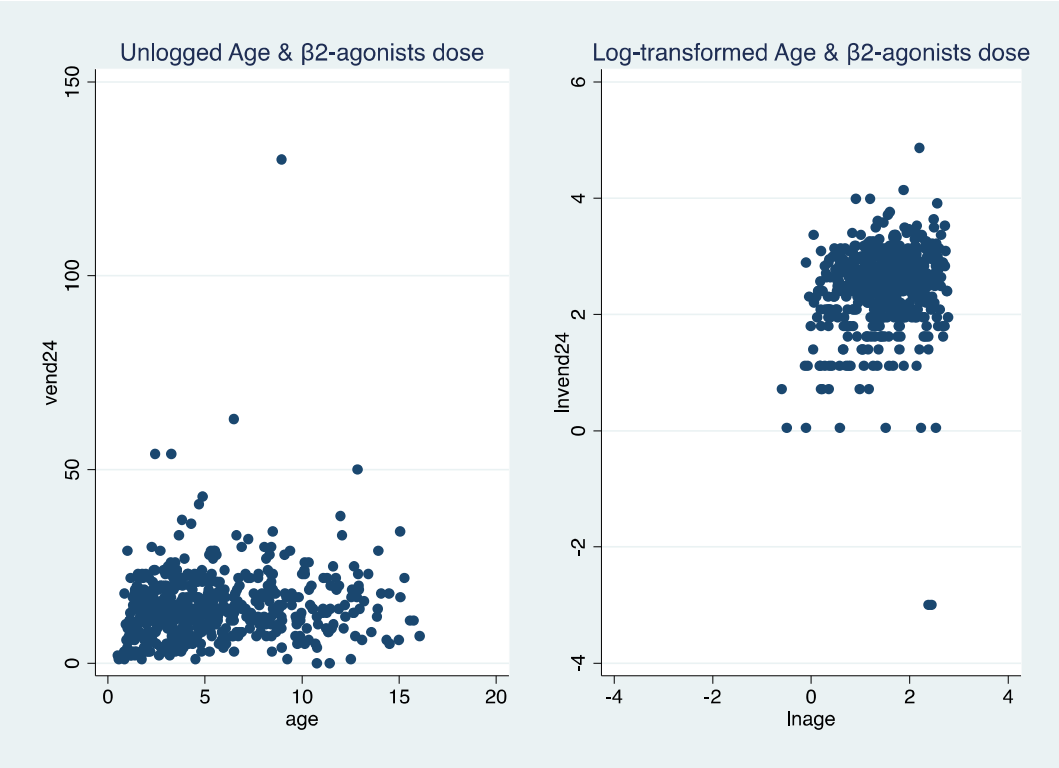


Figure 1b. Age vs. β2-agonist dose within 24 hours

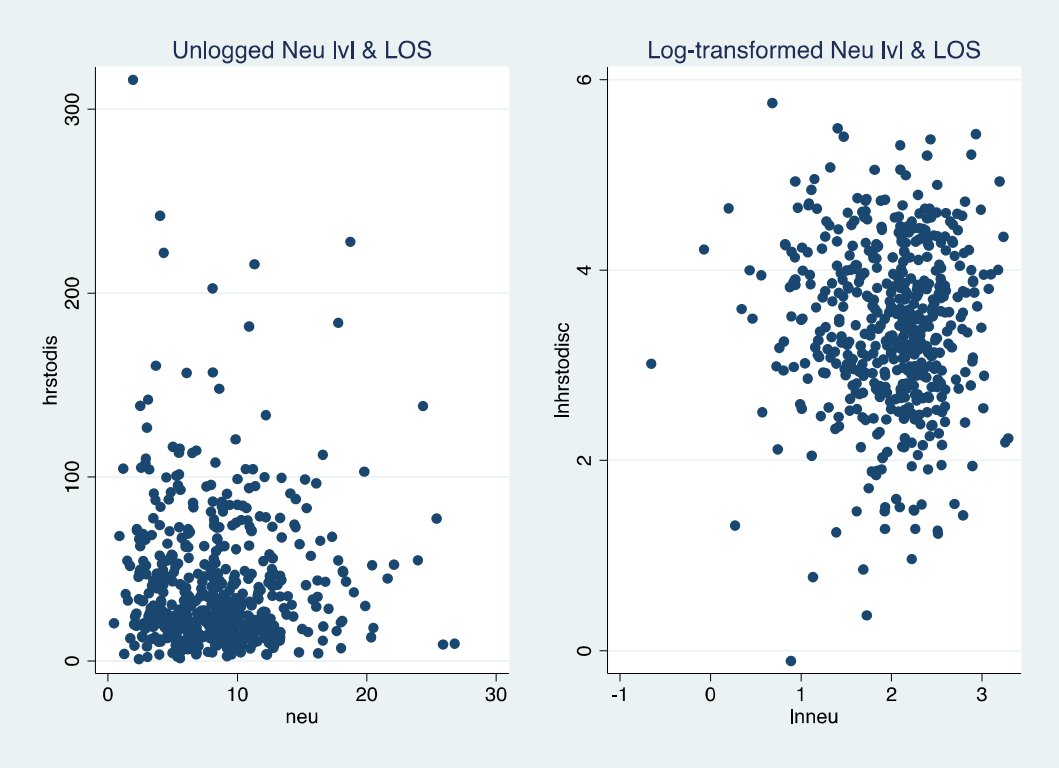


Figure 2a. Neutrophils level vs. Length of hospitalization stay

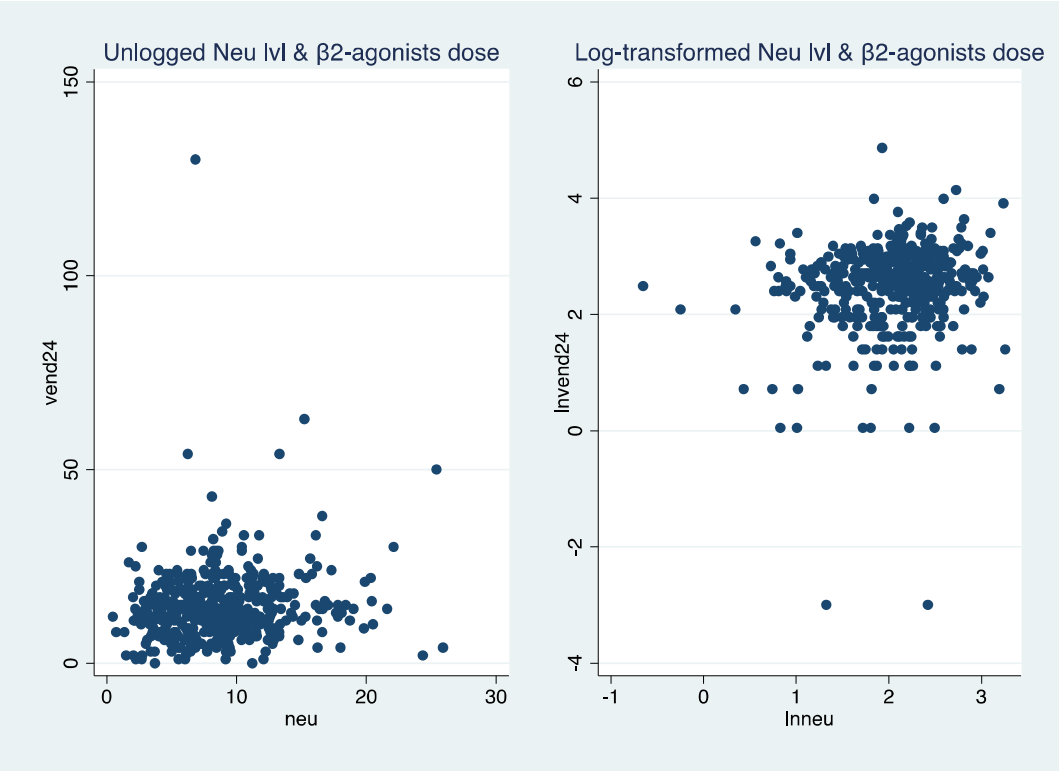


Figure 2b. Neutrophil level vs. β2-agonist dose within 24 hours

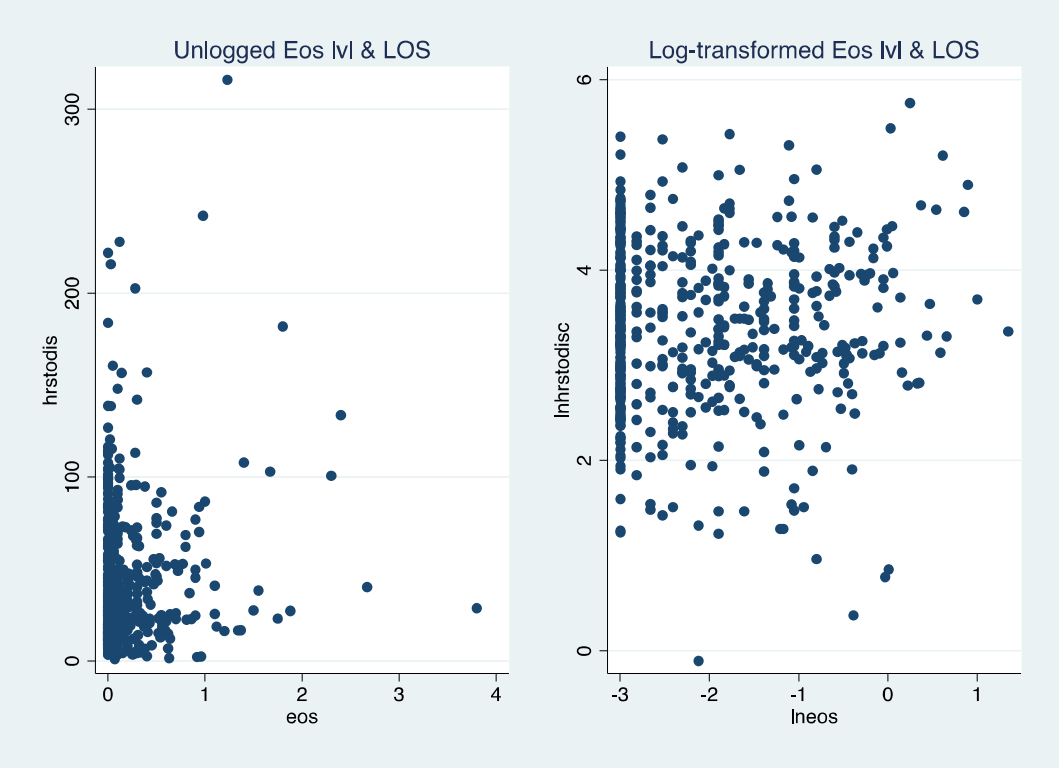


Figure 3a. Eosinophil level vs. Length of hospitalization stay

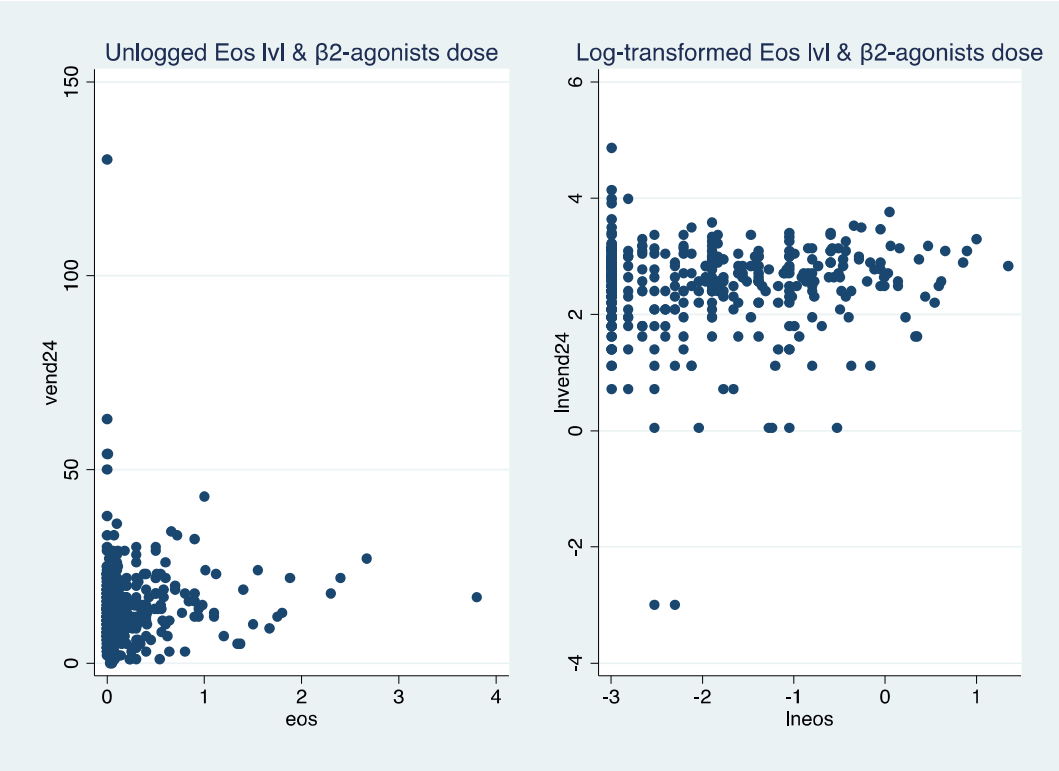


Figure 3b. Eosinophil level vs. β2-agonist dose within 24 hours

Appendix 2

Model checking

Regression diagnostics of the final model with the length of hospitalization stay as the outcome.

1. Normal probability plot

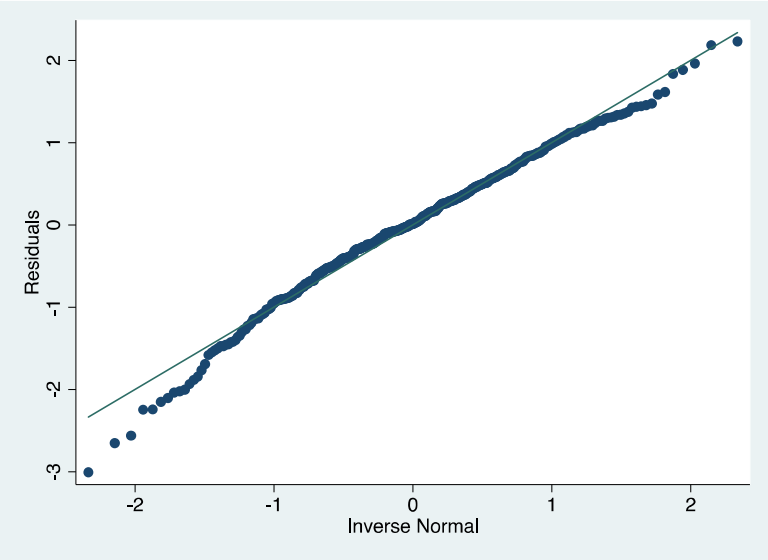


Figure 4. Normal probability plots of the residuals

2. Residuals versus the fitted values

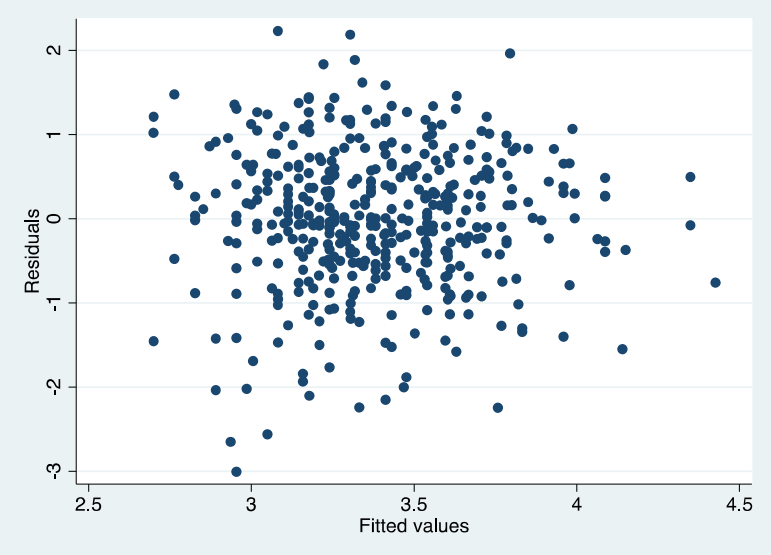


Figure 5. Plot of the residuals versus the fitted values

The assumptions of the linear regression models were tested by examining the normal probability plots of the residuals (Figure 4) and the plot of the residuals versus the fitted values (Figure 5). Both plots showed that there didn't seem to be major violations of normality of the error term in the final model, homoscedasticity of residuals or any indications of extreme non-linearity between the predictors and the outcome.

3. Collinearity diagnostics

Table 22. Collinearity diagnostics

Variable	**VIF**
Gender (Female)	1.01
Positive for RSV	1.08
Mother smoking now	1.02
Previous hospitalisation due to asthma	1.08
Severity of acute exacerbation	1.03
Mean VIF	**1.05**

There was no evidence of collinearity between the predictor variables on the evaluation of the variance inflation factors (VIF) with each of the predictor variable showed VIF value close to one (Table 22).

Based on the above results of the regression diagnostics, it was concluded that the final model was an adequate model.

Regression diagnostics of the final model with the dose of inhaled β2-agonist within the first 24 hours of acute asthma exacerbation as the outcome.

1. Normal probability plot

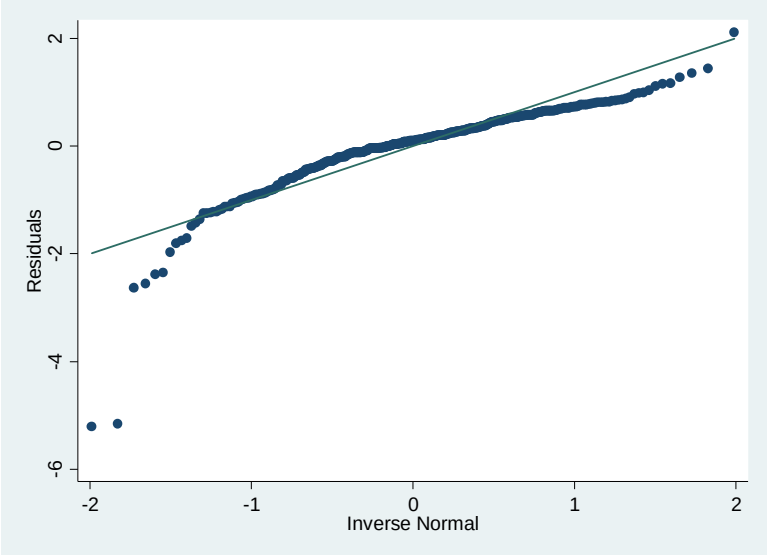


Figure 6. Normal probability plots of the residuals

2. Residuals versus fitted values

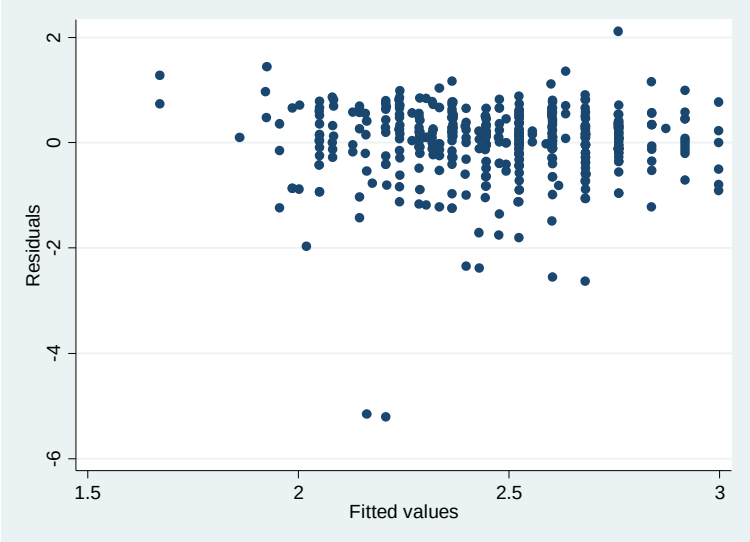


Figure 7. Plot of the residuals versus the fitted values

The assumptions of the linear regression models were tested by examining the normal probability plots of the residuals (Figure 6) and the plot of the residuals versus the fitted values (Figure 7). Both plots showed that there didn't seem to be major violations of normality of the error term in the final model, homoscedasticity of residuals or any indications of extreme non-linearity between the predictors and the outcome.

3. Collinearity diagnostics

Table 23. Collinearity diagnostics

Variable	**VIF**
Positive for any virus	1.01
Positive for aero-atopy and allergy	1.08
Asthmatic	1.01
Severity of acute exacerbation	1.01
Mean VIF	**1.05**

There was no evidence of collinearity between the predictor variables on the evaluation of the variance inflation factors (VIF) with each of the predictor variable showed VIF value close to one (Table 23).

Based on the above results of the regression diagnostics, it was concluded that the final model was an adequate model.

Regression diagnostics for Poisson versus Negative Binomial Regression Models

1. Distribution of long-term outcome

Variable	**Mean**	**Standard deviation**	**Variance**
Inpatient admissions	2.172962	2.393531	5.72899
ED visits	1.284294	2.101198	4.41503

2. Test for over-dispersion

   A null model was run for each long-term outcome and the result is summarized below.

Variable	**α (95% CI)**	***p***
Inpatient admissions	0.383 (0.300 – 0.489)	<0.001
ED visits	1.646 (1.312 – 2.066)	<0.001

The test results showed evidence that both inpatient admissions and ED visits were over-dispersed hence their distribution is not Poisson. Furthermore, Poisson distribution implies that variance equals to the mean. This assumption does not appear to hold in this instance. Thus we opted to use a negative binomial model that is more appropriate for over-dispersed count data.