Spirometric Reference Equations for First Nations Children and Adolescents Living in Rural Saskatchewan

A Thesis Submitted to the

College of Graduate Studies and Research

in Partial Fulfillment of the Requirements

for the degree of Master of Science

in the Collaborative Biostatistics Program of School of Public Health

University of Saskatchewan

Saskatoon

By

Rifat Zahan

Permission to Use

In presenting this thesis in partial fulfilment of the requirements for a Postgraduate degree from the University of Saskatchewan, I agree that the Libraries of this University may make it freely available for inspection. I further agree that permission for copying of this thesis in any manner, in whole or in part, for scholarly purposes may be granted by the professor or professors who supervised my thesis work or, in their absence, by the Head of the Department or the Dean of the College in which my thesis work was done. It is understood that any copying or publication or use of this thesis or parts thereof for financial gain shall not be allowed without my written permission. It is also understood that due recognition shall be given to me and to the University of Saskatchewan in any scholarly use which may be made of any material in my thesis.

Requests for permission to copy or to make other use of material in this thesis in whole or part should be addressed to:

Graduate Chair of the Collaborative Biostatistics Program
107 Wiggins Road
School of Public Health
University of Saskatchewan
Saskatoon, Saskatchewan
Canada
S7N 2Z4

Abstract

Background: The spirometric reference values are of great importance for diagnosis and treatment of lung diseases. At present, there are no spirometric reference values for First Nations children and adolescents living in Canada.

Objectives: The objectives of the present study were (1) to identify the flexible and efficient statistical method to derive lung function reference equations that can be used to obtain the predicted values and Lower Limit of Normal (LLN) for lung function in children and adolescents, and (2) to obtain prediction equations for FVC, FEV_1 and FEV_1/FVC for First Nations children and adolescents living in rural Saskatchewan, Canada.

Methods: Spirometric results from a prospective cohort study, "First Nations Lung Health Project" were used to identify 130 healthy non-smoking children and adolescents. The predicted values and LLN of spirometric indices [Forced Vital Capacity (FVC), Forced Expiratory Volume at one second (FEV_1) and FEV_1 and FVC ratio (FEV_1/FVC)] were calculated for school-going children and adolescents ages 6-17 years. The subjects participating in the study were from two Cree First Nations on-reserve communities located in rural Saskatchewan, Canada. All lung function values were reviewed by a respirologist for acceptability of the test.

Following an extensive literature review, the Generalized Additive Models for Location, Scale and Shape (GAMLSS) was identified as a flexible statistical tool to model the lung function variables. The lung function indices were assumed to follow a Box-Cox-Cole-Green (BCCG) distribution with median, μ, coefficient of variation, σ_L and skewness, ν. Akaike Information Criteria (AIC) approach was used to obtain the reference models. The LLN was calculated by taking the fifth percentile of the prediction equations of the lung function variables. The above approach is recommended for the prediction of lung function of multi-ethnic people aged 3-95 years from different ethnic groups by the Global Lung Function Initiative (GLI).

Results: Significant differences were observed in lung function (FVC, FEV_1 and FEV_1/FVC) and anthropometric measurements between both boys and girls. Therefore, fitting separate equations for both sexes are justified. In GLI, polynomial bases of order 6-7 were used for modeling the μ, σ_L and ν. In this study, lower order polynomial bases (up to order 4) were enough to obtain the reference models. In GLI, the polynomial bases were divided by 100 to let it lie within 0 to 1. In this study, the polynomials were divided by 20 to lie these between 0 and 1. The predicted values of FVC was higher than the values for FEV_1 in both boys and girls. Therefore the values of FEV_1/FVC ratios is less than 100% in this population. In girls, the difference between the curves of FVC and FEV_1 was smaller compared to boys. Thus, the total volume of air for girls during exhalation are close to the volume of air exhaled at the first second. The estimated curves showed that the models fitted the lung function data reasonably well.

Conclusions: The results in this study showed that the optimum model for the prediction of lung function were almost similar to the ones used by GLI for the prediction of lung function of all-age multi-ethnic populations. The predicted values and LLN values of the lung function variables reported in this study can be recommended to health-care providers for the use in diagnosis respiratory diseases in First Nations children and adolescents in rural Saskatchewan. Small sample ($n < 150$) was a limitation of this study. This study limitation can be overcome by including more individuals from the follow-up study, which will be conducted in 2016.

Acknowledgements

All praises go to Almighty Allah, the Lord and Cherisher of all the worlds, who taught humankind everything they knew not, and has given me the strength and courage to do this research work.

It gives me a great pleasure to express my sincere gratitude and deepest appreciation to my supervisors, Dr. Shahedul A. Khan, Assistant Professor, Department of Mathematics and Statistics, University of Saskatchewan (U of S) & Dr. Donna C. Rennie, Professor, College of Nursing, U of S, for their encouragement, professional guidance, and comprehensive support in my studies. Their valuable suggestions regarding literature review, report writing, hints for programming calculations, careful draft revision, helped me a great deal to complete this research in time. Without them I would not have restarted it after the worse days, and I would not have completed the work, either. I desire to express my gratitude and appreciation to my committee member, Dr. Punam Pahwa, Professor, Department of Community Health and Epidemiology, U of S, for helping me to understand the statistical part of spirometry, specially, the recent models used in the literature.

I want to acknowledge the people working for First Nations Lung Health Project (FNLHP), specially, the Principal Investigator, Dr. James A. Dosman, for giving me the permission to use their data. Special thanks to Dr. Chandima Karunanayake, Biostatistician, Canadian Center for Health and Safety in Agriculture (CCHSA) for sending me some important papers relating to my research.

I also would like to give thanks to Dr. Michael Szafron, current Program Chair, Dr. Lisa M. Lix, past Program Chair of Collaborative Biostatistics Program, U of S, Dr. Suresh Tikoo, Graduate Chair, Thesis Based Program, School of Public Health, U of S, for allowing me to go through the present research work and for providing every assistance and support whenever required.

I would like to extend my acknowledgement to First Nations Lung Health Project (FNLHP), Saskatchewan Rural Health Study (SRHS), Western Regional Training Center

(WRTC) of Health Services Research (HSR) for supporting me financially throughout my research. The research assistantship from Dr. Shahedul Khan is highly appreciated. Without these support, it was impossible for me to study here in Canada.

I would also be grateful to the First Nations Cree communities living in rural Saskatchewan for their generous participation the study. I also would like to thank librarians and the computer support provided to me by Department of Mathematics and Statistics of U of S, School of Public Health of U of S and Canadian Center for Health and Safety in Agriculture (CCHSA).

I cannot stop myself from expressing my acknowledgement to my beloved husband, Muhammad Sayedul Islam, a graduate student in the Johnson-Shoyama Graduate School (JSGS) of Public Policy at U of S. When I was overwhelmed with studies and family responsibilities, and could not concentrate on my work, he gave me lots of space, love, support and inspiration to complete this research.

Finally, I wish to express gratitude to my mother, siblings, beloved friends (Fahmida Yeasmin, Kakon Datta and Hasina Islam) and wonderful people of Quantum Foundation and Quantum Meditation Society, for their support, appreciation and help. They always inspired me during my hard times, when I was about to stop. They are the people, who truly have always believed in my dreams, as well as in me. I would not have think about standing up again without the love and courage from these people. I did not take any leave in the last three and half years to see them. They compromised for me, so that I can finish my degree.

The truth is, I took all the help I could possibly get. Without such mass co-operation, this current research would not be possible to present in time.

Rifat Zahan
Saskatoon, SK
CANADA

This thesis is dedicated to my mother

Ms. Latifa Begum (Lily)

Her unconditional love, support and sacrifices made my life beautiful and successful in every way.

Contents

List of Tables

List of Figures

List of Abbreviations

ACCP	American College of Chest Physicians
AIC	Akaike Information Criteria
ATS	American Thoracic Society
ANZSRS	Australian and New Zealand Society of Respiratory Science
APSR	Asian Pacific Society for Respirology
BF	Body Fat
BCCG	Box-Cox-Cole-Green
BCPE	Box-Cox Power Exponential
Bio-REB	Biomedical Research Ethics Board
BMI	Body Mass Index
CIHR	Canadian Institute of Health Research
COPD	Chronic Obstructive Pulmonary Disease
ECC	Expiratory Chest Circumference
ERS	European Respiratory Society
FEF	Forced Expiratory Flow
FEV_1	Forced Expiratory Volume at 1 second
FEV_1/FVC	Ratio of FEV_1 to FVC
FMF	Forced Mid Expiratory Flow
FN	First Nations
FNLHP	First Nations Lung Health Project
FVC	Forced Vital Capacity
GAM	Generalized Additive Model
GAMLSS	Generalized Additive Models for Location, Shape and Scale
GLI	Global Lung Initiative
GLM	Generalized Linear Models
HSR	Health Services Research
ICC	Inspiratory Chest Circumference
ICU	Intensive Care Unit
LMS	Lambda, Mu, Sigma
LLN	Lower Limit of Normal
MADAM	Mean and Dispersion Additive Model
NIH	National Institutes of Health
PEF	Peak Expiratory Flow
PEFR	Peak Expiratory Flow Rate
PFT	Pulmonary Function Testing
SRHS	Saskatchewan Rural Health Study
SD	Standard Deviation
TSANZ	Thoracic Society of Australia and New Zealand

ULN	Upper Limit of Normal
WRTC	Western Regional Training Centre

Chapter 1

Introduction

Respiratory diseases are one of the leading concerns of morbidity for children (Bulkow *et al.*, 2012). The prevalence of pulmonary diseases are unusually high in indigenous children (McCuskee *et al.*, 2014), and a large number of them require repeated hospitalization and admission to the pediatric intensive care unit (ICU) (Banerji *et al.*, 2001). Childhood respiratory problems are associated with chronic lung diseases in adulthood (Singleton, 2000).

Lung function is monitored using a physiological test called spirometry. The test is used to observe lung function, assess the severity of some lung diseases, and response to treatment (Karkhanis & Joshi, 2012). The test consists of measuring flow, time and volume of exhaled air (Moore, 2012). The most important measurements in pulmonary function testing are the Forced Vital Capacity (FVC), Forced Expiratory Volume at the first second (FEV_1), the FEV_1 and FVC ratio (FEV_1/FVC) and Forced Expiratory Flow between 25% and 75% of the FVC ($FEF_{25\%-75\%}$). Spirometric data along with age, height, weight and ethnicity are used to develop reference equations, based on statistical methods involving regression analyses (Veale *et al.*, 1997). Such equations are then used to predict lung function values for an individual given his/her age, sex, height, weight and ethnicity.

The ethnicity of children has an effect on lung function and has been examined by several authors (Azizi & Henry, 1994). For example, the predicted lung function values for African-American children are lower than those for Mexican-American and Caucasian children (Hsu *et al.*, 1979; Hsi *et al.*, 1983; Kirkby *et al.*, 2013). In Australia, Caucasian children have been shown to have higher lung function values compared to Aboriginal Australian children (Watson *et al.*, 1986) and children of European origins tend to have higher lung function values compared to children of Asian origin (Wesley *et al.*, 1989; Johnston *et al.*, 1987). Differences in lung function were also observed between Caucasian, Chinese and Indian populations (Yang *et al.*, 1991). It is evident that ethnic differences in lung function begin in childhood due to different physical stature between different ethnic groups (Yang

et al., 1991). For example, the reasons for predicted values between African-American children and Caucasian children in a study by Hsi *et al.* (1983) was believed to be different because of the shorter sitting heights and longer legs of African-American children compared to Caucasian children of the same standing height.

Quanjer *et al.* (2012) has derived all-age multi-ethnic reference equations that can be used globally for different ethnic groups including Caucasian, African-American, North-East Asian, South-East Asian and Others (people with mixed ethnicity). The authors considered an adjustment for ethnicity while modeling lung function indices. Because of potential differences in physical statures in different ethnic groups, lung function reference equations that are ethnic-specific can be found in the literature.

Some studies on developing spirometric reference equations for Canadians have been conducted so far (Gutierrez *et al.*, 2004; Tan *et al.*, 2011; Karunanayake *et al.*, 2015). All studies were aimed at deriving lung function prediction equations for Caucasian adults living in Canada. Spirometric reference equations for non-First Nations Canadian population may not be useful for First Nations people, as the equations could largely depend on ethnicity. Moreover, the equations for adults may not be useful for children, as lung function increases with age until adulthood and starts to decline with age (Moore, 2012). At present, there is no specific spirometric reference equation available for First Nations people in Canada, including children. This research focuses on developing research questions for First Nations children and adolescents and could fill a gap in spirometry of Canadian First Nations children and adolescents.

1.1 First Nations Children and Adolescents in Canada

According to the National Household Survey (NHS) conducted by Statistics Canada (2011), Aboriginal children ages 14 and less represents 28.0% of the total Aboriginal population and 7% of all Canadian children. First Nations people represent 60.8% of the total Aboriginal population that also includes Inuit and Métis. In Saskatchewan, 38.1% of the First Nations people are ages 14 and less, representing 20.0% of all children in this province (Statistics Canada, 2011).

Health inequalities exist between Aboriginal and non-Aboriginal Canadians (Estey *et al.*, 2007; Wilson *et al.*, 2010). According to the First Nations and Inuit Regional Health Survey (FNIRHS), a significant proportion of Aboriginal children have bronchitis, asthma,

wheeze, ear infections and over-weight problems (MacMillan *et al.*, 2010). Sin *et al.* (2004) conducted a pilot study to assess the prevalence of impaired lung function and asthma in school-going First Nations children. Children living in a rural First Nations reserve located in Northern Alberta were selected for this study. Spirometry tests were performed following parental/guardian consent or child assent. Lung function values were obtained for FVC and FEV_1 of 36 children. The spirometric reference equations derived from the Caucasian population in the United States (Hankinson *et al.*, 1999) were used for First Nations children living in the study. Observed lung function values were compared with the Lower Limit of Normal (LLN) (Hankinson *et al.*, 1999) for the assessment of airflow obstruction. Sin *et al.* (2004) found that 25% of the First Nations children in their study showed evidence of airflow obstruction. First Nations children also had frequent reports of asthma, compared to non-First Nations children (Sin *et al.*, 2004). Sin *et al.* (2004) suggested that asthma is under-diagnosed and under-recognized for First Nations children.

Smoking, which can affect lung health, is prevalent in First Nations communities. The prevalence of daily smoking is higher in mothers of Aboriginal children compared to non-Aboriginal mothers (Gao *et al.*, 2008). According to the study conducted by Sin *et al.* (2004), 73.1% children are exposed to indirect smoke in their households.

Lung function varies with ethnicity along with other demographic characteristics (age, height, sex) (Quanjer *et al.*, 2012); therefore, lung function prediction equations derived from non-First Nations children may not apply to First Nations children. As mentioned earlier, GLI was led by Ph H Quanjer (Quanjer *et al.*, 2012) to develop lung function prediction equations for different ethnic groups globally. Although the research group developed equations for most people from different ethnic backgrounds, they excluded people of mixed-ethnicity, North American Indians or Aboriginal people living in different parts of the world. To accommodate these groups, Quanjer *et al.* (2012) took the average of the equations derived for Caucasian, African-American, North East Asian and South East Asian; and reported the results as spirometric reference values for other populations.

Currently there is an ongoing project on lung health being conducted in Saskatchewan, Canada. The First Nations Lung Health Project (FNLHP)- is a prospective cohort study being conducted in two First Nations communities situated in rural Saskatchewan, Canada (Pahwa *et al.*, 2015). The FNLHP is actively working on potential determinants associated with respiratory outcomes in First Nations peoples. This thesis is a part of the project that is aimed at deriving lung function prediction equations for First Nations. The results from this study will provide the normal values for lung function that can be used to assess respiratory

health in First Nations children and adolescents with more accuracy.

1.2 Spirometry

This section describes key concepts related to spirometry theory, i.e., lung function testing, the use of spirometric reference equations in the calculation of normal values to diagnose lung diseases or abnormalities.

1.2.1 Spirometric Indices

The most common parameters/indices measured in spirometry are forced vital capacity (FVC), forced expiratory volume at the first second (FEV_1), the ratio of FEV_1 and FVC (FEV_1/FVC) and forced expiratory flow (FEF). These parameters are used to assess obstruction or restriction in lung function. *Obstruction* represents airflow limitation in lungs (Moore, 2012). Asthma and COPD are obstructive diseases. *Restrictive disorders* (or restriction) represents a loss of lung volume (Moore, 2012). For example, restriction occurs in pleural disease, chest wall disorder, obesity, pulmonary oedema (Moore, 2012). A brief description of FVC, FEV_1, FEV_1/FVC and $FEF_{25\%-75\%}$ are as follows

- **Forced Vital Capacity (FVC)**. FVC is the total amount (volume) of air that can forcibly blow out as fast as possible after full inspiration (Moore, 2012). FVC is measured in liters. Normal or reduced value of FVC indicates obstructive diseases, whereas, a reduced FVC means restrictive diseases (Moore, 2012).

- **Forced Expiratory Volume (FEV_1)**. FEV_1 measures how much air can be exhaled during a forced breath at the first second (Moore, 2012). This index is also measured in liters. A reduced value of FEV_1 means an obstructive disease (Moore, 2012).

- **Ratio of FEV_1 and FVC (FEV_1/FVC)**. The FEV_1/FVC, also known as Tiffeneau-Pinelli index (Yao *et al.*, 2013) ratio is used in the diagnosis of restrictive and obstructive lung diseases (Swanney *et al.*, 2008; Sahebjami & Gartside, 1996). It represents the proportion (or percentage) of a person's vital capacity that they can expire in the first second of exhalation. In healthy people, this should be above the lower limit of normal (LLN) (Quanjer *et al.*, 2012) (see Section 1.2.2). An FEV_1/FVC below the LLN indicates an obstruction, whereas restriction is characterized by normal-to-high FEV_1/FVC value (Moore, 2012).

- **Forced Expiratory Flow (FEF)**. The FEF is usually expressed as a percentage of vital capacity. The $FEF_{25\%-75\%}$ is also known as *maximal mid-expiratory flow* (Koopman *et al.*, 2011), which is the average flow from the time 25% of the FVC has been exhaled to the time 75% of the FVC has been exhaled (Moore, 2012). Recent research suggests that $FEF_{25\%-75\%}$ or $FEF_{25\%-50\%}$ may be a more sensitive measure than other lung function indices in the detection of obstructive small airway diseases (Simon *et al.*, 2010; Ciprandi & Cirillo, 2010). However, use of this measure in pediatric lung function testing is controversial for children (White, 1994). Coates *et al.* (2013) reported that $FEF_{25\%-75\%}$ depends on FVC and has a high degree of variability; therefore, interpretation of this index in children requires experience.

1.2.2 Spirometric Reference Equations and Lower Limit of Normal (LLN)

Spirometric reference equations are constructed using the lung function parameters of individuals based on their ethnicity and demographic characteristics such as age, height, weight, sex (Quanjer *et al.*, 2012). These equations are developed using statistical techniques involving regression models. Only healthy individuals who are non-smokers are used as a reference to construct these equations. This normal healthy state is then compared with the lung function values of an individual's test results to assess his/her lung function. Thus, deriving appropriate reference values are crucial for interpreting pulmonary function tests and for assessing the lung function and respiratory diseases.

In research and clinical medicine the Lower Limit of Normal (LLN) of lung function is defined as the 5^{th} percentile of a healthy population (Culver, 2012). For example, if a variable follows a normal distribution, the 5^{th} percentile (or LLN) is equivalent to mean - 1.645 $\times$ standard deviation. When the variable follows distribution other than normal, the LLN can be calculated as the 5^{th} percentile of that particular distribution. The observations for healthy individuals are assumed to lie beyond the LLN. This approach for calculating the LLN for spirometric indices are available in the documents of American Thoracic Society/ European Respiratory Society (ATS/ERS) (ATS, 1991; Clausen *et al.*, 1980; Crapo *et al.*, 1981; Pallegrino *et al.*, 2005).

1.3 Lung Function Testing in Young Children

Lung function testing procedures and evaluation in children, particularly in younger children can be different from adults (Coates *et al.*, 2013). Since this testing takes considerable effort, it is helpful to have bright and pleasant environment while conducting the lung function testing in children. Children vary in size and the size of the mouthpieces used need to be considered (Seed *et al.*, 2012). A child size mouthpiece may be required if a child is very young or with a child who has cranio-facial abnormalities (Seed *et al.*, 2012). If the child has missing teeth, particularly the front upper and lower central and lateral incisors, the technician will need to pay close attention on mouth closure when the children is forcibly exhaling during testing. Noseclips are necessary for children so that they do not breath air in/out through their nose during testing. Moreover, the children should have adjustable chairs, so that it allows them to sit straight with both feet planted on the floor or stool during the testing procedure (Coates *et al.*, 2013).

For the acceptability of any lung function test, certain criteria should be met, which are recommended by American Thoracic Society (ATS)/ European Respiratory Society (ERS) (Miller *et al.*, 2005). In children, these criteria sometimes are overriden. For example, the recommendation by ATS/ERS that the minimum exhalation time during the test is 6 seconds for adults is reduced to 3 seconds for children ages less than 10 years. However, if a child finishes exhalation in less than 2 seconds, it may become difficult to resist inhaling before the technician can end the test (Coates *et al.*, 2013). There is another challenge in inhalation manoeuvres for children. In some procedures, the ATS/ERS recommendation is to blow all air out of the lungs and to have a rapid inhalation with no breath hold at the beginning of the inhalation. When the technologist asks a child to inhale, there is a possibility that the child may hold the breath at total lung capacity (Coates *et al.*, 2013).

The test repeatability criteria set by ATS/ERS is that if the observed value of FVC (or FEV_1) is less than 1 litre and the differences between two successive FVC (or FEV_1) is within 100mL, then the test is repeatable (Miller *et al.*, 2005). Similarly, if the observed value of FVC (or FEV_1) is greater than 1 litre and the differences between two successive FVC (or FEV_1) is within 150mL; the test is repeatable (Seed *et al.*, 2012). Repeatability is often not a problem with children because of the greater consistency in lung volumes on repeated blows. Most problems are with technique and length of forcible blows.

Spirometry in children should require special skills for technologists to obtain optimal and useful test results for lung function (Seed *et al.*, 2012). For the current study all research

assistants conducting the spirometry were certified in spirometry by the Lung Association of Saskatchewan.

1.4 Research Objectives and Questions for the Study

The two objectives of the present study are

1. to identify flexible and efficient statistical method to derive lung function prediction equations and LLN for children and adolescents, and

2. to obtain prediction equations for FVC, FEV_1 and FEV_1/FVC for First Nations children and adolescents living in rural Saskatchewan, Canada.

   The objectives of this study lead to two broad research questions, which are,

1. **What are the available spirometric reference equations of lung function variables for children and adolescents for different ethnic groups?**
   (a) What type of statistical methods are used to model FVC, FEV_1 and FEV_1/FVC?
   (b) What are the ages of the children and adolescents considered for prediction of lung function?
   (c) Which variables are considered as predictor(s) to model the lung function indices?
   (d) Which statistical method or formula are used to calculate the LLN?
   (e) Which approach is the best one for prediction of lung function variables?

2. **What would be the reference equations for the prediction of lung function indices (FVC, FEV_1, FEV_1/FVC) for First Nations children and adolescents, living in rural Saskatchewan?**
   (a) What data and variables are considered for the study?
   (b) Which approach/model are considered to model lung function indices?
   (c) How can the optimum models be chosen?
   (d) What is the LLN for each of the spirometric indices?

The first objective of this study was achieved by an extensive literature review on spirometric reference equations and LLN of children and adolescents. Following the literature review, the best approach identified in the first objective was applied to the modeling of spirometric prediction equations and LLN for First Nations children and adolescents living in rural Saskatchewan, Canada.

The next section gives an overall idea of the organization of the study.

1.5 Organization of the study

This study is organized into five chapters:

In Chapter 1, an introduction to spirometric reference equations, a brief description of the respiratory health status of First Nations population and some basic terms used in spirometry is discussed.

In Chapter 2, an extensive literature review is conducted to achieve the first objective of this study. The focus is on the available reference equations used for children and adolescents. Chapter 2 also shows how the spirometric reference equations revolve around regression models, ranging from the simplest form of a simple linear regression model to the more flexible method of generalized additive models for location, scale and shape.

In Chapter 3, a description and estimation method is provided based on the models used for prediction of lung function values for First Nations school-going children and adolescents living in rural Saskatchewan, Canada.

Chapter 4 provides a description of the study design, analysis of the data, application of the model to obtain the spirometric reference values, selection of the best equations for First Nations children and adolescents, calculation of the LLN and interpretations of the results.

In Chapter 5, a brief discussion of the answers to the research questions is provided. The strength, limitations, further scope and conclusion of the study is also given in Chapter 5.

Chapter 2

Literature Review

One of the objectives of this study is to identify a flexible and efficient statistical method to derive lung function prediction equations and LLN for children and adolescents. With this objective in mind, an extensive literature review is presented in this chapter to provide a thorough understanding of the theory of spirometry, and to address the research questions for Objective 1 presented in Section 1.4. This will also help to identify gaps, if any, in the theory of spirometry, both in the contexts of scientific content and statistical method. The scientific concern is the implications of the anthropometric variables to predict lung function values for different ethnic groups, whereas the statistical concern is the use of an appropriate method to predict lung function values.

An overview of the scientific context for spirometric reference equations are summarized by Kory *et al.* (1961), Cole (1975) and Wang *et al.* (1993). As reported in these articles, lung functionality of children changes with their growth. Wang *et al.* (1993) identified growth spurt as an important indicator to understand the status of lung health. In particular, standing height was estimated to be the most important predictor for children's lung function growth. They also reported that the lung function values largely depend on age: FEV_1 and FVC grow linearly until adolescence, generally reach to the peak between the ages of 20 and 30 after which a slow decline is observed throughout adulthood. However, they observed differences in growth spurts between boys and girls: the growth spurts for girls were generally smaller than those for the boys. In fact, since the work by Kory *et al.* (1961), it has become a standard practice to compare the distributions of lung function values in different cohorts, summarized by age, sex and height. Therefore, a common practice is to develop prediction equations by using regression models which relate the predictors like age and height to the lung function values.

Statistical techniques to develop lung function prediction equations revolve around linear regression, ranging from the simplest form of simple linear regression to the more

flexible method of generalized additive models for location, shape and scale. We may broadly classify the statistical methods that are used to generate spirometric reference equations into three categories: (1) linear regression, (2) polynomial regression, and (3) generalized additive models using smooth functions to handle curved relationship. Some significant works on spirometric reference equations based on these three methods are reviewed in Sections 2.1, 2.2 and 2.3, respectively. In the light of the literature review, a discussion about the scientific and the statistical contents is presented in Section 2.4, with the ultimate goal to address the first objective of this study.

2.1 Linear Regression in Spirometry

Many authors considered linear regression to develop spirometric reference equations. Let y_i represents the value of the response variable on the i^{th} subject, and $x_{1i}, x_{2i}, \ldots, x_{pi}$ represent the subject's values on p explanatory variables (i.e., predictors), with $i = 1, 2, \ldots, n$. The linear regression model can be expressed as

$$y_i = \beta_0 + \beta_1 x_{1i} + \ldots + \beta_p x_{pi} + \epsilon_i, \tag{2.1}$$

where $\beta_0, \beta_1, \ldots, \beta_p$ are the unknown regression coefficients and ϵ_i is the random error component. For purposes of testing hypotheses and calculating confidence intervals, it is assumed that the errors are independent and identically distributed as normal with mean zero and constant variance σ^2, that is, $\epsilon_i \sim N(0, \sigma^2)$. The model with only one predictor (i.e., $p = 1$) is commonly known as the simple linear regression model. In the theory of spirometry, one of the lung function indices (e.g., FEV_1) is taken as the response variable, and the anthropometric measurements (e.g., age, height, weight, abdominal girth, BMI, ethnicity, sex, etc.) are considered as the explanatory variables. Linear regression model (2.1) is then fitted to the observed data to predict the response (i.e., the lung function values) given the measurements of the explanatory variables. Transformation of variables [e.g., log (FEV_1), log (height)] are often used to obtain a better fit of the model and/or to remedy the violation of the assumptions underlying the regression model (i.e., linearity, normality and homoscedasticity).

Several studies were conducted in North America to predict lung function values for children and adolescents based on model (2.1). Dickman *et al.* (1971) analyzed lung function data for 482 healthy boys and 468 healthy girls ages 5-18 years to predict FVC and FEV_1.

The sample comprised of subjects from Salt Lake County, Utah, who had no history of respiratory disease, asthma or other chronic lung diseases. For subjects with height < 60 inches, the values of the spirometric measurements were very similar for boys and girls, whereas noticeable differences were observed for taller subjects. For this reason, separate analyses were conducted for subjects grouped by height and gender. Simple linear regression was used for the shorter individuals with height as the sole predictor, whereas multiple linear regression was used for the taller individuals with both height and age as predictors. The authors reported that the lung function values increased until the age of 16 and then became steady for girls. On the other hand, the lung function values for boys increased dramatically at adolescence, peaked at the age of 18 and then started to decrease. In another study, Hsu *et al.* (1979) used the simple linear regression model with logarithmic transformations for both height and lung function indices. The study population consisted of 1,805 healthy Mexican-American, Caucasians, and African-Americans between 7 and 20 years of age from six public schools of Houston, Texas. The lower limit of normal (LLN) was calculated using

$$\text{LLN} = \text{Lung Function}_{\text{Predicted}} \times (1 - \text{Standard Deviation}_{\text{Error}})^2. \tag{2.2}$$

Their results showed that only 2.5% individuals had lung function values less than LLN. Accounting for ethnic differences, the standing height was found to be the most important covariate. The predicted lung function values for the African-American children were generally lower than those for the non-African American children. Wall *et al.* (1982) derived lung function reference equations for 176 North American Indian healthy children (94 girls and 82 boys) ages 7-18 years from the town of Warm Springs in the United States. Simple linear regression with logarithmic transformations for both height and lung function values was used for prediction. They found that the lung function predicted values for these children were higher than those reported by Hsu *et al.* (1979) for Mexican-American and African-American children. Since the ancestral background of the Indians of Warm Springs reserve is very similar to the Indians of Pacific Northwest, the authors recommended to use the proposed reference equations for the American Indian children living in these regions. Coultas *et al.* (1988) conducted a population-based survey in New Mexico Hispanic community to develop spirometric reference equations for 576 children and adolescents ages 6-18 years. The authors suggested that their prediction equations could be used for the Hispanics in New Mexico and southern Colorado as their ancestries were comparable to those of the study subjects.

A few studies on spirometric reference equations were conducted for children and adolescents of European origin. Bjure (1963) analyzed lung function data on 161 healthy Scandinavian boys and girls between 7 and 17 years of age. Fitting simple linear regression model led to the conclusion that the predicted lung function values for boys were higher than those for the girls. Cotes *et al.* (1979) used height, fat-free mass and body fat to predict lung function values for 254 healthy British boys and girls between 8 and 16 years of age. Separate models were considered for sitting height and standing height. The inclusion of body fat and fat-free mass in the analyses helped to reduce the variability due to error and gender differences, respectively. In another study, Roizin *et al.* (1993) considered lung function data from 753 second and third generations of Israeli children of different ethnic groups (European, Iraqi, North African, Indian, Yemenite and Georgian). Children ages 7-14 years who had no history of chest or heart diseases, wheezing, chest surgery, spinal deformity, concurrent upper or lower airway infection and smoking, and whose parents were from the same ethnic group were considered in the final analyses. The final sample consisted of 471 children from six different ethnic groups. Separate analyses were conducted for FVC, FEV_1, and the ratio FEV_1/FVC. The authors concluded that only standing height could be used to predict the lung function values for all but the Georgian and the Indian ethnic groups. Roizin *et al.* (1993) also concluded that there was substantial variability in lung function values in different ethnic groups, and such variability diminished after childhood. They suggested to use different prediction equations for Indian and Georgian children. Piccioni *et al.* (2007) performed spirometry on a sample of 960 healthy children ages 3 - 6 years from Turin, Italy. They used age, sex, standing height, weight and body mass index (BMI) of these children to predict their lung function values.

Until the study of Veale *et al.* (1997), a little was known about the spirometry of Australian Aboriginal people. Veale *et al.* (1997) aimed at developing normal range for spirometric indices for this particular ethnic group. They conducted a cross-sectional population-based study of four rural Aboriginal communities from Queensland, Northern Territory and South Australia. There were 261 healthy children between 7 and 19 years of age and 332 healthy adults between 20 and 80 years of age. Multiple linear regression with covariates age, standing height, abdominal girth and the interaction between age and abdominal girth was used to derive prediction equations for FEV_1 and FVC. The findings of the study were compared with those reported by Gore *et al.* (1995) for Australian Caucasians. Veale *et al.* (1997) concluded that their equations were more accurate for prediction of lung function values compared to all previously published equations for Australian children and adults. The

rural Australian Aboriginals exhibited low FEV_1 and FVC compared to those of Caucasians. As a result, Australian Caucasians had relatively higher values for FEV_1/FVC ratios.

Several studies on lung function prediction equations for African children and adolescents can be found in the literature. Shamssain *et al.* (1988) conducted a study for children and adolescents between 6 and 19 years of age in Libya. They proposed to consider only age and standing height to predict FVC and FEV_1 for these children, and reported significantly higher predicted values for the boys than those for the girls. Weight was not found significant in predicting the lung function values. Although reference values were recommended for the Libyan children, they suggested to conduct an analysis of the lung function values for other parts of Africa. In fact, Shamssain (1991) in a subsequent study considered 2,000 non-smoking healthy African school-going children and adolescents between 6 and 19 years of age. The study was conducted in Umtata in the Republic of Transkei in Southern Africa. This study consisted of the largest number of African children compared to any other previous studies (Schoenberg *et al.*, 1978; Dockery *et al.*, 1983; Miller *et al.*, 1977; Huizinga & Glanville, 1968). Only height was reported to be significantly associated with lung function values for both boys and girls. The findings of this study also indicated that the forced expiratory indices for Africans were lower than those for the United States counterparts. The predicted lung function values obtained in this study were recommended to use for the South African school-going children and adolescents. Bougrida *et al.* (2012) stressed the need of spirometric reference values for children living in Constantine. Subsequently, data on lung function values (FVC, FEV_1, FEV_1/FVC, $FEF_{25\%}$, $FEF_{50\%}$ and $FEF_{75\%}$) and covariates (age, height, weight, BMI and body surface area) were obtained from 208 (107 boys and 101 girls) healthy children and adolescents ages 5 - 15 years. Bougrida *et al.* (2012) applied multiple linear regression to predict lung function values using all these covariates. The estimated equations were recommended for Constantine children and adolescents ages 5 - 15 years.

People living in India come from diverse ethnic groups, and therefore considerable variation in spirometric reference equations is observed in the literature. Chowgule *et al.* (1995) predicted lung function values (FVC, FEV_1 and $FEF_{25\%-75\%}$) for 632 healthy children and adolescents ages 6-15 years from Bombay, a city in the western coast of India. Separate analyses for boys and girls were carried out, with the conclusion that age and weight were not significantly associated with lung functionality, whereas height was the most important covariate to predict the lung function values. Vijayan *et al.* (2000) obtained information on pulmonary function tests for 246 boys and 223 girls between 7 and 19 years of age from the

southern part of India. Covariate data on sex, age, ethnicity, standing height, weight and smoking status were also collected for these children. Use of the multiple linear regression model suggested only height and weight to be significantly associated with the lung function values. Budhiraja *et al.* (2010) analyzed lung function data for healthy boys and girls ages 6 - 15 years from the district of Ludhiana, located in northern part of India. Age, height, weight and sex were taken as covariates to predict lung function values. Substantial differences were observed between the findings of this study and some of the previously published studies in India, including Chowgule *et al.* (1995), Vijayan *et al.* (2000) and Rajkapoor *et al.* (1997). Budhiraja *et al.* (2010) mentioned that environmental factors and regional variations could have contributed to such differences. Doctor *et al.* (2010) estimated reference values for FEV_1 and FVC using data from 655 healthy boys and girls (ages 8 - 14 years) living in south Gujarat, western India. Age, height, weight, body surface area and sex were used to predict lung function indices. Choudhuri & Sutradhar (2015) estimated reference equations for adolescents from Tripura, located in the northeastern part of India. The study population consisted of 640 healthy non-smoking adolescents between 10 and 14 years of age (320 ethnic tribal and 320 non-ethnic Bengali). The authors recommended multiple linear regression model, using weight, BMI, waist-to-hip ratio and waist-to-height ratio to predict lung function values for this study population.

Burity *et al.* (2013) selected 135 healthy preschool children (ages 3 - 6 years) from the metropolitan areas of Brazil. Spirometric data were obtained on FVC, FEV_1, $FEV_{0.5}$, $FEF_{25\%-75\%}$, FEV_1/FVC, $FEV_{0.5}$/FVC and $FEF_{25\%-75\%}$/FVC. There were Caucasians, African and mixed-race participants in the study. For boys, age and height were used to predict FVC, FEV_1 and $FEV_{0.5}$. However, these lung function indices were better explained using weight and height as covariates for the girls. On the other hand, $FEF_{25\%-75\%}$, FEV_1/FVC, $FEF_{25\%-75\%}$/FVC and $FEV_{0.5}$/FVC were predicted using simple linear regression with height being the only covariate.

One study was conducted by Miller *et al.* (1977) with Jamaican children, including African, Afro-American and European descents. The sample consisted of 54 urban children (29 boys and 25 girls) and 54 rural children (30 boys and 24 girls) ages 7-14 years. Multiple linear regression was used to predict the logarithmic transformations of FVC and FEV_1 using sex, ethnicity and the logarithmic transformations of age and height. Miller *et al.* (1977) found that (1) the average FVC was 3% higher for children living in the rural areas compared to those living in the urban areas, (2) the average FVC was 7% lower for the girls compared to those for the boys, and (3) the average FVC was 16% higher in children of European

descents compared to those of African origin.

In summary, the main points of interest for spirometric reference equations based on linear models are described as follows.

- Linear regression is the most popular statistical method to develop spirometric reference equations (Quanjer *et al.*, 1995), perhaps because of its simplicity and ease of interpretation.

- Transformations of variables are often necessary to obtain a better fit of the model to the observed data.

- A spirometric reference equation cannot be generalized for all ethnic groups. This is because the diversity in social, environmental and economic factors often leads to substantial variability in the lung functions, as well the predictors which describe the lung function values (Trabelsi *et al.*, 2008).

- Lung function values can differ substantially by gender; therefore, the usual recommendation is to construct separate equations for boys and girls.

- In many studies, standing height was identified as the most important covariate to predict lung function values.

- In general, the lung function values do not change linearly over time, and therefore many authors developed separate reference equations for children stratified by age.

2.2 Polynomial Regression in Spirometry

In many cases, the dependent variable in a regression setting might have a non-linear relationship with an independent variable. For example, FEV_1 often exhibits a curved relationship with age: an increasing trend before puberty, followed by a sudden increase during puberty and a decreasing trend thereafter (Moore, 2012). This type of curved relationship between a lung function measure and a predictor (e.g., age) can be modelled more appropriately using the polynomial regression model. Mathematically, a polynomial regression model of order p in one independent variable x (e.g., age) can be expressed as

$$y_i = \beta_0 + \beta_1 x_i + \beta_2 x_i^2 + \ldots + \beta_p x_i^p + \epsilon_i, \tag{2.3}$$

where y_i is the response (e.g., FEV_1) for the i^{th} individual, $\beta_0, \beta_1, \ldots, \beta_p$ are the regression coefficients and ϵ_i is the error component. How large an order of polynomial model to consider depends on the problem being studied and type of data being collected. For example, a first-order model (i.e., $p = 1$, leading to simple linear regression model) has no bends; a second-order model has no more than one bend, and each higher order term adds another potential bend. Extension of (2.3) by incorporating other predictors which are linearly related with the dependent variable is straightforward. For example, suppose the relationship between y and x_1 is linear, whereas the relationship between y and x_2 is parabolic (i.e., $p = 2$). The model can be written as

$$y_i = \beta_0 + \beta_1 x_{1i} + \beta_2 x_{2i} + \beta_3 x_{2i}^2 + \epsilon_i. \tag{2.4}$$

Polynomial models are special cases of the general multiple regression model (2.1). For example, if we denote x_1, x_2 and x_2^2 by z_1, z_2 and z_3, respectively, then Equation (2.4) becomes

$$y_i = \beta_0 + \beta_1 z_{1i} + \beta_2 z_{2i} + \beta_3 z_{3i} + \epsilon_i, \tag{2.5}$$

which is simply a multiple linear regression model with three independent variables.

Many authors used polynomial regression to develop spirometric reference equations. For example, Rosenthal *et al.* (1993) analyzed a dataset consisting of 772 Caucasian children and adolescents (455 males and 317 females) between 4.6 to 18.8 years of age living in the United Kingdom. Predicted values for FEV_1, FVC and FEV_1/FVC were obtained using the polynomial regression model. The authors recommended to use a polynomial of order five in height to predict the lung function values. In another study, Parma *et al.* (1996) analyzed data from 897 Caucasian boys between the ages of 7 and 18 years from the city of Rome, Italy, to derive spirometric reference equations. The authors used weight, BMI, inspiratory chest circumference (ICC), expiratory chest circumference (ECC), the increment in chest and a polynomial of order two on age to predict the logarithm of the lung function indices.

2.3 Generalized Additive Models in Spirometry

A family of regression models which can more flexibly characterize the curved relationship is the so called generalized additive models (GAM) (Hastie & Tibshirani, 1986, 1990). This family includes models for different types of response variables such as categorical and continuous. The basic idea is to use a flexible smooth function of a covariate to model the curved relationship. One of the special cases of GAM is the classical normal regression model (2.1) which assumes independent and identically distributed normal errors with mean zero and constant variance σ^2. Suppose that a flexible representation of the covariate x_i is necessary to capture the nonlinearity. Under the normality assumption of the error components, the GAM can be expressed as

$$y_i = \beta_0 + \beta_1 x_{1i} + \cdots + \beta_p x_{pi} + f(x_i) + \epsilon_i, \tag{2.6}$$

where $f(\cdot)$ is a smooth function of x. As described in Section 2.2, $f(\cdot)$ can be defined using the polynomial model (or polynomial basis), and therefore polynomial regression is also a special case of GAM. Another very flexible function is the cubic spline, made up of sections of cubic polynomial, joined together so that they are continuous in value as well as first and second derivatives. The points at which the sections join are called the *knots* of the spline. One of the appealing features of the spline regression is that they can be shown to have good theoretical properties. The cubic spline can also be defined using other basis functions rather than the polynomials. Example include *natural spline* and *B-spline*; Different types of basis functions can be found in Wood (2006).

The regression spline mechanism offers an amenable way to model compound dependence of lung function on two growth parameters, height and age. Wypij *et al.* (1993) analyzed data collected from 5,030 Caucasian boys of ages 10 to 18 years living in the United States. The study subjects were followed for 15 years to observe the changes in their lung function indices. The authors considered GAM to predict $\text{FEF}_{25\%-75\%}$ using the polynomial spline on age and an interaction between this spline and the logarithmic transformation of height. Hankinson *et al.* (1999) selected a sample of 7,429 lifelong non-smoking individuals from the third National Health and Nutrition Examination Survey (NHANES III, 1996) conducted during 1988-1994. They first formed four groups classified by age and sex, and then used the second order polynomial basis to derive lung function prediction equations for each group. These equations were recommended for prediction of spirometric

indices for the population of the United States, including Caucasians, African-Americans and Mexican-Americans.

Until the study conducted by Stanojevic *et al.* (2008), the equations derived by Hankinson *et al.* (1999) had been used for the population of the United States. However, Hankinson *et al.*'s findings were limited to people over 8 years of age. Stanojevic *et al.* (2008) considered a sample of 3,598 non-Hispanic Caucasian subjects of 4 to 80 years of age living in the United States. In addition to a curved relationship between the lung function values and age and height, the authors observed the response variable to exhibit non-uniform dispersion around the mean (i.e., non-constant variance), as well as skew in the shape of the distribution (i.e., potential violation of the normal assumption). The so-called LMS (lambda-mu-sigma or λ-μ-σ) method proposed by Cole & Green (1992) uses the ideas of GAM but does not require a specific distributional assumption such as normality. Therefore, this class of models is more flexible in the sense that it can deal with skewed distributions. The basic idea is to transform the response variable with the goal of obtaining a new variable which is approximately normal. The LMS method is based on the Box-Cox transformation which involves three parameters: (1) λ to correct skewness, (2) μ, the median of the response variable, and (3) σ_L, the coefficient of variation which is a standardized measure for dispersion. LMS method is also known as the Box-Cox-Cole Green (BCCG) model (Cole & Green, 1992) or the generalized additive models for location, shape and scale (GAMLSS). The estimated equations proposed by Stanojevic *et al.* (2008) were based on the LMS method, and were recommended for the prediction of lung function values for the non-Hispanic white subjects between 4 and 80 years of age living in the United States. In another study, Koopman *et al.* (2011) used the LMS method to analyze spirometric data from 1,042 healthy children and adolescents between 4 and 18 years of age living in Utrecht, the Netherlands.

Researchers stressed the need of all-age multi-ethnic prediction equations that can be used globally. With this motivation, a large number of centers have been sharing lung function data with Philip H. Quanjer (a researcher in physiology) since 2006. Later the Global Lung Function Initiative (GLI) was established in Berlin in September 2008, which subsequently acquired the European Respiratory Society (ERS) Task Force status in April 2010. The GLI was authorized for developing reference equations by several respiratory societies, including American Thoracic Society (ATS), Australian and New Zealand Society of Respiratory Science (ANZSRS), Asian Pacific Society for Respirology (APSR), the Thoracic Society of Australia and New Zealand (TSANZ) and American College of Chest Physicians (ACCP). The Task Force derived spirometric prediction equations using data from 74,187

healthy individuals from 72 centers in 33 countries. These equations were recommended to use for different ethnic groups (Caucasians, African-Americans, North-East Asians and South-East Asians) covered in this study. The work of Quanjer *et al.* (2012) is perhaps the most influential work in spirometry theory to date, conducted under this task force. They derived spirometric reference equations and the LLN of lung function indices for people between 3 and 95 years of age coming from different ethnic groups, including Caucasians, African-Americans and North and South East Asians. The equations were derived using the LMS method as described in Stanojevic *et al.* (2008). Specifically, the following model was considered to predict lung function indices for both males and females:

$$y_i = \beta_0 + \beta_1 \text{Height}_i + \beta_2 \text{Age}_i + \beta_3 \text{Ethnicity}_i + \beta_4 (\text{Age}_i \times \text{Ethnicity}_i) + f(\text{Age}_\text{i}) + \epsilon_i, \quad (2.7)$$

where y_i is the response (lung function value with or without log transformation) for the i^{th} individual, and $f(\text{Age})$ is the polynomial spline on age. Depending on the type of data, Quanjer *et al.* (2012) suggested to use the logarithmic transformation of the response and/or one or more independent variables to obtain a better fit of the model. Note that model (2.7) enables a smooth connection of the lung function values among children, adolescents and adults. Some of the important findings of this widely acceptable work are summarized below.

- The FEV_1/FVC ratio is independent of the ethnic groups.
- FVC and FEV_1 differ between the Caucasian people and other ethnic groups.
- People with mixed ethnic origins have different lung function values compared to the rest of the study population. One recommendation was to predict the lung function values for people with mixed ethnic origins by taking the average of the predicted values for other ethnic groups.
- The LLN is calculated as the 5^{th} percentile of the BCCG distribution, which can be expressed as $\mu(1 - 1.645\lambda\sigma_L)^{1/\lambda}$ (see Chapter 3 for detail).

Although the work of Quanjer *et al.* (2012) covers several ethnic groups, the proposed method may not be applicable for each and every ethnicities around the globe. For example, Rochat *et al.* (2013) analyzed lung function data for the central European populations, consisting of 118,891 individuals between 8 and 90 years of age with 51% of them female. The authors realized that a more general form of the model that takes into account both skewness

and kurtosis would have a better fit to the observed data. For this, they recommended to use GAM based on Box-Cox Power Exponential (BCPE) (Rigby & Stasinopoulos, 2004) transformation of the lung function values.

2.4 Conclusion

Researchers have shown that the choice of prediction equations can have substantial impact on the clinical interpretations of the spirometry results (Rosenfeld *et al.*, 2001; Subbarao *et al.*, 2004). Therefore, rather than depending on the default lung function reference values of commercial spirometer, a sophisticated choice should be made for different ethnic groups (Pittman & Rosenfeld, 2011). The following scientific facts revealed through our literature review might be of importance to the researchers.

- Anthropometric variables such as height and age can be used to predict lung function values quite accurately.
- Prediction equations are derived from lung function data of healthy individuals.
- There is no unique reference equation to use for all ethnic groups. This is because there could be substantial variability in the lung function values and anthropometric variables among people of different ethnic origins.
- Height is the most important predictor of lung function values.
- Lung function values can differ substantially by gender; therefore, the usual recommendation is to construct separate equations for males and females.
- Lung function values change over time, and therefore the effects of age should be taken into account while deriving prediction equations.

Statistical methods to derive prediction equations are primarily based on regression models. Depending on the nature of the data, different regression models are recommended. Linear regression is used when the association between the lung function values and each of the predictors can be reasonably approximated by a straight-line relationship. However, lung function values often exhibit a curved relationship with age. Polynomial regression and GAM are usually preferred in such situations to characterize a curved relationship. GAM

has become more popular recently because of its flexibility and attractive theoretical properties. For skewed lung function data, the GAMLSS method is more appropriate to develop prediction equations, as this method does not require a specific distributional assumption for the response.

No work has been done so far on spirometry reference equations for First Nations children and adolescents of Canada. As discussed in the Chapter 1, the prevalence of respiratory diseases is high in Indigenous children, and therefore reference equations for other ethnic groups may not be appropriate for them. In this study, we develop lung function prediction equations for First Nations children and adolescents. We will use the GAMLSS method because of its flexibility of modelling different types of data. Our findings will provide a better understanding and assessment of lung functionality of First Nations children and adolescents of Canada.

Chapter 3

Statistical Methods

This chapter focuses on introducing the Generalized Additive Model for Location, Scale, and Shape (GAMLSS). GAMLSS will be used to analyze the lung function data for the First Nations children and adolescents living in rural Saskatchewan, Canada. As discussed in Chapter 2, the GAMLSS is used for the all-age multi-ethnic people to predict their lung function indices and to define the lower limit of normal values. In Section 3.1, an introduction to GAMLSS will be described. In Section 3.2, a nonlinear semiparametric additive GAMLSS will be described, which is used for all-age multi-ethnic population for modeling the lung function indices. In this chapter, the estimation procedure of GAMLSS and how the reference equations was chosen based on the model comparison criteria will be discussed. In Section 3.3, the method for calculating the lower limit of normal of the lung function indices will be described.

3.1 Introduction to Generalized Additive Model for Location, Scale and Shape (GAMLSS)

Linear regression analysis is one of the most commonly used statistical method for modeling the relationship between a response (dependent) variable and explanatory (independent) variables. Linear regression modeling assumes that errors are independently and identically distributed with zero mean and constant variance. More recently, Generalized Additive Models (GAMs) (Hastie & Tibshirani, 1986, 1990) and Generalized Linear Models (GLMs) (Nelder & Wedderburn, 1972) have become more popular. In GAM and GLM, the normal distribution of response variable Y_i is replaced by an exponential family of distribution; and a link function μ_i (the mean of y_i) to the linear predictor, X_i. GAMLSS are extension of GAM and GLM, which replaces the exponential family distribution by general distribution family

that can model both skewness and kurtosis. GAMLSS uses the concept of LMS method (L (λ)-skewness; M (μ)-mean; S (σ_L)-coefficient of variation) (Cole, 1988), which is a popular and highly cited technique for age-varying reference ranges (Cole *et al.*, 2009) for skewed data. In other words, this method models the location, shape and scale, simultaneously, that can capture the age-varying changes in the response variables. GAMLSS were first introduced by Rigby & Stasinopoulos (2001, 2005); Akantziliotou *et al.* (2002) and was adopted by many authors to construct growth references in many countries (Cole, 1998; Fredriks *et al.*, 2000; Kuczmarski *et al.*, 2002). GAMLSS are semi-parametric regression models, as they require parametric distribution assumption for the dependent variable. And some non-parametric smoothing functions are required while estimating the parameters of the distribution (which are function of the independent variables). In GAMLSS, the distribution of the dependent variable does not have to belong to the exponential family; GAMLSS can cover highly skewed and kurtotic continuous and discrete distributions (Rigby & Stasinopoulos, 2010).

Some of different sub-models of GAMLSS are: semi parametric additive, parametric linear, non-linear semi-parametric additive, and non-linear parametric. The most popular form of GAMLSS, to predict lung function indices is non-linear semiparametric additive GAMLSS, which will be described in section 3.2.

3.2 Estimation Technique of GAMLSS

GAMLSS is a framework for modelling the response variable (e.g., FVC, FEV_1 or FEV_1/FVC) following a wide range of family of distributions, which may depend non-linearly on covariates (e.g., age, height, weight, abdominal girth, etc.).

Let us consider the following matrices,

$$\boldsymbol{Y} = \begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{pmatrix}, \quad \boldsymbol{X} = \begin{pmatrix} x_1, x_2, \ldots, x_n \end{pmatrix}, \quad \boldsymbol{\theta} = \begin{pmatrix} \theta_1, \theta_2, \theta_3, \theta_4 \end{pmatrix}$$

Let (x_i, y_i), for $i = 1, 2, \ldots, n$ be the covariate and response for individual i. We can assume that the distribution of $\boldsymbol{Y}$ depends on parameters $\boldsymbol{\theta}$, such that $\boldsymbol{\theta}$ can be modeled as a function of covariate and the response variable. The first two parameters of $\boldsymbol{\theta}$ are the location and scale. The remaining parameters are shape parameters, i.e., skewness and kurtosis. Here

we are emphasizing on a popular special case of GAMLSS, the LMS model (Cole & Green, 1992). For $\boldsymbol{Y} > 0$, $\boldsymbol{Y}|\boldsymbol{X}$ follows a Box-Cox Cole and Green (BCCG) distribution (Box & Cox, 1964) with parameters $[\theta_1(x), \theta_2(x), \theta_3(x)] = [\mu(x), \sigma_L(x), \nu(x)]$. Which means,

$$\boldsymbol{Y}|\boldsymbol{X} \sim BCCG\ [\mu(x), \sigma_L(x), \nu(x)]$$

with the transformed response

$$Z = \begin{cases} \frac{[\frac{Y}{\mu(x)}]^{\nu(x)}-1}{\sigma_L(x)\nu(x)}, & \nu(x) \neq 0 \\ \frac{1}{\sigma_L(x)} \log[\frac{Y}{\mu(x)}], & \nu(x) = 0 \end{cases} \tag{3.1}$$

for $0 < \boldsymbol{Y} < \infty$, where $\mu(x) > 0, \sigma_L(x) > 0$ and $-\infty < \nu(x) < \infty$. Z is assumed to follow a truncated standard normal distribution with $-\frac{1}{\sigma_L(x)\nu(x)} < Z < \infty$, if $\nu(x) > 0$ and $-\infty < Z < -\frac{1}{\sigma_L(x)\nu(x)}$, if $\nu(x) < 0$. Hence the probability density function (pdf) of $\boldsymbol{Y}$ is given by

$$f_Y(y) = \frac{y^{\nu(x)-1} \exp(-\frac{1}{2}z^2)}{[\mu(x)]^{\nu(x)}\, \sigma_L(x)\sqrt{2\pi}\phi(\frac{1}{\sigma_L(x)|\nu(x)|})} \tag{3.2}$$

where z is given by equation 3.1 and $\phi(.)$ is the distribution function of a standard normal distribution. Cole & Green (1992) assumed that Z follows a normal distribution; therefore, the truncation probability $\phi(\frac{1}{\sigma_L(x)|\nu(x)|})$ is negligible. Here $\mu(x)$ is the median of $\boldsymbol{Y}$ for covariate x (e. g., age), which can be modeled with log-link as follows:

$$\begin{aligned} \log[\mu(x)] &= [b_\mu(x)]^T \beta_\mu \\ &= \sum_{j=0}^{p} b_{\mu_j}(x)\beta_{\mu_j} \\ &= \beta_{\mu_0} + \beta_{\mu_1} x_i + \beta_{\mu_2} x_i^2 + \ldots + \beta_{\mu_p} x_i^p. \end{aligned}$$

Hence, as a simple example, we have assumed that μ is believed to have a p^{th} order polynomial basis. Thus the space of polynomial of order p and below contains μ. A basis for this space is, $b_{\mu_0}(x) = 1, b_{\mu_1}(x) = x_i, b_{\mu_2}(x) = x_i^2 \ldots, b_{\mu_p}(x) = x^p$. Similarly, the scale (coefficient of

variation) with log-link can be modeled as follows:

$$\begin{aligned}\log\left[\sigma_L(x)\right] &= \left[b_{\sigma_L}(x)\right]^T \beta_\sigma \\ &= \sum_{j=0}^{p} b_{\sigma_{Lj}}(x)\beta_{\sigma_{Lj}} \\ &= \beta_{\sigma_{L0}} + \beta_{\sigma_{L1}} x_i + \beta_{\sigma_{L2}} x_i^2 + \ldots + \beta_{\sigma_{Lp}} x_i^p.\end{aligned}$$

The shape (skewness) parameter with identity link can be modeled as follows:

$$\begin{aligned}\nu(x) &= \left[b_\nu(x)\right]^T \beta_\nu \\ &= \sum_{j=0}^{p} b_{\nu_j}(x)\beta_{\nu_j} \\ &= \beta_{\nu_0} + \beta_{\nu_1} x_i + \beta_{\nu_2} x_i^2 + \ldots + \beta_{\nu_p} x_i^p,\end{aligned}$$

where $\beta_\mu, \beta_{\sigma_L}, \beta_\nu \in \mathbb{R}$ are vectors of polynomial coefficients. This type of likelihood function was first proposed by Green (1987) in a general semi-parametric regression setting, and then further used by Cole & Green (1992) for LMS method. The log-likelihood function, l derived from Equation 3.1 for n independent cases for (y_i, x_i) is given by,

$$\begin{aligned}l &= l(\mu, \sigma_L, \nu) \\ &= \log\left(\prod_{i=1}^{n} \frac{[\frac{y_i}{\mu(x_i)}]^{\nu(x_i)} - 1}{\sigma_L(x_i)\nu(x_i)}\right) \\ &= \sum_{i=1}^{n}\left(\nu(x_i)\log(\frac{y_i}{\mu(x_i)}) - \log(\sigma_L(x_i)) - \frac{1}{2}z_i^2\right)\end{aligned}$$

The GAMLSS package of `R` was proposed by Rigby & Stasinopoulos (2010), where the Rigby and Statsinopoulos (RS) algorithm can carry out the iterative procedure to maximize the log-likelihood function to obtain the estimate of the models of median, coefficient of variation and skewness of the data.

3.3 Lower Limit of Normal (LLN) for Lung Function

This section focuses on calculating the Lower Limit of Normal (LLN) of the lung function indices. As described in Quanjer *et al.* (2012), for spirometric tests, the suggestion is to use 90% (NOT 95%), and two-sided (NOT one-sided) for calculating the LLN. Thus the remaining 10% are equally distributed in the two tails with 5% each side. Use of 90% lead to the fact that 5% of the population are considered to have too low values as opposed to 2.5%, that is, a larger portion of the population (5%) have too low values.

The formula for calculating the LLN (i.e., 5^{th} percentile of BCCG distribution) involves the predicted values of median, skewness and coefficient of variation. For $\nu(x) \neq 0$,

$$
\begin{aligned}
z &= \frac{\left[\frac{y}{\mu(x)}\right]^{\nu(x)} - 1}{\sigma_L(x)\nu(x)} \\
&=> \quad z\sigma_L(x)\nu(x) = \left[\frac{y}{\mu(x)}\right]^{\nu(x)} - 1 \\
&=> \quad 1 + z\sigma_L(x)\nu(x) = \left[\frac{y}{\mu(x)}\right]^{\nu(x)} \\
&=> \quad \log(1 + z\sigma_L(x)\nu(x)) = \nu(x)\log\left[\frac{y}{\mu(x)}\right] \\
&=> \quad \log\left[\frac{y}{\mu(x)}\right] = \frac{\log(1 + z\sigma_L(x)\nu(x))}{\nu(x)} \\
&=> \quad \log(y) - \log[\mu(x)] = \frac{\log(1 + z\sigma_L(x)\nu(x))}{\nu(x)} \\
&=> \quad \log(y) = \log[\mu(x)] + \frac{\log(1 + z\sigma_L(x)\nu(x))}{\nu(x)} \\
&=> \quad \log(y) = \log[\mu(x)] + \log[1 + z\sigma_L(x)\nu(x)]^{\frac{1}{\nu(x)}} \\
&=> \quad \log(y) = \log\left(\mu(x)[1 + z\sigma_L(x)\nu(x)]^{\frac{1}{\nu(x)}}\right) \\
&=> \quad y = \mu(x)[1 + z\sigma_L(x)\nu(x)]^{\frac{1}{\nu(x)}}
\end{aligned}
$$

For $\nu(x) = 0$, the formula for calculating LLN involves the pdf of z without the component for skewness, as described in Equation 3.1.

$$
\begin{aligned}
z &= \frac{1}{\sigma_L(x)} \log\left[\frac{y}{\mu(x)}\right] \\
&=> \quad z\sigma_L(x) = \log\left[\frac{y}{\mu(x)}\right] \\
&=> \quad \exp\left[z\sigma_L(x)\right] = \frac{y}{\mu(x)} \\
&=> \quad y = \mu(x)\left[\exp(z\sigma_L(x))\right]
\end{aligned}
$$

Therefore, the formula for calculating lower 5^{th} percentile becomes

$$
y_{0.05} = \begin{cases} \mu(x)\big(1 - \sigma_L(x)\nu(x)z_{0.05}\big)^{\frac{1}{\nu(x)}}, & \nu(x) \neq 0 \\ \mu(x)\exp(-\sigma_L(x)z_{0.05}), & \nu(x) = 0 \end{cases} \tag{3.3}
$$

Using this formula, the LLN can be calculated for spirometric indices based on the reference model.

3.4 Selection of the Reference Model

The Akaike information criterion (AIC), which was proposed by Akaike (1973) is widely used for model selection. For a set of models for a given data, AIC estimates a criteria, which gives the minimum value for the reference model. It is calculated based on the maximized value of the likelihood function. Suppose $l = \log(L)$ be the log-likelihood estimate of the selected model, where L is the likelihood function, and k be the number of estimated parameters in the model. Thus the value of AIC of the particular model is as follows:

$$\text{AIC} = 2k - 2l.$$

Given a set of model, the best model can be chosen based on the minimum value of AIC. There might be some models, whose values are almost close to each other. In such case, the models can be compared in the following way. We can consider,

$$\Delta\text{AIC} = \Delta_i = \text{AIC}_i - \text{AIC}_{\min},$$

where AIC_i is the value of the i^{th} model; i.e., the model of interest and $\text{AIC}_{\min}$ is the AIC

value of a model, with the lowest value of AIC. As a rule of thumb, $\Delta_i < 2$ suggests significant evidence for the model of interest. If $3 \leq \Delta_i \leq 7$, it indicates that the model of interest has considerably less support, whereas, $\Delta_i > 10$ indicates a strong support against the model of interest (Burnham & Anderson, 2002).

This chapter gives an idea about GAMLSS and how it can be applied to a data, where the relationship between the response and predictor variables is non-linear, the response variable is skewed, and the error term does not necessarily follow a normal distribution. The next chapter 4 is about how the study is designed, the methods for selecting individuals, how the spirometry was performed, what exclusion criteria is considered to choose healthy subjects, how the data were analyzed and the interpretations of the LLN for lung function values.

Chapter 4

Data Analyses: Predictions Equations for the First Nations Children and Adolescents

Lung function data analyses for the First Nations children are presented in this chapter. We begin with a description of the data in Section 4.1. The model selection procedure is discussed in Section 4.2, and then the results (model fits, prediction curves and LLNs) are presented in Section 4.3. We conclude this chapter by summarizing the important findings in Section 4.4.

4.1 Data and Variables

The results from the Saskatchewan First Nations Lung Health Project (FNLHP) were used to derive the reference equations for this study. FNLHP is an ongoing prospective cohort study to identify factors associated with respiratory outcomes for the First Nations people living on reserves (Pahwa *et al.*, 2015). The project is being carried out by Prof. James A. Dosman (Distinguished Research Chair, Canadian Center for Health and Safety in Agriculture) and his team, and is funded by the Canadian Institute of Health Research (MOP: 246983-11829). The study began in 2012 with a cohort of 874 First Nations Cree people (428 males and 446 females) from two reserves in Saskatchewan. They will be followed longitudinally until 2017 with the long-term goals to implement potential intervention programs and to address issues that have been identified by the baseline data on respiratory health and will be re-evaluated. The first phase of the study was completed during 2012-2013, involving survey questionnaire and clinical assessment (see Appendix A for survey questionnaire and Appendix B for clinical assessment form). In this study, lung function data for children and adolescents (ages between 6 to 17 years) from FNLHP were used to derive reference equations, with the ultimate goal to derive a scientific mechanism in predicting lung function values for the population of the

First Nations Cree children and adolescents living in two on-reserve communities. This study was approved by the University of Saskatchewan Biomedical Research Ethics Board (Bio# 15-69; see Appendix C).

There were a total of 351 children and adolescents (47% males and 53% females), who completed the baseline survey. The survey questionnaire included questions regarding their past and current health conditions, lifestyles, personal and family history of chronic diseases. The clinical component of the study assessed anthropometric variables (standing height, weight and waist circumference), blood pressure, pulmonary function testing (FVC, FEV_1 and FEV_1/FVC) and allergy skin status through skin prick testing. Parental report of the date of birth was used to calculate the age. Height of the individual and height was measured using a fixed tape with a participant standing on a hard surface without wearing shoes (Chen *et al.*, 2005). Sensormedics (Anaheim, CA) dry rolling seal spirometer was used to measure the lung function values. Lung function testing was conducted according to criteria set by the American Thoracic Society (ATS) (Miller *et al.*, 2005). All lung function testing was conducted by registered nurses certified in Spirometry by the Lung Association of Saskatchewan. The spirometer was calibrated at the early morning and mid-day to meet the ATS guideline. The lung function testing results were assessed by a respirologist. Results from children who did not have acceptable curves on the first testing were retested and the best curves were retained in the study for further analysis. A letter accompanying the consent described the survey procedures (Appendix D). A consent form was completed by the parent and assent form by the study participants who were of 16 years of age or over, who were not living with a parent or guardian were considered to be emancipated and able to complete the consents for themselves.

As described in Chapter 2, the most important covariates to predict lung function values are height and age. Moreover, the general recommendation is to develop prediction equations based on data from healthy individuals. Therefore, a reduced dataset of healthy children and adolescents was considered, and height and age were used in a regression model to derive lung function prediction equations for FVC, FEV_1 and FEV_1/FVC. Participants with missing observations for any of FVC, FEV_1, age and height were excluded from the study, leading to a total of 302 children and adolescents with complete data. An individual was then considered healthy if the answer was "no" to all of the questions stated below.

- Has your child ever had a dry cough at night or first thing in the morning, in the past 12 months?

- Does this child usually have tightness in the chest or bring up phlegm or mucus apart from colds?
- In the past 12 months, has this child had a wheeze or whistling noise that comes from the chest?
- Has this child ever been diagnosed by a doctor as having asthma?
- Has a doctor ever said this child had tonsillitis?
- Has a doctor ever said this child had bronchitis?
- Has a doctor ever said this child had pneumonia?
- Has a doctor ever said this child had croup?
- Has a doctor ever said this child had sleep apnea?
- Does this child/adolescent smoke today? If yes, how many years has this child/adolescent been smoking? [A participant with a history of smoking for more than one year was excluded from the study]

The systematic procedure of selecting a sample of healthy children and adolescents is described in Figure 4.1, which leads to a sample of 130 children and adolescents (53 boys and 77 girls).

Figure 4.1: The process of selecting a sample of healthy individuals with no missing observations for FVC, FEV_1, age and height.

The summary statistics of the study variables (age, height, FVC, FEV_1 and FEV_1/FVC) are displayed in Table 4.1. The measurement of skewness justifies the use of GAMLSS to analyze the data: (a) the distribution of age is right skewed for both males and females (skewness is 0.40 and 0.48, respectively), (b) the distribution of standing height is right skewed for males (skewness = 0.22) and very close to symmetric for females (skewness = -0.03), (c) the distributions of FVC and FEV_1 are right skewed for both males and females (for FVC, skewness = 0.73 and 0.58 for males and females, respectively; and for FEV_1, skewness = 1.06 and 0.58 for males and females, respectively), and (d) the distribution of FEV_1/FVC is left skewed for both males and females (skewness = -0.31 and -0.66, respectively). These facts are also evident from the histograms of these variables displayed in Figures 4.2 and 4.3. The t tests comparing the means of each of the study variables for boys and girls resulted in significant differences between groups. The mean age of boys in this study was older than girls. As a result, the height and lung function were significantly higher in boys as well. This suggests the need for separate reference equations as well.

Table 4.1: Summary statistics of the study variables for a sample of 130 First Nations healthy children and adolescents ages 6-17 years.

Variables	Males (n = 53)			Females (n = 77)			Two sample t-test	
	Mean	SD	Skewness	Mean	SD	Skewness	t	p-value
Age (years)	11.04	2.67	0.40	9.97	2.98	0.48	2.09	0.02
Standing Height (cm)	151.74	15.84	0.22	146.00	15.98	- 0.03	2.01	0.02
FVC (liters)	3.14	1.02	0.73	2.68	0.93	0.58	2.66	<0.01
FEV_1 (litres)	2.72	0.93	1.06	2.37	0.83	0.58	2.24	<0.01
FEV_1/FVC	86.60	5.26	- 0.31	88.50	5.23	- 0.66	-2.03	0.02

4.2 Model Selection and Reference Values

Suppose that the distribution of the response variable y is defined by three parameters μ, σ_L and ν such that the transformed variable

$$z = \frac{(y/\mu)^{\nu} - 1}{\nu \times \sigma_L}$$

is a z score with distribution close to N(0,1). Here μ, σ_L and ν are the location (median), scale (coefficient of variation) and shape (skewness) parameters, respectively. Any skewness in y can be removed by a suitable choice of ν. The distribution of Z is called the Box-Cox-Cole-Green (BCCG) distribution. Following Quanjer *et al.* (2012), we consider GAM based on the BCCG distribution (i.e., GAMLSS) to develop prediction equations for FVC, FEV_1 or FEV_1/FVC. We apply log links for μ and σ_L and an identity link for ν, so that the models can be expressed as

$$\log(\mu) = \beta_0 + \beta_1 \log(\text{Height}) + \beta_2 \log(\text{Age}) + \left\{\alpha_0 + \alpha_1 \left(\frac{\text{Age}}{20}\right) + \ldots + \alpha_p \left(\frac{\text{Age}}{20}\right)^p\right\},$$

$$\log(\sigma_L) = \beta_0 + \beta_1 \log(\text{Age}) + \left\{\alpha_0 + \alpha_1 \left(\frac{\text{Age}}{20}\right) + \alpha_2 \left(\frac{\text{Age}}{20}\right)^2 + \ldots + \alpha_p \left(\frac{\text{Age}}{20}\right)^p\right\},$$

$$\nu = \beta_0 + \beta_1 \log(\text{Age}) + \left\{\alpha_0 + \alpha_1 \left(\frac{\text{Age}}{20}\right) + \alpha_2 \left(\frac{\text{Age}}{20}\right)^2 + \ldots + \alpha_p \left(\frac{\text{Age}}{20}\right)^p\right\},$$

where p is the order of a polynomial. Note that age is divided by 20 to scale down the polynomial bases; since the ages of the participants range from 6 to 17, age/20 $\in$ (0,1).

Initially, we consider several models based on the order p of the polynomial ($p = 1, 2, \ldots, 7$). Model selection procedure reveals that the smallest AIC values are achieved for models as displayed in Table 4.2. Then, the data are analyzed using the values of p presented in Table 4.2 to develop reference equations for FVC, FEV_1 and FEV_1/FVC. As an example, the AIC value is 22.01 for FVC (males) of the model when the order of the polynomial for ν is 1. On the other hand, the AIC value is 23.84 for FVC (males) when the polynomial order for ν is 2. Therefore, the the former model with lower AIC value is selected as the reference model. (AIC values for different combinations of polynomials of the lung function indices are presented in Appendix E; see pages 87, 90, 94, 97, 101, 104 and 105).

Once the estimates of the regression coefficients for each of the μ, σ_L and ν models are obtained, we can estimate μ, σ_L and ν for different values of height and age. For a particular height-age combination, the estimates of μ, σ_L and ν can then be used to obtain the median reference value and LLN: given μ, σ_L and ν, the median reference value and the LLN are the 50^{th} and the 5^{th} percentiles of the BCCG distribution, respectively. The height-age adjusted median reference curve and the LLN curve are then constructed using the estimated median reference values and LLNs for different combinations of heights and ages. The GAMLSS package (Stasinopoulos & Rigby, 2007) in `R` (R Core Team, 2015) is used to fit the models.

Table 4.2: Order, p, of the polynomials for which the smallest AIC values are achieved.

Lung Function	Males			Females		
	p for the μ Model	p for the σ_L Model	p for the ν Model	p for the μ Model	p for the σ_L Model	p for the ν Model
FVC	4	3	1	3	3	1
FEV_1	3	3	1	3	3	1
FEV_1/FVC	3	3	2	3	2	1

All the curves are then constructed using the *plot()*, *lines()* and *contour()* functions in R. The relevant R codes are presented in Appendix E.

4.3 Results

The distribution of anthropometric measurements and lung functions of both males and females are given in Figure 4.2 and 4.3. The ages of the healthy First Nations children and adolescents are skewed to the right and there are only few participants of 16 years of age or older. The distributions of standing height for both males and females are symmetric. The modal standing height for the males is 155 cm, indicating the most frequently appeared height in the data set, which comprises of 26.5% of the male participants. The modal standing height for females is 145 cm, which comprises of approximately 21.5% of the data set of healthy female participants. Figure 4.3 indicates that the distributions of FVC and FEV_1 are both skewed to the right, whereas, the distributions of FEV_1/FVC for both males and females are slightly skewed to the left. As indicated by the scatterplots of Figure 4.4, the observed values of FVC and FEV_1 increase with both age and standing height. However, the FEV_1/FVC ratios exhibit an irregular shape and large variability. Overall, there exists a curved relationship between each of the lung function values (FVC, FEV_1 and FEV_1/FVC) and the anthropometric variables (i.e., age and standing height). The application of GAMLSS with polynomial bases captures such curved relationships.

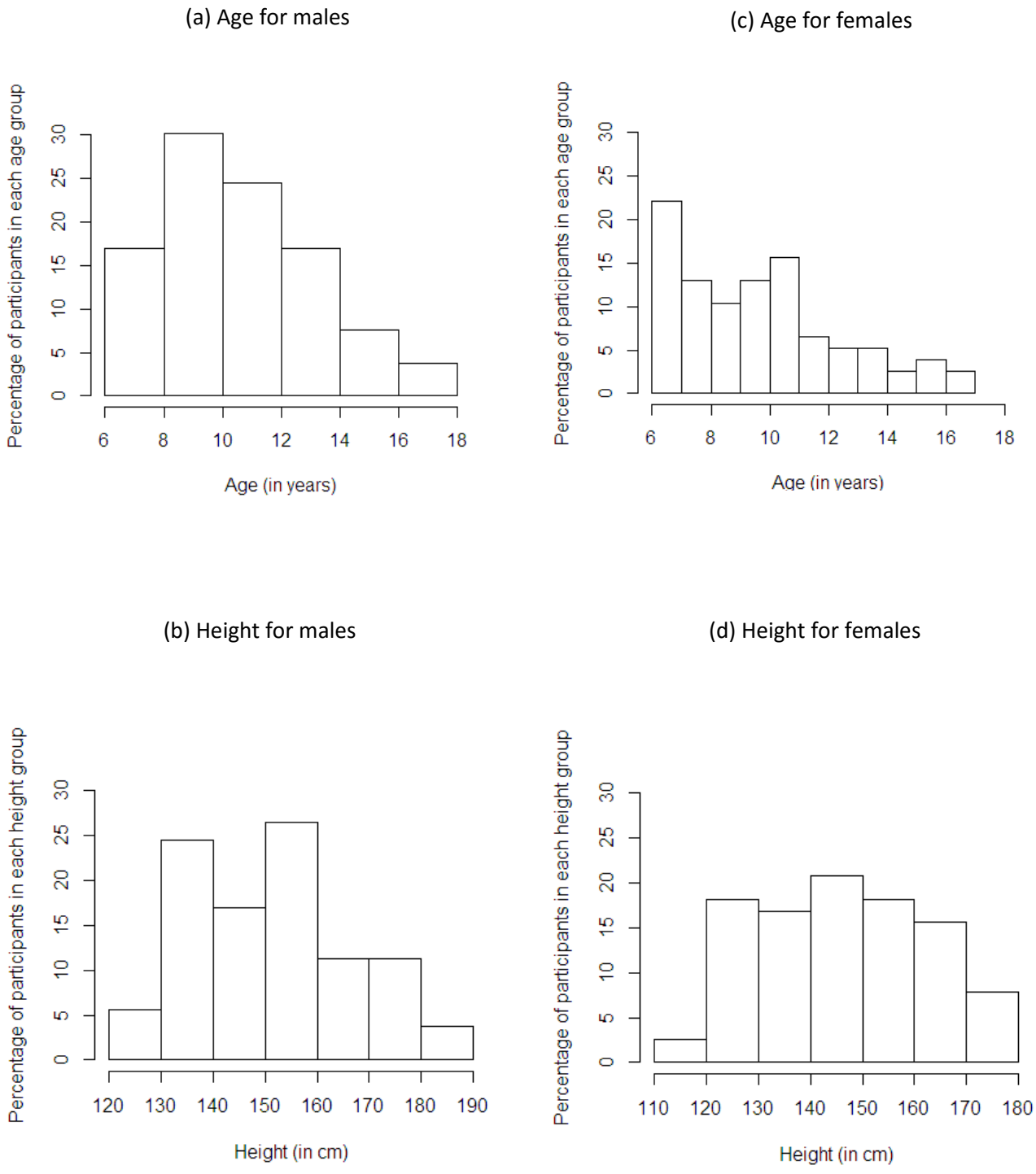
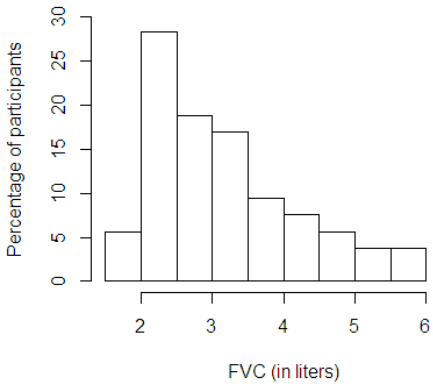
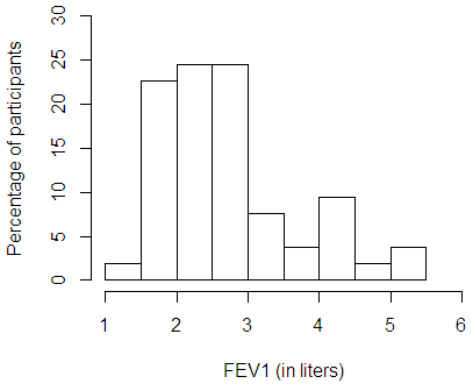


Figure 4.2: Histograms of the anthropometric measurements for 130 healthy participants.

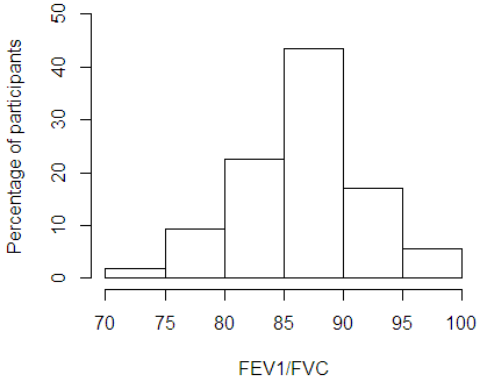
(a) FVC for males



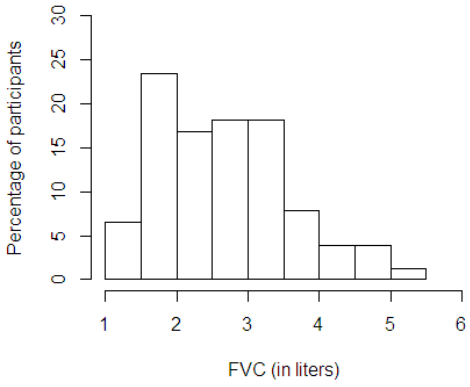
(b) FEV_1 for males



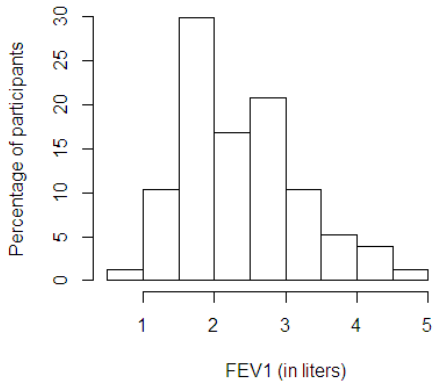
(c) FEV_1/FVC for males



(d) FVC for females



(e) FEV_1 for females



(f) FEV_1/FVC for females

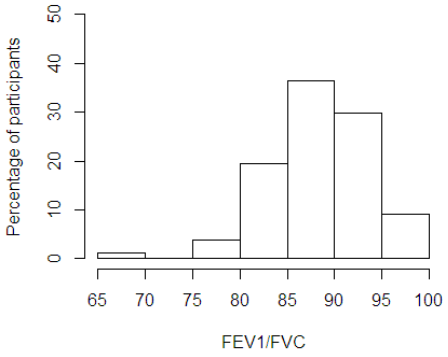


Figure 4.3: Histograms of the lung function indices for 130 healthy participants.

4.3.1 Model Fits and Prediction Equations

The predicted or reference models for lung function indices of First Nations boys and girls are given below:

The predicted equations for log(FVC) for males are

$$\begin{aligned}
\log(\mu) &= 6.79 + 3.05\log(\text{Height}) - 101.59\log(\text{Age}) \\
&\quad + \left\{809.65\left(\frac{\text{Age}}{20}\right) - 1169.67\left(\frac{\text{Age}}{20}\right)^2 + 967.58\left(\frac{\text{Age}}{20}\right)^3 - 326.46\left(\frac{\text{Age}}{20}\right)^4\right\} \\
\log(\sigma_L) &= -396.32 + 720.52\log(\text{Age}) \\
&\quad + \left\{3946.15\left(\frac{\text{Age}}{20}\right) + 3500.69\left(\frac{\text{Age}}{20}\right)^2 + 1339.53\left(\frac{\text{Age}}{20}\right)^3\right\}, \\
\nu &= -514.91 + 375.72\log(\text{Age}) - 694.24\left(\frac{\text{Age}}{20}\right).
\end{aligned}$$

The predicted equations for log(FVC) for females are

$$\begin{aligned}
\log(\mu) &= -4.29 + 2.04\log(\text{Height}) - 12.62\log(\text{Age}) \\
&\quad + \left\{78.56\left(\frac{\text{Age}}{20}\right) - 76.48\left(\frac{\text{Age}}{20}\right)^2 + 31.99\left(\frac{\text{Age}}{20}\right)^3\right\}, \\
\log(\sigma_L) &= 60.32 - 148.30\log(\text{Age}) \\
&\quad + \left\{909.35\left(\frac{\text{Age}}{20}\right) - 883.73\left(\frac{\text{Age}}{20}\right)^2 + 363.70\left(\frac{\text{Age}}{20}\right)^3\right\}, \\
\nu &= 55.34 - 39.35\log(\text{Age}) + 71.32\left(\frac{\text{Age}}{20}\right).
\end{aligned}$$

The predicted equations for log(FEV_1) for males are

$$\begin{aligned}
\log(\mu) &= -1.12 + 2.90\log(\text{Height}) - 68.82\log(\text{Age}) \\
&\quad + \left\{562.72\left(\frac{\text{Age}}{20}\right) + 833.86\left(\frac{\text{Age}}{20}\right)^2 + 705.75\left(\frac{\text{Age}}{20}\right)^3 - 242.63\left(\frac{\text{Age}}{20}\right)^4\right\}, \\
\log(\sigma_L) &= -420.40 + 784.29\log(\text{Age}) \\
&\quad + \left\{4357.64\left(\frac{\text{Age}}{20}\right) + 3925.63\left(\frac{\text{Age}}{20}\right)^2 - 1526.00\left(\frac{\text{Age}}{20}\right)^3\right\}, \\
\nu &= -449.94 + 322.91\log(\text{Age}) - 581.99\left(\frac{\text{Age}}{20}\right).
\end{aligned}$$

The predicted equations for $\log(\text{FEV}_1)$ for females are

$$\begin{aligned}
\log(\mu) &= -6.58 + 1.68\log(\text{Height}) - 3.34\log(\text{Age}) \\
&\quad + \left\{21.32\left(\frac{\text{Age}}{20}\right) - 19.59\left(\frac{\text{Age}}{20}\right)^2 + 7.96\left(\frac{\text{Age}}{20}\right)^3\right\}, \\
\log(\sigma_L) &= 65.22 - 148.93\log(\text{Age}) \\
&\quad + \left\{886.83\left(\frac{\text{Age}}{20}\right) - 84.92\left(\frac{\text{Age}}{20}\right)^2 + 339.52\left(\frac{\text{Age}}{20}\right)^3\right\}, \\
\nu &= 46.15 - 32.53\log(\text{Age}) + 60.10\left(\frac{\text{Age}}{20}\right).
\end{aligned}$$

The predicted equations for $\log(\text{FEV}_1/\text{FVC})$ for males are

$$\begin{aligned}
\log(\mu) &= 3.54 - 0.37\log(\text{Height}) + 7.90\log(\text{Age}) \\
&\quad - \left\{52.36\left(\frac{\text{Age}}{20}\right) + 54.41\left(\frac{\text{Age}}{20}\right)^2 - 23.31\left(\frac{\text{Age}}{20}\right)^3\right\}, \\
\log(\sigma_L) &= -564.84 + 975.72\log(\text{Age}) \\
&\quad - \left\{5170.95\left(\frac{\text{Age}}{20}\right) + 4428.07\left(\frac{\text{Age}}{20}\right)^2 - 1638.41\left(\frac{\text{Age}}{20}\right)^3\right\}, \\
\nu &= -1294.20 + 1420.90\log(\text{Age}) - \left\{4837.50\left(\frac{\text{Age}}{20}\right) + 1883.80\left(\frac{\text{Age}}{20}\right)^2\right\}.
\end{aligned}$$

The predicted equations for $\log(\text{FEV}_1/\text{FVC})$ for females are

$$\begin{aligned}
\log(\mu) &= 2.57 - 0.24\log(\text{Height}) + 6.80\log(\text{Age}) - \left\{40.61\left(\frac{\text{Age}}{20}\right) + 39.02\left(\frac{\text{Age}}{20}\right)^2\right\}, \\
\log(\sigma_L) &= 3.79 - 7.07\log(\text{Age}) + \left\{24.51\left(\frac{\text{Age}}{20}\right) - 11.53\left(\frac{\text{Age}}{20}\right)^2\right\}, \\
\nu &= -20.57 + 24.79\log(\text{Age}) - 64.72\left(\frac{\text{Age}}{20}\right).
\end{aligned}$$

From the above equations, we can see that the polynomial order for the model of μ in males are higher compared to girls for all the lung function indices. Therefore, the AIC values are higher in the models of males compared to females. The reason behind this is that the plots of lung function data in terms of anthropometric measurements in males exhibited more curved patterns compared to females in Figure 4.4. Another interesting fact about the

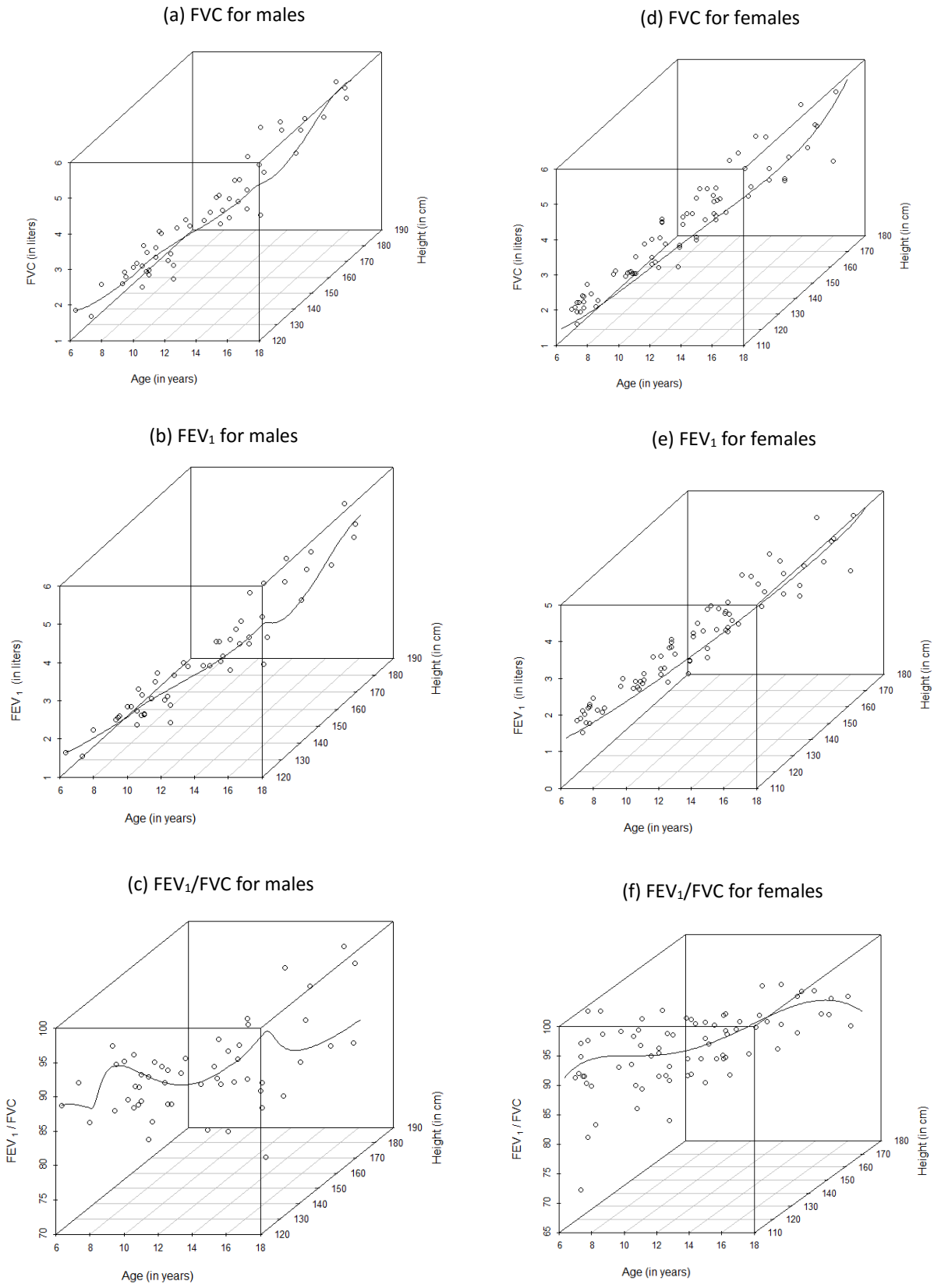


Figure 4.4: Observed data overlaid by the fitted curves for FVC, FEV_1 and FEV_1/FVC.

Table 4.3: Results from the GAMLSS analyses for log(FVC), log(FEV_1) and log(FEV_1/FVC) with covariates height and age.

		Males			Females		
Lung Function	Parameters	μ Model Estimate (p-value)	σ_L Model Estimate (p-value)	ν Model Estimate (p-value)	μ Model Estimate (p-value)	σ_L Model Estimate (p-value)	ν Model Estimate (p-value)
log(FVC)	Intercept	6.79 (0.08)	−396.32 (< 0.01)	−514.91 (< 0.01)	−4.29 (0.70)	60.32 (0.15)	55.34 (0.63)
	log(Height)	3.05 (< 0.01)			2.04 (< 0.01)		
	log(Age)	−101.59 (< 0.01)	720.52 (< 0.01)	375.72 (< 0.01)	−12.62 (0.54)	−148.30 (0.06)	−39.35 (0.64)
	Age/20	809.65 (< 0.01)	−3946.15 (< 0.01)	−694.25 (< 0.01)	78.56 (0.53)	909.35 (0.04)	71.32 (0.67)
	$(Age/20)^2$	−1169.67 (< 0.01)	3500.69 (< 0.01)		−76.48 (0.51)	−883.73 (0.02)	
	$(Age/20)^3$	967.58 (< 0.01)	−1339.53 (< 0.01)		31.99 (0.50)	363.70 (0.02)	
	$(Age/20)^4$	−326.46 (< 0.01)					
log(FEV_1)	Intercept	-1.12 (0.79)	-420.40 (< 0.01)	-449.94 (< 0.01)	-6.58 (0.33)	65.22 (0.04)	46.15 (0.29)
	log(Height)	2.90 (< 0.01)			1.68 (< 0.01)		
	log(Age)	-68.82 (< 0.01)	784.29 (< 0.01)	322.91 (< 0.01)	-3.34 (< 0.01)	-148.93 (0.38)	-32.53 (0.04)
	Age/20	562.72 (< 0.01)	4357.64 (< 0.01)	-581.99 (< 0.01)	21.32 (0.80)	886.83 (0.04)	60.10 (0.48)
	$(Age/20)^2$	-833.86 (< 0.01)	3925.63 (< 0.01)		-19.59 (0.81)	-840.92 (0.04)	
	$(Age/20)^3$	705.75 (< 0.01)	-1526.00 (< 0.01)		7.96 (0.82)	339.52 (0.04)	
	$(Age/20)^4$	-242.63 (< 0.01)					
log(FEV_1/FVC)	Intercept	3.54 (< 0.01)	-564.84 (< 0.01)	-1294.20 (< 0.01)	2.57 (< 0.01)	3.79 (0.04)	-20.57 (0.70)
	log(Height)	-0.37 (< 0.01)			-0.24 (< 0.01)		
	log(Age)	7.90 (< 0.01)	975.72 (< 0.01)	1420.90 (< 0.01)	6.80 (< 0.01)	-7.07 (< 0.01)	24.79 (0.60)
	Age/20	-52.36 (< 0.01)	-5170.95 (< 0.01)	-4837.50 (< 0.01)	-40.61 (< 0.01)	24.51 (< 0.01)	-64.72 (0.54)
	$(Age/20)^2$	54.41 (< 0.01)	4428.07 (< 0.01)	1883.80 (< 0.01)	39.02 (< 0.01)	-11.53 (< 0.01)	
	$(Age/20)^3$	-23.31 (< 0.01)	-1638.41 (< 0.01)				

estimated equations is that the signs of the polynomials are alternatively positive (+) and negative (−). This is because the observed curves are having upward and downward peaks in Figure 4.4 to adequately represent the observed data. Fitted curves overlaid with the observed data are displayed in Figure 4.4, and the estimates of the regression coefficients in Table 4.3. The estimated curves (Figure 4.4) show that the models fit the lung function data reasonably well for all the lung function indices.

All the slope parameters for males are significant ($p < 0.01$), indicating that both standing height and age have substantial effects in predicting FVC. Moreover, significant slopes for the polynomial bases suggest a curved relationship between age and FVC. For females, only standing height is significantly associated with FVC ($p < 0.01$), and insignificant slopes for the polynomial bases suggest a linear relationship between FVC and age. We see similar FEV_1 results for males: both standing height and age have significant effects in predicting FEV_1 ($p < 0.01$), and there is a sharp increase in FEV_1 at around age 13 (Figure 4.4(b)). However, standing height is found significantly associated with FEV_1 for females ($p < 0.01$), though age is not significant and there is approximately a linear relationship between age and FEV_1 (polynomial bases are not significant; also see Figure 4.4 (e)). For both males and females, significant slopes in the μ and σ_L models suggest a curved relationship and large variability in FEV_1/FVC. This is also evident in Figures 4.4 (c) and (f). There is significant variability in each of FVC, FEV_1 and FEV_1/FVC for males, whereas a significant variability only in FEV_1/FVC is observed for females.

A comparison among the fitted median curves for FVC and FEV_1 is shown in Figure 4.5. The predicted FVC values for males (solid in red) are higher than those for females (solid in green) for ages between 6 and 17. On the other hand, we see lower FEV_1 for females until around age 11 (dashed in green), and then very similar between 11 and 13. We see a sudden drop in FEV_1 for males at around age 13 (dashed in red), followed by a transition until around age 15. During this transition phase, FEV_1 values are higher for females than those for males. There is a sharp increase in FEV_1 for males after around age 15, leading to higher FEV_1 values for males compared to females.

4.3.2 Predicted Curves

The contour plots in Figure 4.6 represents the fitted curves for median lung function indices adjusted for both age and height. This plot is helping us to read the 3-dimensional relationship in a 2-dimensions. The X-axis represents the two covariates age and height. The Y-axis represents the predicted median values for lung function index, which is represented by contour lines. The predicted values for lung function index can be obtained from the

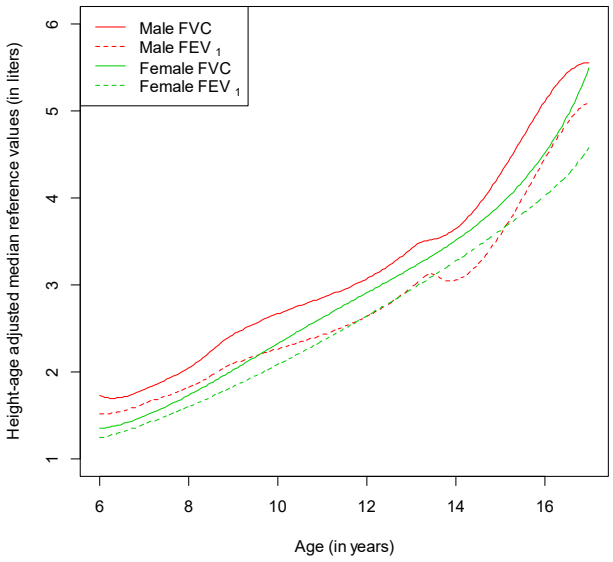


Figure 4.5: Height-age adjusted median reference values for FVC and FEV_1; solid red and green curves indicate FVC for males and females, respectively; dashed red and green curves indicate FEV_1 for males and females, respectively.

contour plots. For example, the predicted FVC value for a 6 years old boy with 130 cm height is 2.0 liters (as the corresponding age and height fall in the line of 2.0). Similarly, the predicted FVC value for a 6 years old girl, having 130 cm height will be approximately 1.8 liters (as the corresponding age and height fall in the line of 1.8).

The contour lines for boys are exhibiting more curved pattern compared to girls. The same phenomena was observed in fitted curves of Figure 4.4. This is because of the fact that boys' lung function values had a sudden decrease at the age of 9. Again at the age of 13, the boys' lung function values had a sudden increase. The predicted contour plots were able to capture such sudden increase and decrease; therefore, giving us a better fit to the data. The density plots of Figure 4.7 confirms the residuals follow approximately normal distribution.

4.3.3 Lower Limit of Normal (LLN)

The fitted curves for the lower 5^{th} percentile of the BCCG distribution or the lower limit of normal (LLN) are given in Figure 4.8. One can easily find the predicted values and LLN values for lung function indices from the red and green curves, respectively. The red curves represent height-age adjusted values for median FVC, whereas, the green curves represent height-age adjusted LLN. For example, the predicted value for median FVC for

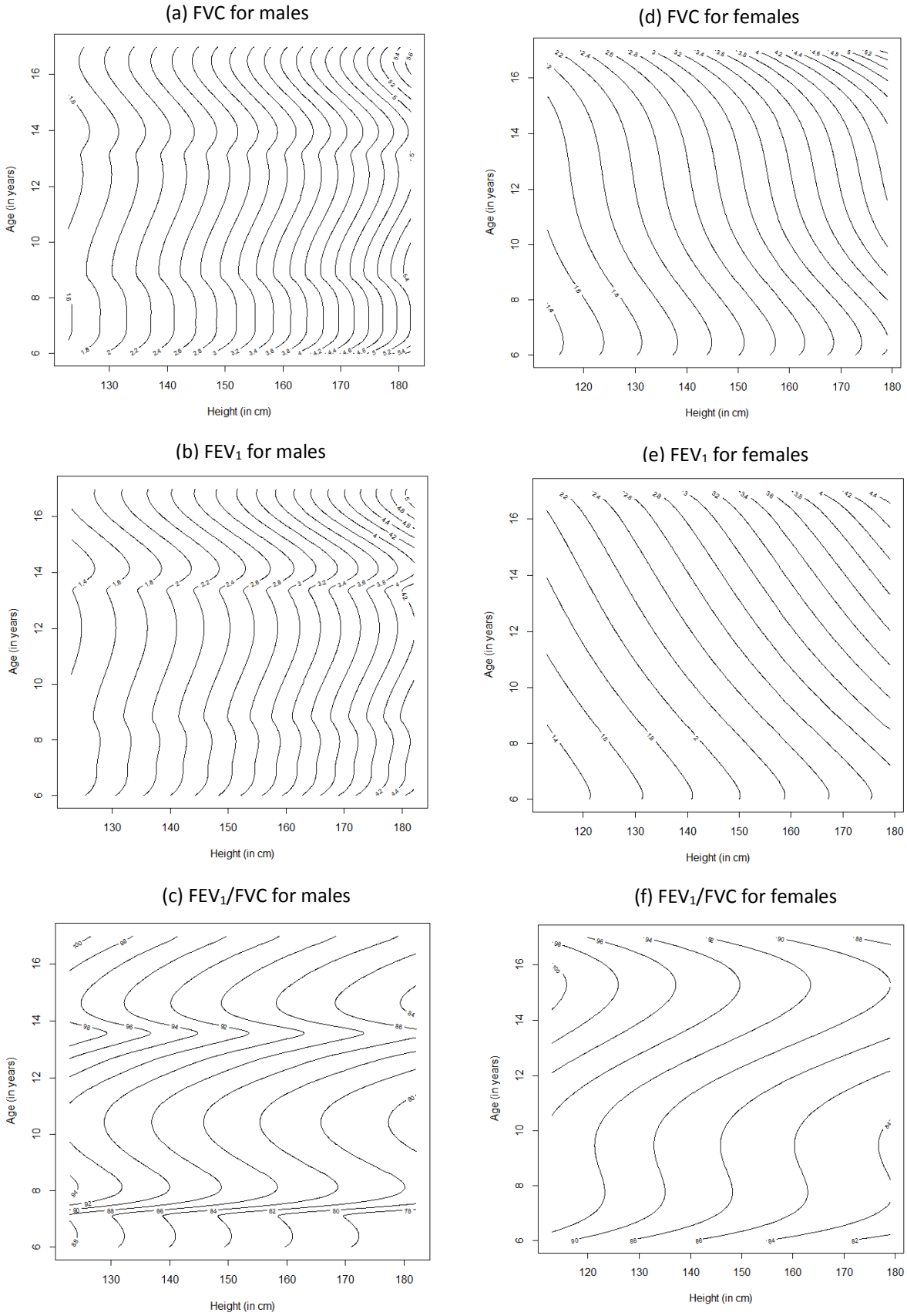
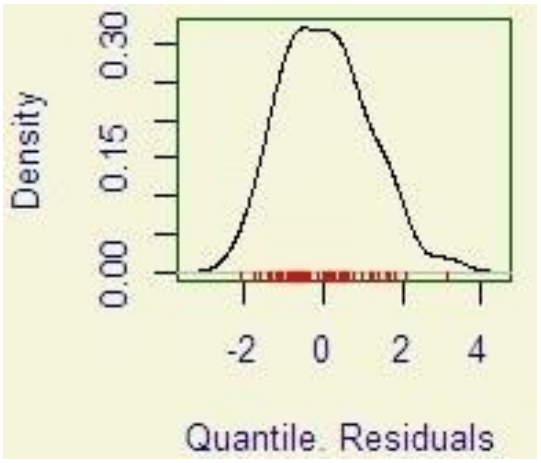
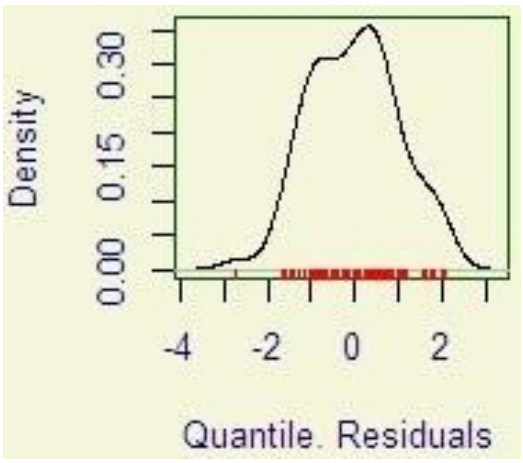


Figure 4.6: Contour plots for the fitted median lung function adjusted for age and height.

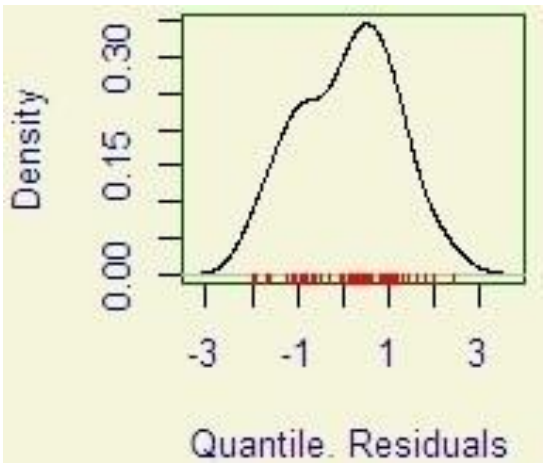
(a) FVC for males



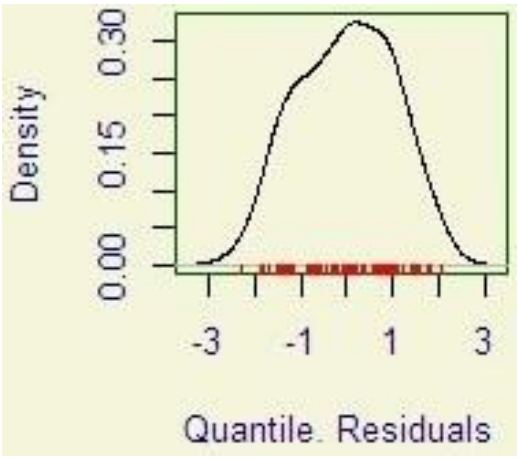
(b) FVC for females



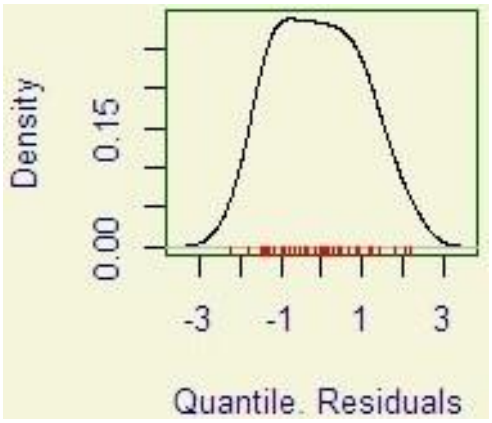
(c) FEV_1 for males



(d) FEV_1 for females



(e) FEV_1/FVC for males



(f) FEV_1/FVC for females

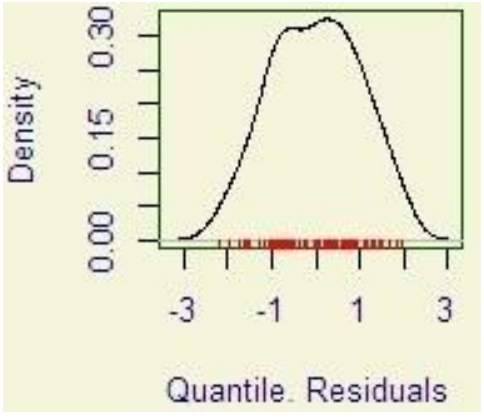


Figure 4.7: Residual plots for the fitted models of lung functions (FVC, FEV_1 and FEV_1/FVC).

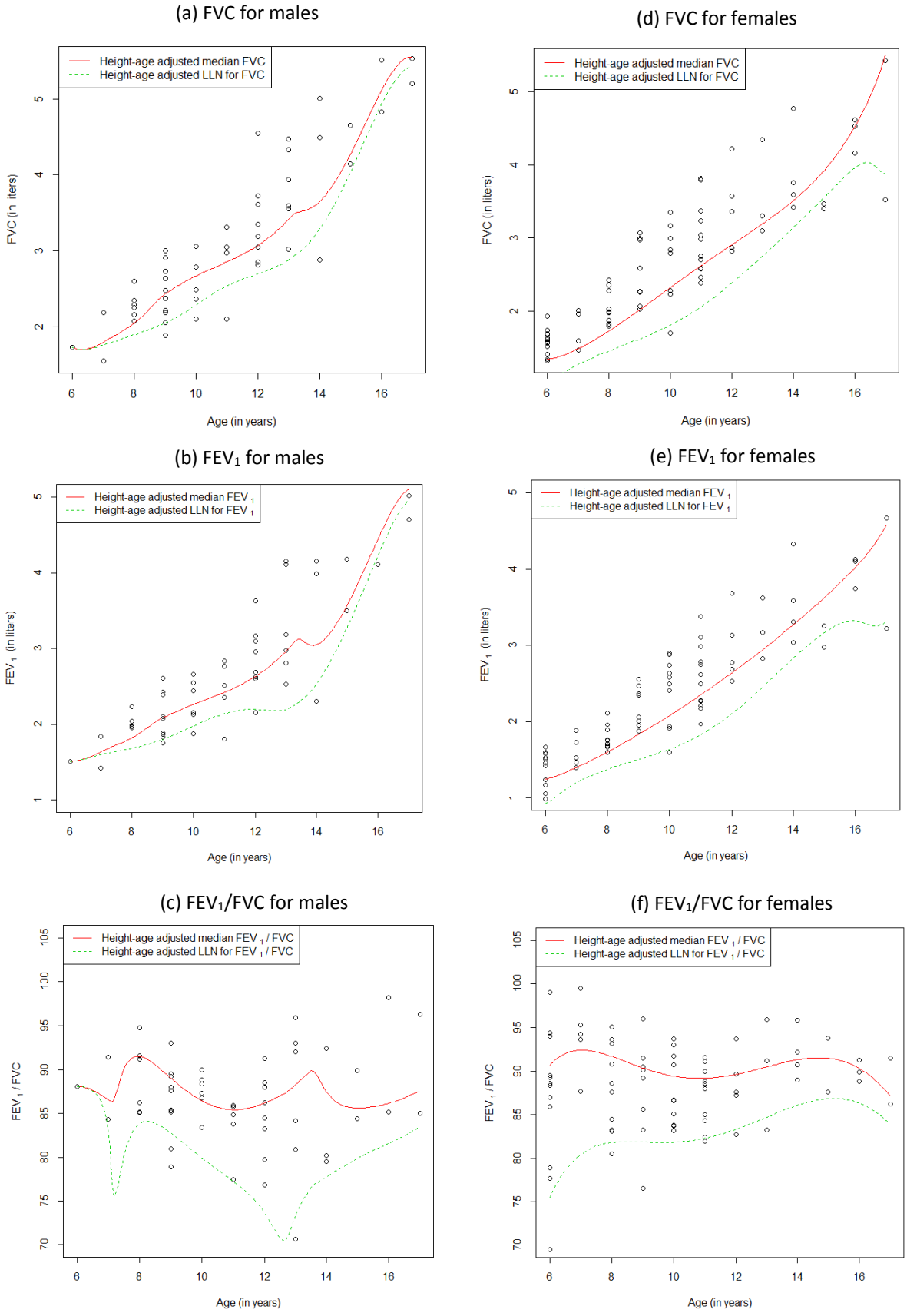


Figure 4.8: Lower Limit of Normal (LLN) for the median of lung function indices.

boys of 8 years is approximately 2.0 liters, and have LLN value of approximately 1.7 liters. Similarly for 10 years old girls, the predicted median FEV_1 is approximately 2.0 liters with a LLN value at 1.5 liters.

Only few observations were below the LLN curve, indicating that the LLN comprised of healthy individuals. In boys, there was only one observation at the age of 6 for all the lung function indices; therefore both the predicted curve and LLN curve intersect each other. This means that the predicted lung function value and LLN are same for a boy ages 6 years old. In reality, this should not happen. The same pattern was not observed in girls, as there were more than one observations in the data at the age of 6. At age 17 in girls, there were outlier both in the plot of FVC and FEV_1. This outlier bend the curve towards it. As a result, girls' LLN had a sudden decrease at the age of 17. There were outliers in FEV_1/FVC ratio for both boys and girls (identified by visual examination). Those outliers are pulling the LLN curve towards them, which was not observed in the predicted median curve.

4.4 Conclusion

The results in this chapter give new reference values for First Nations children and adolescents living in rural Saskatchewan, Canada. The results from this study is compared with the results from other studies (Hankinson *et al.*, 1999; Quanjer *et al.*, 2012). The prediction equations by Hankinson *et al.* (1999) of lung function (FVC, FEV_1 and FEV_1/FVC) for healthy Caucasian males < 20 years and females <18 years are given below-

$$\begin{aligned}
\text{FVC (Males)} &= -0.2584 - 0.20415\ \text{Age} + 0.010133\ \text{Age}^2 + 0.00018642\ \text{Height}^2 \\
\text{FVC (Females)} &= 1.2082 + 0.05916\ \text{Age} + 0.00014815\ \text{Height}^2 \\
\text{FEV}_1\text{ (Males)} &= -0.7453 - 0.04106\ \text{Age} + 0.004477\ \text{Age}^2 + 0.00014098\ \text{Height}^2 \\
\text{FEV}_1\text{ (Females)} &= -0.8710 + 0.06537\ \text{Age} + 0.00011496\ \text{Height}^2 \\
\text{FEV}_1/\text{FVC (Males)} &= 88.066 - 0.2066\ \text{Age} \\
\text{FEV}_1/\text{FVC (Females)} &= 90.809 - 0.2125\ \text{Age}
\end{aligned}$$

These equations are different than the reference equations derived for First Nations children and adolescents. The parameter estimates differed in both the studies. The fitted plots of the lung function indices based on this study and Hankinson *et al.* (1999) are presented in Figure 4.9. In all the plots, it is obvious that the fitted curve derived from Hankinson *et*

al. (1999) is more linear and their equations is not able to capture the sudden change in lung function during adolescent period. The model derived from GAMLSS is able to capture nonlinear relationship between lung function and anthropometric measurements. The fitted curve by Hankinson *et al.* (1999) on FEV_1 for girls underestimated the data, since it was going further away from the data. The LLN of FVC for a 13 years old boy with 128.63 cm height is 1.76 litre based on our study, whereas, it is 1.89 litre based on the equations derived for Caucasians. Therefore, the assessment of lung function of the First Nations children and adolescents based on the reference values derived from Caucasians will not appropriate.

The lung function prediction equations by Quanjer *et al.* (2012) of FVC, FEV_1 and FEV_1/FVC for healthy males and females ages 3-95 years of mixed/other ethnicity are given below-

The predicted equations for log(FVC) for males are

$$\begin{aligned}\log(\mu) &= -11.2281 + 2.4135\log(\text{Height}) - 0.0865\log(\text{Age}) - 0.0825 + f(\text{Age})\\ \log(\sigma_L) &= -2.2963 + 0.0718\log(\text{Age}) - 0.0503 + f(\text{Age})\\ \nu &= 0.9481,\end{aligned}$$

The predicted equations for log(FVC) for females are

$$\begin{aligned}\log(\mu) &= -10.4030 + 2.2633\log(\text{Height}) + 0.0234\log(\text{Age}) - 0.0833 + f(\text{Age})\\ \log(\sigma_L) &= -2.3549 + 0.1017\log(\text{Age}) - 0.0503 + f(\text{Age})\\ \nu &= 0.8236\end{aligned}$$

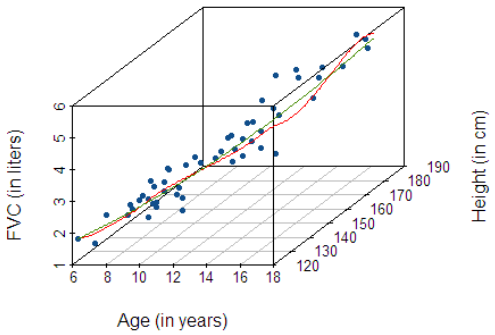
The predicted equations for log(FEV_1) for males are

$$\begin{aligned}\log(\mu) &= -10.3420 + 2.2196\log(\text{Height}) + 0.0574\log(\text{Age}) - 0.0708 + f(\text{Age})\\ \log(\sigma_L) &= -2.3268 + 0.0798\log(\text{Age}) + 0.0114 + f(\text{Age})\\ \nu &= 0.886600 + 0.085000\log(\text{Age})\end{aligned}$$

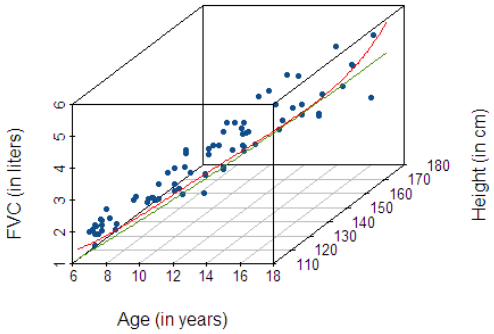
The predicted equations for log(FEV_1) for females are

$$\begin{aligned}\log(\mu) &= -9.6987 + 2.1211\log(\text{Height}) - 0.0270\log(\text{Age}) - 0.0708 + f(\text{Age})\\ \log(\sigma_L) &= -2.3765 + 0.0972\log(\text{Age}) + 0.0114 + f(\text{Age})\\ \nu &= 1.1540\end{aligned}$$

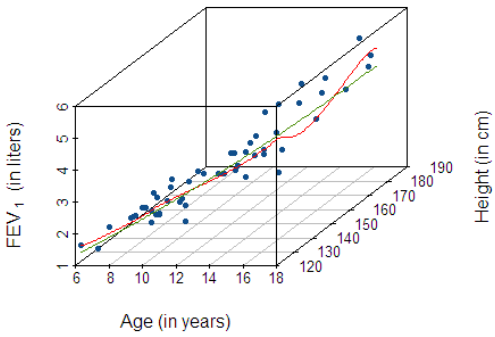
(a) FVC for males



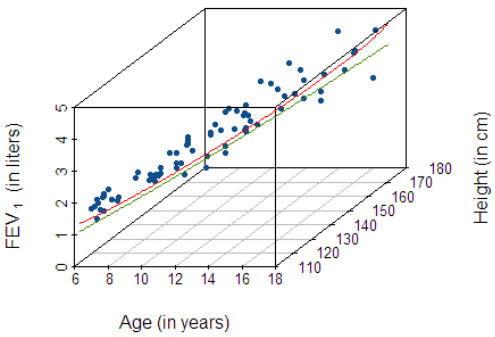
(d) FVC for females



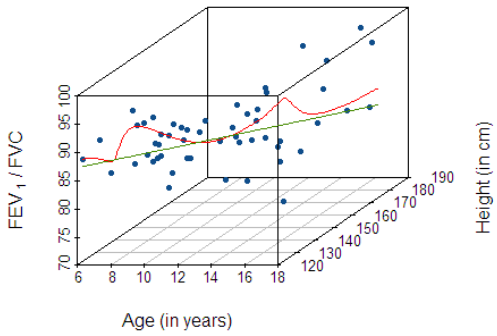
(b) FEV_1 for males



(e) FEV_1 for females



(c) FEV_1/FVC for males



(f) FEV_1/FVC for females

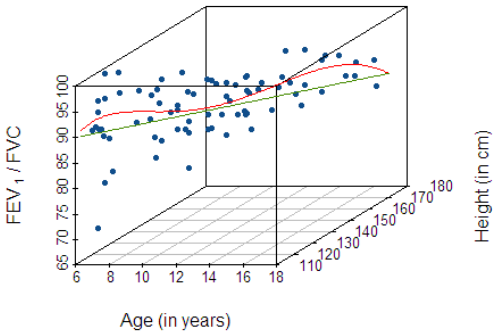


Figure 4.9: Comparison of the fitted plots of lung function indices (FVC, FEV_1) with other study; solid red curves indicate predicted median lung function based on GAMLSS; solid green curves indicate predicted mean lung function based on the study of Hankinson *et al.* (1999)

The predicted equations for $\log(\text{FEV}_1/\text{FVC})$ for males are

$$\begin{aligned}
\log(\mu) &= 0.7403 - 0.1595\log(\text{Height}) - 0.0366\log(\text{Age}) + 0.0106 + f(\text{Age}) \\
\log(\sigma_L) &= -2.9595 + 0.1156\log(\text{Age}) - 0.0860 + f(\text{Age}) \\
\nu &= 4.7101 - 0.6774\log(\text{Age})
\end{aligned}$$

The predicted equations for $\log(\text{FEV}_1/\text{FVC})$ for females are

$$\begin{aligned}
\log(\mu) &= 0.5506 - 10.78\log(\text{Height}) - 0.05044\log(\text{Age}) + 0.0106 + f(\text{Age}) \\
\log(\sigma_L) &= -3.2395 + 0.1850\log(\text{Age}) - 0.0860 + f(\text{Age}) \\
\nu &= 7.032 - 1.197\log(\text{Age})
\end{aligned}$$

The estimated prediction equations from GLI for 3-95 years age range are different from the reference equations derived by this study. The LLN of FVC for a 13 years old boy with 128.63 cm height is 1.76 litre based on our study, whereas, it is 1.0043 litre based on the equations derived for mixed ethnic group by Quanjer *et al.* (2012). The reason for such a difference is that there is no polynomial basis for skewness (ν) in GLI, whereas, in our study there is a coefficient for (Age/20) in the model for skewness. The specific form of the spline basis for children and adolescents in not provided by any document of GLI. Quanjer *et al.* (2012) provided some calculated values for the splines of age for people having <25 years of age. The calculated values for the splines of μ, σ_L and ν can be found from the lookup tables.

There are some significant findings from our study:

- Significant differences were observed in lung functions (FVC, FEV_1 and FEV_1/FVC) and anthropometric measurements between both boys and girls. Therefore, fitting separate equations for both the sexes are justified.

- The approach proposed by Quanjer *et al.* (2012) in Global Lung Function Initiative (GLI) was useful in prediction of lung functions for First Nations children and adolescents living in rural Saskatchewan, Canada.

- In GLI, polynomial bases of order 6-7 were used for modeling the μ, σ_L and ν. In this thesis, lower order polynomial bases (upto order 4) were enough to obtain the reference models.

- The equations derived by Hankinson *et al.* (1999) for Caucasians may not be applicable to the lung function data of Canadian First Nations children and adolescents.
- The predicted values of FVC was higher than the values for FEV_1 in both boys and girls (as presented in Figure 4.5). Therefore the values of FEV_1/FVC ratios is less than 100% in this population. In girls, the difference between the curves of FVC and FEV_1 was smaller compared to boys. Thus, the total volume of air for girls during exhalation are close to the volume of air exhaled at the first second.
- The predicted curves were able to capture the curved relationship between lung functions and anthropometric measurements.
- The LLN values obtained in this study can be used for the assessment of lung function, their response to treatment and diagnosis of respiratory illness in First Nations children and adolescents living in rural Saskatchewan, Canada.

Chapter 5

Discussion

The high prevalence of respiratory diseases among First Nations children and adolescents is a main concern in the public health sector of Canada (Bulkow *et al.*, 2012; McCuskee *et al.*, 2014), as it requires repeated hospitalization and admission to the paediatric intensive care unit (ICU) (Banerji *et al.*, 2001). Respiratory diseases are still under-recognized and under-diagnosed in First Nations children and adolescents living in Canada (Sin *et al.*, 2004). Proper statistical measurements (i.e., lung function reference equations) are needed to assess lung function and respiratory diseases in the above mentioned population, which motivated this research.

The two objectives of this study were

- to identify flexible and efficient statistical method to derive lung function prediction equations and LLN for children and adolescents, and

- to obtain prediction equations for FVC, FEV_1 and FEV_1/FVC for First Nations children and adolescents living in rural Saskatchewan, Canada.

An extensive literature review of spirometric prediction equations was conducted for children and adolescents, which are available for different ethnic groups. Statistical techniques to develop such equations revolved around linear regression, ranging from the simplest form of simple linear regression to more flexible method of generalized additive models for location, scale and shape (GAMLSS). In addition, transformation of the dependent variables were often used to remedy if assumptions associated with the regression model (linearity, normality and homoscedasticity) were not satisfied. The main focus of the literature review was to introduce each of these techniques in the context of modeling lung function values for prediction. Researchers mentioned that selecting the best prediction equations for lung function indices is important as it has great impact on the clinical interpretations of the results (Rosenfeld *et al.*, 2001; Subbarao *et al.*, 2004). Therefore, rather than depending on

the default reference equations, which are available with commercial spirometer, a sophisticated choice should be made (Pittman & Rosenfeld, 2011). In Chapter 2, it was seen that how spirometric reference equations were improved and generalized day-by-day.

In the review of the literature, most of the studies considered participants ages 5-20 years to be children and adolescents (Bjure, 1963; Dickman *et al.*, 1971; Miller *et al.*, 1977; Cotes *et al.*, 1979; Hsu *et al.*, 1979; Wall *et al.*, 1982; Coultas *et al.*, 1988; Shamssain *et al.*, 1988; Shamssain, 1991; Roizin *et al.*, 1993; Wypij *et al.*, 1993; Chowgule *et al.*, 1995; Parma *et al.*, 1996; Veale *et al.*, 1997; Rajkapoor *et al.*, 1997; Vijayan *et al.*, 2000; Budhiraja *et al.*, 2010; Rochat *et al.*, 2013; Choudhuri & Sutradhar, 2015). Few studies derived lung function prediction equations for children less than 6 years [(Bougrida *et al.*, 2012), (Rosenthal *et al.*, 1993), (Stanojevic *et al.*, 2008) and (Koopman *et al.*, 2011)]. Only two studies (Quanjer *et al.*, 2012) and (Burity *et al.*, 2013) considered children of 3 years of age able to perform lung function testing and to model spirometric indices. The covariates or anthropometric measurements considered for different studies included age, height, weight, abdominal girth, chest circumference, BMI, sex and ethnicity. Separate equations were fitted for both boys and girls in all the studies. Age and height were found to be most significant variables to model lung function reference equations. This was because in children and adolescents the biological growth rate changes more rapidly with age what is observed with adults (Wang *et al.*, 1993), having an important impact on the development of lungs, as well as lung function. For several years, authors used height (or logarithmic transformation of height) covariate along with the logarithmic transformation of lung functions. Age was then included as a potential covariate along with logarithmic, polynomial or exponential transformation in the model. Recently, splines of age Are being considered to capture the non-linear relationship between age and lung function indices.

In most of the studies, the following formula was used to calculate the LLN:

$$\text{LLN} = \text{Lung Function}_{\text{Predicted}} \times (1 - \text{Residual Standard Deviation})^2$$

Since the GLI approach was introduced, the lower 5^{th} percentile of the predicted lung function values are being used. If the response variable follows a normal distribution, the LLN becomes:

$$\text{LLN} = \text{Lung Function}_{\text{predicted}} - 1.645 \times \text{SD}_{\text{residuals}} \tag{5.1}$$

When the lung function variable follows a Box-Cox-Cole-Green distribution with median μ,

coefficient of variation σ_L and skewness ν, the following formula is used to calculate LLN:

$$\mu(x)\big(1-\sigma_L(x)\nu(x)z_{0.05}\big)^{\frac{1}{\nu(x)}}, \quad \text{for } \nu(x)\neq 0 \tag{5.2}$$

In this study, the focus was on spirometric reference equations of the children and adolescents. Models should be flexible enough to generalize for different scenarios, i.e., deriving equations for all age groups from different ethnic background. Keeping this in mind, Global Lung Function Initiative (GLI) was formed, where lung function data from different ethnic groups of all ages people were collated together to identify a flexible approach (Quanjer *et al.*, 2012). GAMLSS was used to model the spirometric indices for all-age multi-ethnic people. The GLI incorporated only logarithm of height and age, and regression spline of age to model the lung function variables.

Following the extensive literature review, the flexible and efficient statistical method was chosen. The method was applied to obtain the prediction equations for lung function indices. Following the approach provided by the GLI, the GAMLSS was applied to model the lung function indices for First Nations school-going children and adolescents. The lung function indices were assumed to follow a BCCG distribution. This approach gave reliable results for this study. Logarithmic transformation and polynomial bases of age were considered to capture the non-linear relationship between anthropometric and lung function variables. Logarithmic transformation of standing height and lung function values were also included in the models. From different combination of the order of the polynomial, the models giving the lowest values of AIC were selected as reference models. The equations 5.1 and 5.2 were used to calculate the LLN for lung function values. The calculation of LLN are given in Appendix E. One can easily find out the LLN value for FVC, FEV_1 and FEV_1/FVC for boys and girls ages 6-17 years and within the range of the heights (123 cm - 182 cm for boys; 113 - 179 for girls).

Significant differences were observed in lung functions (FVC, FEV_1 and FEV_1/FVC) and anthropometric measurements between both boys and girls. Therefore, fitting separate equations for both the sexes are justified. The approach proposed by Quanjer *et al.* (2012) in Global Lung Function Initiative (GLI) was useful in prediction of lung functions for First Nations children and adolescents. In GLI, polynomial bases of order 6-7 were used for modeling the μ, σ_L and ν. In this thesis, lower order polynomial bases (upto order 4) were enough to obtain the reference models. In girls, the difference between the curves of FVC and FEV_1 was smaller compared to boys. Thus, the total volume of air for girls during exhalation

are close to the volume of air exhaled at the first second. The predicted curves were able to capture the curved relationship between lung function and anthropometric measurements. The LLN values obtained in this study can be used for the assessment of lung function and their response to treatment in First Nations children and adolescents.

5.1 Strength, Limitations and Further Scope

Strengths of the study:

This is the first study to derive spirometric reference equations for First Nations children and adolescents living in Canada. The equations derived by Hankinson *et al.* (1999) for Caucasians may not be applicable to the lung function data of Canadian First Nations children and adolescents. The recent model (i.e., GAMLSS), which is used globally to model lung function values was applied in this study to obtain the reference models for lung function indices. The predicted values and LLN values can be provided to clinicians to assess lung function, severity of respiratory diseases and their response to treatment of the First Nations children and adolescents.

To ensure usable lung function information in this study, all lung function assessments were reviewed by a respirologist to confirm their acceptability for use in deriving prediction equations. Lung function testing was conducted by nurses who were certified in lung function testing.

Limitations of the study:

The study has a smaller sample size that is recommended for the prediction of lung function indices. According to the recommendation by GLI, to derive equations for the people of new ethnic group, a minimum group size should be 150 males and 150 females (Quanjer *et al.*, 2015). In this study, it was only 53 males and 77 females, which implies that the sample size is smaller than the GLI recommendation; therefore, the results of this study cannot be incorporated in the software of GLI. Height was not included in the model of coefficient of variation (σ_L) and skewness (ν). σ_L and ν involving height is hard to detect unless the sample size is relatively large (Cole *et al.*, 2007), as the change in height is more variable than age.

There was no specific equation form for the splines of lung function reference equations for younger age group in the studies of Quanjer *et al.* (2012). Therefore, graphical comparison between the Quanjer *et al.* study and this study is not possible.

There is tremendous diversity among First Nations peoples in Canada. This work considered two Cree First Nations people in Saskatchewan. The results from this study may not be generalizable to other First Nations groups in Canada, including Cree people in other regions.

Lastly, smoking rates were high in the study sample. Thus, obtaining a non-smoking healthy sample was challenging for this particular population. We cannot guarantee that children included in this analysis, although non-smokers, were not indirectly exposed to significant amounts of second-hand smoke.

Further scope of this study:

Further studies with other Canadian First Nations communities should be conducted to derive spirometric reference equations for children and adolescents.

In future study, it will be useful to increase the sample size of the children and adolescents or by including those 3 to 5 years, adults and older people. That way the sample size can be increased and the model can then be generalized for most First Nations ages. In the GLI model people from 3-95 were considered, whereas, in this study people ages 6-17 years were considered. Thus another model including the people of this age group and beyond the age range may generalize the model and can be recommended to the GLI software to predict lung function values for First Nations people. Moreover, considering a large sample size may reduce the variability in height. Studies should be conducted to see if the the modeling of coefficient of variation and/or skewness involve height for any of the spirometric indices.

Forced Expiratory Flow ($FEF_{25\%-75\%}$) and Peak Expiratory Flow Rate (PEFR) are important lung function indices in children (Coates *et al.*, 2013). There are numerous number of studies available for deriving the reference equations for FEF and PEFR. A new study may consider modeling of FEF and PEFR for First Nations children and adolescents to calculate the LLN.

Human growth pattern changes with time. Today's First Nations people may not have the same physical stature 10 years later. Therefore, new reference range should be re-evaluated for First Nations children and adolescents in future for a better assessment of the lung function.

More advanced study can be done to define ethnicity based on genetic information rather than self-declaration (Race, Ethnicity, and Genetics Working Group, 2005; Hunt, 2007).

This is a baseline study of the First Nations Lung Health Project (FNLHP). There will be follow-up study of the FNLHP- which will include additional children, adolescents

and adults along with the current participants. Thus another sets of spirometric reference equations can be derived from the next follow-up study and compared with the findings from the current study. As well these equations should be used with other First Nations children and adolescents studies across Canada where they can be re-evaluated or confirmed.

In conclusion, spirometric reference values (FVC, FEV_1 and FEV_1/FVC), predicted values and LLN values are now available for Cree First Nations children and adolescents living in rural Saskatchewan, Canada. Until other equations are obtained, the results from this study can be used to assess the lung function and their response to treatment in children and adolescents.

References

Akaike, H. (1973). Information theory and an extension of the maximum likelihood principle. In: B. N. Petrov, and F. Csáki, eds., *2nd International Symposium on Information Theory*, Akadémia Kiadó, Budapest, pp. 267–281.

Akantziliotou, K., Rigby, R. A. & Stasinopoulos, D. M. (2002). The R implementation of Generalized Additive Models for Location, Scale and Shape. In: Stasinopoulos, M. and Touloumi, G. (eds.), *Statistical modelling in Society: Proceedings of the 17th International Workshop on statistical modelling* (pp. 75-83). Chania, Greece.

American Thoracic Society & others (1991). Lung function testing: selection of reference values and interpretative strategies. *The American Review of Respiratory Disease.* **144(5)**, 1202–1218.

Azizi, B. H. O., & Henry, R. L. (1994). Ethnic differences in normal spirometric lung function of Malaysian children. *Respiratory Medicine*, **88**, 349–356.

Banerji, A., Bell, A., Mills, E. L., McDonald, J., Subbarao, K., Stark, G., Eynon, N., & Loo, V. G. (2001). Lower respiratory tract infections in Inuit infants on Baffin Island. *Canadian Medical Association Journal.* **164(13)**, 1847–1850.

Bjure, J. (1963). Spirometric Studies in Normal Subjects IV. Ventilatory Capacities in Healthy Children 7-17 Years of Age. *Acta paediatrica*, **52(3)**, 232–240.

Bougrida, M., Bourahli, M.K., Aissaoui, A., Rouatbi, S., Mehdioui, H., & Ben Saad, H. (2012). Spirometric reference values for children living in Constantine (Eastern region of Algeria). *La Tunisie Médicale*, **90(1)**, 51–61.

Box, G.E.P. & Cox, D.R. (1964). An analysis of transformations. *Journal of the Royal Statistical Society, Series B.*, **26**, 211–252.

Budhiraja, S., Singh, D., Pooni, P. A., & Dhooria, G., S. (2010). Pulmonary functions in normal school children in the age group of 6-15 years in North India. *Iranian Journal of Pediatrics*, **20(1)**, 82–90.

Bulkow, L. R., Singleton, R. J., Debyle, C., Miernyk, K., Redding, G., Hummel, K. B., Chikoyak, L., & Hennessy, T. W. (2012). Risk factors for hospitalization with lower respiratory tract infections in children in rural Alaska. *Pediatrics*, **129(5)**, 1220–1227.

Burity, E. F., Pereira, C. A. C., Rizzo, J. A., Brito, M. C. A., & Sarinho, E. C. S. (2013). Reference values for spirometry in preschool children. *Journal de Pediatria*, **89(4)**, 374–380.

Burnham, K. P., & Anderson, D. R. (2002). Model Selection and Multimodel Inference: a practical information-theoretic approach, 2nd edition. Springer-Verlag, New York.

Chen, Y., Rennie, D., Cormier, Y., & Dosman, J.A. (2005). Sex specificity of asthma associated with objectively measured body mass index and waist circumference: the Humbolt Study. *Chest*, **128(4)**, 3048–3054.

Choudhuri, D., & Sutradhar, B. (2015). Pulmonary function of adolescents from Tripura, a North-eastern state of India. *Lung India*, **32(4)**, 353–358.

Chowgule, R. V., Shetye, V. M., & Parmar, J. R. (1995). Lung function tests in normal Indian children. *Indian Pediatrics*, **32(2)**, 185–191.

Ciprandi, G., & Cirillo, I. (2010). Forced expiratory flow between 25% and 75% of vital capacity may be a marker of bronchial impairment in allergic rhinitis. *Journal of Allergy and Clinical Immunology*, **127(2)**, 549-549.

Coates, A. L., Graham, B. L., McFadden, R. G., McParland, C., Moosa, D., Provencher, S., & Road, J. (2013). Spirometry in primary care. *Canadian Respiratory Journal*, **20(1)**, 13–22.

Cotes, J.E., Dabbs, J. M., Hall, A. M., & Heywood, C. (1979).Sitting height, fat-free mass and body fat as reference variables for lung function in healthy British children: Comparison with stature. *Annals of Human Biology*, **6(4)**, 307–314.

Cole, T.J. (1975). Linear and Proportional Regression Models in the Prediction of Ventilatory Function. *Journal of the Royal Statistical Society. Series A (General)*, **138(3)**, 297–338.

Cole, T.J. (1988). Fitting smoothed centile curves to reference data (with Discussion). *Journal of the Royal Statistical Society. Series A.*, **151**, 385–418.

Cole, T.J., & Green, P.J. (1992). Smoothing reference centile curves: the LMS method and penalized likelihood. *Statistics in Medicine*, **11**, 1305–1319.

Cole, T.J., Freeman, J.V., & Preece, M.A. (1998). British 1990 growth reference centiles for weight, height, body-mass index and head circumference fitted by maximum penalized likelihood. *Statistics in Medicine*, **17**, 407–429.

Cole, T.J., Cortina-Borja, M., Sandhu, J., Kelly, F. P., & Pan, H. (2007). Nonlinear growth generates age changes in the moments of the frequency distribution: the example of height in puberty. *Biostatistics*, **9(1)**, 159–171.

Cole, T.J., Stanojevic, S., Coates, A.L., Hankinson, J.L., & Wade, A.M. (2009). Age- and size-related reference ranges: A case study of spirometry through childhood and adulthood. *Statistics in Medicine*, **28**, 880–898.

Coultas, D. B., Howard, C.A., Skipper, B.J., & Samet, J.M. (1988). Spirometric prediction equations for Hispanic children and adults in New Mexico. *American Review of Respiratory Disease*, **138(6)**, 1386–1392.

Clausen, J. L. (1980). Pulmonary function testing: guidelines and controversies. New York: Academic Press.

Crapo, R.O., Morris, A.H., & Gardner, R.M. (1981). Reference spirometric values using techniques and equipment that meet ATS recommendations. *American Review of Respiratory Disease.* **123(6)**, 659–664.

Culver, B. H. (2012). How should the lower limit of the normal range be defined? *Respiratory Care.* **57(1)**, 136–143.

Dickman, M.L., Schmidt, C. D., & Gardner, R. M. (1971). Spirometric standards for noraml children and adolescents (ages 5 years through 18 years). *The American Review of Respiratory Disease*, **104(5)**, 680–687.

Dockery, D.W., Berkey, C.S., Ware, J.H., Speizer, F.E., & Ferris, B.G., Jr. (1983). Distribution of forced vital capacity and forced expiratory volume on one second in children 6 to 11 years of age. *The American Review of Respiratory Disease*, **131**, 511–522.

Doctor, T. H., Trivedi, S. S., & Chudasama, R. K. (2010). Pulmonary function test in healthy school children of 8 to 14 years age in south Gujarat region, India. *Lung India*, **27(3)**, 145–148.

Estey, E. A., Kmetic, A. M., & Reading, J. (2007). Innovative approaches in public health research: Applying life course epidemiology to Aboriginal health research. *The Canadian Journal of Public Health*, **98**, 444–446.

Fredriks, A.M., van Buuren, S., Burgmeijer, R.J., Meulmeester, J.F., Beuker, R.J., Brugman, E., Roede, M.J., Verloove Vanhorick, S.P., & Wit, J.M. (2000). Continuing positive secular growth change in The Netherlands 19551997. *Pediatric Research*, **47**, 316–323.

Gao, Z., Rowe, B. H., Majaesic, C., OHara, C., & Senthilselvan, A. (2008). Prevalence of asthma and risk factors for asthma-like symptoms in Aboriginal and non-Aboriginal children in the northern territories of Canada. *Canadian Respiratory Journal*, **15(3)**, 139–145.

Gore, C.J., Crockett, A.J., Pederson, D.G., Booth, M.L., Bauman, A., & Owen, N. (2012). Spirometric standards for healthy adult lifetime nonsmokers in Australia. *European Respiratory Journal*, **8(5)**, 773–782.

Green, P.J. (1987). Penalized likelihood for general semiparametric regression model. *International Statistical Review*, **55**, 245–259.

Gutierrez, C., Ghezzo, R. H., Abboud, R. T., Cosio, M. G., Dill, J. R., Martin, R. R., McCarthy, D. S., Morse, J. L., & Zamel, L. (2004). Reference values of pulmonary function tests for Canadian Caucasians. *Canadian Respiratory Journal* **11(6)**, 414–424.

Hankinson, J.L., Odencrantz, J.R., & Fedan, K.B. (1999). Spirometric reference values from a sample of the general U.S. population. *American Journal of Respiratory and Critical Care Medicine*, **159(1)**, 179–187.

Hastie, T., & Tibshirani, R. (1986). Generalized additive models (with discussion). *Statistical Science*, **1**, 297–310.

Hastie, T., & Tibshirani, R. (1990). *Generalized Additive Models*. Chapman and Hall.

Hsi, B. P., Hsu, K. H., & Jenkins, D. E. (1983). Ventilatory functions of normal children and young adults: Mexican-American, white, and black. III. Sitting height as a predictor. *The Journal of Pediatrics*, **102**, 860–865.

Hsu, K.K.H., Jenkins, D.E., Hsi, B.P., Bourhofer, E., Thompson, V., Hsu, F.F.C., & Jacob, S.C (1979). Ventilatory functions of normal children and young adults–Mexican-American, white, and black. II. Wright peak flowmeter. *The Journal of Pediatrics*, **95(2)**, 183–190.

Huizinga, J., & Glanville, E. V. (1968). Vital capacity and timed vital capacity in the Kurumba from Upper Volta. *South African Journal of Science*, **64**, 125–133.

Hunt, L. M., & Megyesi, M. S. (2007). The ambiguous meanings of the racial/ethnic categories routinely used in human genetics research. *Social Science and Medicine*, **66**, 349–361.

Wesley, A. G., Pather, M., & Becker, P. (1987). Ethnic variation in respiratory morbidity and lung function in childhood. *Thorax*, **42**, 542–548.

Karkhanis, V.S., & Joshi, J.M. (2012). Spirometry in Chronic Obstructive Lung Disease (COPD). *Journal of the Association of Physicians in India*, **60**, 22–26.

Karunanayake, C. P., Dosman, J. A., Hagel, L., Rennie, D. C., Lawson, J. A., Pahwa, P., & Saskatchewan Rural Health Study Group (2015). Reference values of pulmonary function tests for rural Canadians. *International Journal of Respiratory and Pulmonary Medicine*, **2:021**.

Kirkby, J., Bonner, R., Lum, S., Bates, P., Morgan, V., Strunk, R.C., Kirkham, F., Sonnappa, S. & Stocks, J. (2013).Interpretation of pediatric lung function: Impact of ethnicity. *Pediatric Pulmonology* **48(1)**, 20–26.

Kory, R.C., Callahan, R., Boren, H.G., & Syner, J.C. (1961). The Veterans Administration-Army cooperative study of pulmonary function. I. Clinical spirometry in normal men. *The American Journal of Medicine*, **30**, 243–258.

Koopman, M., Zanen, P., Kruitwagen, C.L., van der Ent, C.K., & Arets, H.G. (2011). Reference values for paediatric pulmonary function testing: The Utrecht dataset. *Respiratory Medicine*, **105(1)**, 15–23.

Kuczmarski, R.J., Ogden, C.L., Guo, S.S., Grummer-Strawn, L.M., Flegal. K.M., Mei, Z., Wei, R., Curtin, L.R., Roche, A.F., & Johnson, C.L. (2002). *2000 CDC Growth Charts for the United States: Methods and Development*. National Center for Health Statistics: Hyattsville, MD, 2002.

MacMillan, H., L., Jamieson, E., Walsh, C., Boyle, M., Crawford, A., & MacMillan, A. (2010). The health of Canada's Aboriginal children: results from the First Nations and Inuit Regional Health Survey. *International Journal of Circumpolar Health*, **69(2)**, 158–167.

McCuskee, S., Kirlew, M., Kelly, L., Fewer, S., & Kovesi, T. (2014). Bronchiolitis and pneumonia requiring hospitalization in young first nations children in Northern Ontario, Canada. *Pediatric Infectious Disease Journal.* **33(10)**, 1023–1026.

Miller, G. J., Saunders, M. J., Gilson, R. J. C., & Ashcroft, M. T. (1977). Lung function of healthy boys and girls in Jamaica in relation to ethnic composition, test exercise performance, and habitual physical activity. *Thorax*, **32**, 486–496.

Miller, M.R., Hankinson, J., Brusasco, V., Burgos, F., Casaburi, R., Coates, A., Crapo, R., Enright, P., van der Grinten, C.P.M., Gustafsson, P., Jensen, R., Johnson, D.C., Maclntyre, N., McKay, R., Navajas, D., Pedersen, O.F., Pellegrino, R., Viegi, G., & Wanger, J. (2005). Standardisation of spirometry. *European Respiratory Journal*, **26**, 319-338.

Moore, V. C. (2012). Spirometry: step by step. *European Respiratory Society*, **8(3)**, 233–240.

Nelder, J. A., & Wedderburn, R. W. M. (1972). Generalized linear models. *Journal of the Royal Statistical Society (A).*, **135**, 370–384.

National Center for Health Statistics (1996). NHANES III: Reference Manuals and Reports. Data Dissemination Branch. Hyattsville, MD. CD-ROM No. 6-0178(1096)

Pahwa, P., Abonyi, S., Karunanayake, C., Rennie, D. C., Janzen, B., Kirychuk, S., Lawson, J. A., Katapally, T., McMullin, K., Seeseequasis, J., Naytowhow, A., Hagel, L., Dyck, R. F., Fenton, M., Senthilselvan, A., Ramsden, V., King, M., Koehncke, N., Marchildon, G., McBain, L., Smith-Windsor, T., Smylie, J., Episkenew, J. A., & Dosman, J. A. (2015). A community-based participatory research methodology to address, redress, and reassess disparities in respiratory health among First Nations. *BMC Research Notes.* **8(199)**. doi: 10.1186/s13104-015-1137-5.

Pallegrino, R., Viegi, G., Brusasco, V., Crapo, R.O., Burgos, F., Casaburi, R., Coates, A., van der Grinten, C.P., Gustafsson, J., & Hankinson, J. (2005). Interpretative strategies for lung function tests. *European Respiratory Journal.* **26(5)**, 948–968.

Parma, A., Magliocchetti, N., Spagnolo, A., Di Monaco, A., Migliorino, M.R., & Menotti, A. (1997). Spirometric prediction equations for male Italians 7-18 years of age. *Jornal Brasileiro de Pneumologia*, **33(4)**, 397–406.

Piccioni, P., Borraccino, A., Forneris, M. P., Migliore, E., Carena, C., Bignamini, E., Fassio, S., Cordola, G., Arossa, W., & Bugiani, M. (2007). Reference values of Forced Expiratory Volumes and pulmonary flows in 3-6 year children: a cross-sectional study. *Respiratory Research*, **8(14)**, doi:10.1186/1465-9921-8-14.

Pittman, J. E., & Rosenfeld, M. (2011). Appropriate pediatric spirometry reference equations and interpretation. *Pediatric Allergy, Immunology, and Pulmonology*, **24(2)**. doi: 10.1089/ped.2011.0075.

Quanjer, P.H., Borsboom, G.J.J.M., Brunekreef, B., Zach, M., Forche, G., Cotes, J.E., Sanchis, J., & Paoletti, P. (1995). Spirometric reference values for white European children and adolescents: Polgar revisited. *Pediatric Pulmonology*, **19**, 135–142.

Quanjer, P.H., Hall, G.L., Stanojevic, S., Cole, T.J., Stocks, J. & Global Lungs Initiative (2012). Age- and height-based prediction bias in spirometry reference equations. *European Respiratory Journal*, **40(1)**, 190–197.

Quanjer, P.H., Stanojevic, S., Cole, T.J., Baur, X., Hall, G.L., Culver, B.H., Enright, P.L., Hankinson, J.L., Ip, M.S.M., Zheng, J., Stocks, J., & the ERS Global Lung Function Initiative (2012). Multi-ethnic reference values for spirometry for the 3-95-yr age range: the global lung function 2012 equations. *The European Respiratory Journal*, **40(6)**, 1324–1343.

Quanjer, P.H., Stanojevic, S., Stocks, J. & Cole, T.J., (2012). GLI-2012 All-Age Multi-Ethnic Reference Values for Spirometry: Advantages and Disadvantages. URL: `http://www.lungfunction.org/`.

Quanjer, P.H., Stanojevic, S., Cole, T.J., & Stocks, J. (2015). Implementing GLI-2012 regression equations. *The European Respiratory Society*, Version 19 July 2015.

R Core Team (2015). R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing, Vienna, Austria. URL https://www.R-project.org/.

Race, Ethnicity, and Genetics Working Group (2005). The use of racial, ethnic, and ancestral categories in human genetics research. *American Journal of Human Genetics*, **77(4)**, 519–532.

Rajkapoor, K. K., Mahajan, A., & Mahajan, A. (1997). Ventilatory lung function tests in school children of 6-13 years. *Indian Journal of Chest Diseases and Allied Sciences*, **39(2)**, 97–105.

Rigby, R.A., & Stasinopoulos, D.M. (2001). The GAMLSS project: a flexible approach to statistical modelling. In: Klein, B. and Korsholm, L. (eds.), *New Trends in Statistical Modelling: Proceedings of the 16th International Workshop on Statistical Modelling* (pp. 249–256). Odense, Denmark.

Rigby, R.A., & Stasinopoulos, D.M. (2004). Smooth centile curves for skew and kurtotic data modelledusing the BoxCox power exponential distribution. *Statistics in Medicine*, **23**, 3053–3076.

Rigby, R.A., & Stasinopoulos, D.M. (2005). Generalized additive models for location, scale and shape (with Discussion). *Applied Statistics*, **54**, 507–544.

Rigby, R.A., & Stasinopoulos, D.M. (2010). A flexible regression approach using GAMLSS in R. *University of Athens*. URL: http://www.gamlss.org/wp-content/uploads/2013/01/Lancaster-booklet.pdf

Rochat, M. K., Laubender, R. P., Kuster, D., Braendli, O., Moeller, A., Mansmann, U., Mutius, E., & Wildhaber, J. (2013). Spirometry reference equations for central European populations from school age to old age. *PLoS One*, **8(1)**: e52619. doi: 10.1371/journal.pone.0052619.

Roizin, H., Szeinberg, A., Tabachnik, E., Molho, M., Benzaray, S., Augarten, A., Har-Even, D., Barzilay, Z., & Yahav, J. (1993). Ethnic differences in lung function in Israeli children. *Thorax*, **48(9)**, 906–910.

Rosenfeld, M., Pepe, M. S., Longton, G., Emerson, J., FitzaSimmons, S., & Morgan, W. (2001). Effect of choice of reference equation on analysis of pulmonary function in cystic fibrosis patients. *Pediatric Pulmonology*, **31**, 227–237.

Rosenthal, M., Cramer, D., Bain, S.H., Denison, D., Bush, A., & Warner, J. O. (1993). Lung function in white children aged 4 to 19 years: II-Single breath analysis and plethysmography. *Thorax*, **48**, 803–808.

Sahebjami, H., & Gartside, P.S. (1996). Pulmonary function in obese subjects with a normal FEV1/FVC ratio. *Chest*, **110(6)**, 1425–9.

Schoenberg, J. B., Beck, G. J., & Bouhuys, A. (1978). Growth and decay of pulmonary function in healthy blacks and whites. *Respiratory Physiology*, **33**, 367–393.

Seed, L., Wilson, D., & Coates, A. L. (2012). Children should not be treated like little adults in the PFT lab. *Respiratory Care*, **57(1)**, 61–71.

Shamssain, M. H., Thompson, J., & Ogston, S. A. (1988). Forced expiratory indices in normal Libyan children aged 6-19 years. *Thorax*, **43**, 467–470.

Shamssain, M. H. (1991). Forced expiratory indices in normal black Southern African children aged 6-19 years. *Thorax*, **43**, 467–470.

Simon, M.C., Chinchilli, V.M., Phillips, B.R., Sorkness, C.A., Lemanske Jr., Robert, F., Szefler, Stanley, J., Taussing, L., Bacharier, L.B., & Morgan, W. (2010). Forced expiratory flow between 25% and 75% of vital capacity and FEV1/forced vital capacity ratio in relation to clinical and physiological parameters in asthmatic children with normal FEV1 values. *Journal of Allergy and Clinical Immunology*, **126(3)**, 527–534.

Singleton, R., Morris, A., Redding, G., Poll, J., Holck, P., Martinez, P., Kruse, D., Bulkow, L. R., Petersen, K. M., & Lewis, C. (2000). Bronchiectasis in Alaska Native children: causes and clinical courses. *Pediatric Pulmonology.* **29(3)**, 182–187.

Sin, D. D., Sharpe, H. M., Cowie, R. L., & Man, S. F.; Alberta Strategy to Help Manage Asthma Executive Committee (2004). Spirometric findings among school-aged First Nations children on a reserve: a pilot study. *Canadian Respiratory Journal*, **11(1)**, 45–48.

Stanojevic, S., Wade, A., Stocks, J., Hankinson, J., Coates, A.L., Pan, H., Rosenthal, M., Corey, M., Lebecque, P., & Cole, T.J. (2008). Reference ranges for spirometry across all ages: a new approach. *American Journal of Respiratory and Critical Care Medicine*, **177(3)**, 253-260.

Stasinopoulos, D.M. & Rigby, R.A. (2007). Generalized additive models for location scale and shape (GAMLSS). *Journal of Statistical Software*, **23(7)**.

Statistics Canada (2011). *Aboriginal Peoples in Canada: First Nations People, Métis and Inuit*. Statistics Canada Catalogue no. 99-011-X2011001. Ottawa, Ontario. May 8. http://www12.statcan.gc.ca/nhs-enm/2011/as-sa/99-011-x/99-011-x2011001-eng.cfm (accessed October 20, 2015).

Subbarao, P., Lebecque, P., Corey, M., & Coates, A.L. (2004). Comparison of spirometric reference values. *Pediatric Pulmonology*, **37(6)**, 515–522.

Swanney, M.P., Ruppel, G., Enright, P.L., Pedersen, O.F., Crapo, R.O., Miller, M.R., Jensen, R.L., Falaschetti, E., Schouten, J.P., Hankinson, J.L., Stocks, J., & Quanjer, P.H. (2008). Using the lower limit of normal for the FEV1/FVC ratio reduces the misclassification of airway obstruction. *Thorax*, **63(12)**, 1046–1051.

Tan, W.C., Bourbeau, J., Hernandez, P., Chapman, K., Cowie, R., FitzGerald, M.J., Aaron, S., Marciniuk, D.D., Maltais, F., O'Donnell, D.E., Goldstein, R., & Sin, D.; LHCE study investigators (2011). Canadian prediction equations of spirometric lung function for Caucasian adults 20 to 90 years of age: Results from the Canadian Obstructive Lung Disease (COLD) study and the Lung Health Canadian Environment (LHCE) study. *Canadian Respiratory Journal*, **18(6)**, 321–326.

Trabelsi, Y., Pariés, J., Harrabi, I., Zbidi, A., Tabka, Z., Richalet, J. P., & Buvry, A. (2008). Factors affecting the development of lung function in Tunisian children. *American Journal of Human Biology*, **20**, 716–725.

Veale, A.J., Peat, J.K., Salome, C.M., Woolcock, A.J., & Thompson, J.E. (1997). 'Normal' lung function in rural Australian aborigines. *Australian and New Zealand Journal of Medicine*, **27(5)**, 543–549.

Vijayan, V.K., Reetha, A.M., Kuppurao, K.V., Venkatesan, P. & Thilakavathy, S. (2000). Pulmonary function in normal South Indian children aged 7 to 19 years. *Indian Journal of Chest Disease and Allied Science*, **42**, 147–156.

Wall, M.A., Olson, D., Bonn, B.A., Creelman, T., & Buist, A.S. (1982). Lung function in North American Indian children: reference standards for spirometry, maximal expiratory flow volume curves, and peak expiratory flow. *The American Review of Respiratory Disease*, **125(2)**, 158–162.

Wang, X., Dockery, D.W., Wypij, D., Fay, M.E., & Ferris, B.G. Jr (1993). Pulmonary function growth velocity in children between 6 to 18 years of age. *Pediatric Pulmonology*, **15(2)**, 75–88.

Watson, D. S., Watson, R. M., & Siskind, V. (1986). Respiratory function in aboriginal children. *Medical Journal of Australia.* **144**, 11–13.

Wesley, A. G., Pather, M., & Becker, P. (1989). Normal values for simple lung function tests in South African Asian children. *Annals of Tropical Paediatrics*, **9(2)**, 70–74.

White, D. (1994). Pediatric Pulmonary Function Testing. In Ruppel, G. (Ed.), *Manual of pulmonary function testing* (pp 273-315). *London, United Kingdom: Mosby.*

Wilson, K., Rosenberg, S., Abonyi, S., & Lovelace, R. (2010). Aging and health: An examination of differences between older Aboriginal and non-Aboriginal people. *The Canadian Journal on Aging*, **29(3)**, 369–382.

Wood, S. N. (2006). *Generalized additive models : an introduction with R.* Chapman & Hall/CRC.

Wypij, D., Pugh, M., & Ware, J. H. (1993). Modeling pulmonary function growth with regression splines. *Statistica Sinica*, **3**, 329-350.

Yang, T. S., Peat, J., Keena, V., Donnelly, P., Unger, W., & Woolcock, A. (1991). A review of the racial differences in the lung function of normal Caucasian, Chinese and Indian subjects. *European Respiratory Journal*, **4(7)**, 872–880.

Yao, G. Y., Ma, Y. L., Zhang, M. Q., & Gao, Z. C. (2013). Macrolide therapy decreases chronic obstructive pulmonary disease exacerbation: a meta-analysis. *Respiration*, **86(3)**, 254-260.

Appendix A

Survey Questionnaire

First Nations Lung Health Project

UNIVERSITY OF SASKATCHEWAN

TANSI PARENTS AND CARE GIVERS:

Two First Nations communities and the University of Saskatchewan wish to learn more about breathing problems in First Nations children. Families [redacted] First Nation are participating.

PLEASE COMPLETE ONE SURVEY FOR EACH CHILD ATTENDING GRADES 1 TO GRADE 12 WHO ARE AGES 6-17 YEARS. This survey has 4 parts. Please try to answer all of the questions, but you don't have to answer questions that you choose not to or feel uncomfortable to answer.

Part One asks about contact information.

Part Two asks about the child's or young person's past and present health.

Part Three asks about the child's or young person's lifestyle and surroundings.

Part Four asks questions about the child's or young person's personal and family history.

ALL INFORMATION WILL BE KEPT CONFIDENTIAL.

WHEN YOU HAVE FINISHED, PLEASE PLACE THE SURVEY IN THE BROWN ENVELOPE, SEAL THE ENVELOPE AND RETURN IT TO THE SCHOOL.

Teniki

INSTRUCTIONS

The questions can be answered by checking the best answer or by filling in the blank with a number or words.

EXAMPLE 1:

Does your child usually have a cough at night ?

Yes ✓ No ___

EXAMPLE 2:

How many years has your child lived in this home?

6 Years

PART ONE – CONTACT INFORMATION

Child's First Name _____

Last Name _____

Date of Birth (Month/Day/Year) _____

Mother's name_____ Father's name _____

House Number Phone Number

Location of home: _____[redacted] _____ _____

_____[redacted] _____ _____

_____Other

Person completing questionnaire:

Mother _____

Father _____

Caregiver _____ Relationship to child _____

1

PART TWO – HEALTH OF THIS CHILD

Cough

1. Has your child ever had a dry cough at night or first thing in the morning **NOT** because of a cold or chest infection? *Tick all that apply*

   ___Yes, past 12 months

   ___Yes**, before** the last 12 months

   ___No

2. Does this child usually cough at all during the rest of the day?

   No ___ Yes ___ Don't know ___

3. **In the past 12 months**, has this child woken up because of a cough?

   No ___ Yes ___ Don't know ___

Congestion and Phlegm

4. Does this child usually have tightness in the chest or bring up phlegm or mucus **apart from colds?**

   No ___ Yes ___ Don't know ___

   If **YES**, has this congestion or phlegm been present for 3 months in a row? *Tick all that apply*

   ___ Yes, past 12 months

   ___ Yes, before last 12 months

   ___ No

Wheezing

5. Has this child **ever** had a wheeze or whistling noise that comes from the chest?

   No ___ Yes ___ Don't know ___

   If **YES**, at what age did this child **first start** to wheeze?

   ___ years

6. **In the past 12 months**, has this child had a wheeze or whistling noise that comes from the chest?

   No ___ Yes ___ Don't know ___

   *****IF **NO,** at what age did this child stop wheezing?

   ___ years GO TO **QUESTION 10.**

   *****IF **YES,** CONTINUE ON TO **QUESTION 7.**

7. Does the wheezing or whistling in the chest happen:

   ___ without colds?

   ___ with colds?

   ___ with and without colds?

8. How many attacks of wheezing or whistling in the chest has this child had **in the past 12 months**?

___ none

___ 1-3

___ 4-12

___ more than 12

9. Does wheezing or whistling in the chest happen most nights or days?

___ Nights only

___ Days only

___ Both nights and days

10. Has your child's chest ever sounded wheezy during or after play/exercise/sports? *Tick all that apply*

___ Yes, past 12 months

___ Yes, before last 12 months

___ No

Asthma

11. Has this child ever been diagnosed by a doctor as having asthma?

No ___ Yes ___ Don't know ___

IF NO or DON'T KNOW, Please go to **Question 18.**

IF YES, continue at **Question 12.**

12. At what age was the asthma first diagnosed?

___ years of age

13. **In the past 12 months,** has this child required care for asthma from the following places:

Hospital inpatient	No__ Yes __
Emergency room outpatient	No__Yes __
Reserve Health Centre	No__ Yes __
Doctor's office	No__ Yes __

14. **In the past 12 months**, how many asthma experiences has your child had?

___ (number of experiences)

An asthma experience could be shortness of breath or wheezing with or without coughing

15. Did your child receive medicine for these experiences?

No ___ Yes ___

If YES, what was the medicine?

_____

16. **In the past 12 months, how many times has this child taken medicine for asthma:**

__ Never in the past 12 months

__ At least once in the past 12 months

__ At least once per month

__ At least once per week

__ Every day

17. Has your child's sleep been bothered by asthma **in the past 12 months:**

No ___ Yes ___

If yes, how many times: _____

Allergic disease

18. Has this child ever had an allergy (hives, runny nose, itchiness)?

No ___ Yes ___

19. Has this child ever had an allergy (e.g. hives, runny nose, sneezing and/or wheezing) to any of the following:

House dust	No □ Yes □
Grain dust	No □ Yes □
Pollen	No □ Yes □
Trees	No □ Yes □
Grasses	No □ Yes □
Mold or mildew	No □ Yes □
Dog	No □ Yes □
Cat	No □ Yes □
Birds/feathers	No □ Yes □
Foods	No □ Yes □

If **YES**, what food(s)? _____

20. Has your child ever had a problem with sneezing, or a runny, or a blocked nose when he/she did **NOT** have a cold or the flu?

No ___ Yes ___

Other Illness and past illness

21. Is this child **regularly** taking medicine that your doctor prescribed for a breathing problem?

No ___ Yes ___ Don't know ___

If **YES**, please name the medicine(s) below:

_____

_____

_____

22. **In the past 12 months** has the child been given antibiotics for breathing problems?

No ___ Yes ___ Don't know ___

23. **In the past 12 months** has this child been kept at home from school for 3 or more days with a chest sickness?

No ___ Yes ___ Don't know ___

24. Has the parent's or caregiver's sleep been bothered because of this child's asthma in the past 12 months?

No ___ Yes ___

25. Has a doctor ever said this child had any of the following illnesses:

Tonsillitis	No ___ Yes___
Bronchitis	No ___ Yes___
Pneumonia	No ___ Yes___
Eczema	No ___ Yes___
Croup	No ___ Yes___
Ear infection	No ___ Yes___
Sleep apnea (breathing stops during the sleep)	No ___ Yes___
Diabetes (high blood sugar)	No ___ Yes___
Heart condition	No ___ Yes___
Whooping cough	No ___ Yes___
Sinus trouble	No ___ Yes___

26. Has this child ever been hospitalized because of breathing problems?

No ___ Yes___

If YES, how many times? _____

At what ages: _____

_____

27. Has this child had an operation to remove the tonsils (adenoids)?

No ___ Yes___

28. Does this child snore when sleeping?

No ___ Yes___

29. On school days this child usually goes to bed at _____pm and gets up at _____am.

On weekends and holidays this child usually goes to bed at _____pm and gets up at _____am.

30. Has this child ever fallen asleep in school?

No_____ Yes_____

31. **In the past 12 months**, did you ever experience any difficulties getting the regular or on-going healthcare for this child?

No ___ Yes ___ Don't know ___

32. How far do you travel (in one direction) to get ongoing health care for this child? ____ km

33. How far do you travel (in one direction) to receive emergency health services for this child? ____ km

PART THREE – LIFESTYLE & ENVIRONMENT

34. On a normal day, is the MAIN part of your child's trip to school made by…? (**Please tick one box only**)

___ Walking

___ Bicycle

___ School bus

___ Other way

Please specify _____

35. How many days per month would this child eat wild meat (deer, moose, birds, rabbit)?

_____days

36. Does this child's father smoke today?

No ___ Yes ___

If No but he has smoked, what year did he quit smoking? ____

37. Does this child's mother smoke today?

No ___ Yes ___

If No but she has smoked, what year did she quit smoking? ___

38. Does this child/adolescent smoke today?

No ___ Yes ___ Don't know ___

If yes how many years_____

39. Do any of this child's/adolescent's friends smoke in front of her/him?

No ___ Yes ___ Don't know ___

40. Do any people who live in your house smoke in the house?

No ___ Yes ___ Don't know ___

41. How many people regularly smoke cigarettes in the house?

_____number of persons who usually live in the house

_____ number of regular visitors

42. On average, how many cigarettes are smoked in your home a day? (Please think of everyone who smokes in your home)

_____ Cigarettes/day

43. Do people smoke while this child is in the car?

No ___ Yes ___ Don't know ___

44. During a normal week, how many days was this child physically active for at least 60 minutes per day?

____ days

45. During a normal week, how many hours a day (24 hours) does your child watch TV or play video games?

□ Less than 1 hour

□ 1 hour but less than 3 hours

□ 3 hours but less than 5 hours

□ 5 hours or more

46. How many times per week would this child eat chips, candy or pop?

_____times

47. How does this child like to spend their time after school?

The child's home

48. How long has your child lived in this house?

__ years

49. Which best describes the type of housing unit in which your family lives?

□ one family house

□ other, please specify: _____

50. How many rooms are there in the home (not including bathrooms, porches or hallways)?

____ number

51. How many people live in the home?

____ number

52. In your house, what fuel is usually used for heating?

□ Natural gas/central heating

□ Electricity

□ Wood

□ Other, please specify _____

53. **In the past 12 months**, have you had any problems with mice or pests in your home?

No ___ Yes ___ Don't know ___

54. Do you have any of the following in your home?

Air conditioners	No □	Yes □
Air filter	No □	Yes □
Humidifier (adds moisture)	No □	Yes □
Dehumidifier (takes away dampness)	No □	Yes □
Wood fireplace	No □	Yes □

55. Does your house have any damage caused by dampness (e.g., wet spots on walls, floors)?

No ___ Yes ___ Don't know ___

56. Are there signs of mold or mildew in any living areas in your home?

No ___ Yes ___ Don't know ___

57. **During the past 12 months**, has there been water or dampness in your home from broken pipes, leaks, heavy rain, or floods?

No ___ Yes ___ Don't know ___

58. **In the past 12 months**, have you had any of the following pets living in your home?

Cat	□ No	□ Yes
Dog	□ No	□ Yes
Bird	□ No	□ Yes

PART FOUR – THIS CHILD AND THE FAMILY HISTORY

59. Child's sex: Male____ Female____

60. Child's age: _____

61. How tall is your child? (*For best results please use a tape measure against a wall*)

_____ feet _____ inches

62. How much does your child weigh?

_____ pounds

63. Do you consider your child to be:

Underweight? ___

Just about right weight? ___

Overweight? ___

64. What was the child's weight at birth?

___ pounds ___ ounces or ____ kg

65. Was your child born by caesarean section operation?

_____No _____ Yes

66. What was the mother's age at the time of birth of this child? _____

67. Is this child the first born child in the family?

____ No ____ Yes

68. How many children are in the family?

____children

69. Was this child breastfed?

No ___ Yes ___ Don't know ___

If **YES,** at what age did breast feeding end?

_____months or _____ years

70. From birth, how many times did your child move on and off your reserve/First Nation community?

____Never

___ Once

___ 2-3 times

___ 4-5 times

___ 6 or more times

71. At the end of the month, how much money do you have left over?

□ Some money

□ Just enough money

□ Not enough money

72. Type of household

□ Single parent home

□ Two parent /partner home

.

73. Did this child's mother smoke during the pregnancy of this child? *(Please check all that apply)*

____ No ____ Yes _____Don't know

74. Does the child's <u>birth</u> mother or father have any of the following conditions? *Tick any that apply*

	Mother	Father
Asthma		
Hayfever		
Allergies		
Eczema		
Don't know		

75. What is this child's mother's/father's/caregiver's highest level of education?

	Mother	Father	Caregiver
Less than Grade 12			
Completed Grade 12 or higher			

THE END

Thank you for completing the questionnaire.
Please make any comments you wish below.

Appendix B

Clinical Assessment

CHILDREN'S CLINICAL ASSESSMENT FORM

FIRST NATIONS LUNG HEALTH PROJECT CLINICAL ASSESSMENT – CHILDREN'S EVALUATION

Background information Date:_____
month/day/year

Study ID:_____ Date of Birth _____Sex_____
month/day/year

Name:_____

Height:_____(cm) Location:_____

Weight:_____ (kg) Time:_____

Abdominal girth:_____(cm) Tester's Initials:_____

Blood Pressure: 1st systolic _____ (mmHg) diastolic _____ (mmHg)
2nd systolic _____ (mmHg) diastolic_____ (mmHg)

Pulmonary Function Test

Exclusion Criteria	**YES**	**NO**
Do you have a cold or a cough today? (**If yes, rebook tests**)		
Have you used a blue inhaler in the last 2 hours? (**If yes, rebook PFT in 4 hours**)		
Have you taken an allergy pill or cough syrup? (**If yes, rebook allergy test**)		
Do you have a headache today? (If yes, rebook the PFT)		
Have you ever stayed in the hospital over night? (**If yes, call parents to obtain further information and then consult with investigators before testing**)		
Do you see your doctor a lot? (**If yes, call parents to obtain further information and then consult with investigators before testing**)		

Lung function testing was:
_____ Completed **(ATTACH SPIROMETRY RESULTS TO THIS FORM)**
_____ Not completed; If **NOT**, then why: _____ Subject could not perform test
_____ Refused
_____ Other

Nurses' Comments (enter comments from results sheet into database as well):

Allergy Skin Prick Test

Exclusion Criteria	**YES**	**NO**
Have you taken an allergy pill or cough syrup today? (**If yes, rebook allergy test**)		
Does the child have eczema or a rash? (**If Yes, do not test if the skin is not intact or feels rough or irritated**)		

Results:

Histamine Control	____mm by _____mm	Alternaria	____mm by _____mm
Cat dander	____mm by _____mm	House dust mite	____mm by _____mm
Local grasses	____mm by _____mm	Cladosporium	____mm by _____mm
Aspergillus	____mm by _____mm	Saline Control	____mm by _____mm

Skin testing was:

_____ Completed
_____ Not completed; If **NOT**, then why:
_____ Subject could not perform test
_____ Refused
_____ Other

Nurses' Comments (enter comments from results sheet into database as well):

Appendix C

Ethical Approval of the Study

UNIVERSITY OF SASKATCHEWAN

Biomedical Research Ethics Board (Bio-REB)

Certificate of Approval

PRINCIPAL INVESTIGATOR	DEPARTMENT	Bio #
James A. Dosman	Canadian Centre for Health and Safety in Agriculture	15-69

INSTITUTION(S) WHERE RESEARCH WILL BE CARRIED OUT
University of Saskatchewan

Saskatoon SK

SUB-INVESTIGATOR(S)
Punam Pahwa, Donna Rennie, Shahedul Khan

STUDENT RESEARCHER(S)
Rifat Zahan

FUNDER(S)
CANADIAN INSTITUTES OF HEALTH RESEARCH (CIHR)

TITLE
Spirometric Reference Equations for Aboriginal Children in Saskatchewan

ORIGINAL REVIEW DATE	APPROVED ON	APPROVAL OF	EXPIRY DATE
17-Mar-2015	25-Mar-2015	Research project as outlined in the revised Application to access Existing Health Data for Research (rec'd 24-Mar-2015)	24-Mar-2016

Delegated Review ☒ Full Board Meeting ☐

CERTIFICATION
The study is acceptable on scientific and ethical grounds. The Bio-REB considered the requirements of section 29 under the Health Information Protection Act (HIPA) and is satisfied that this study meets the privacy considerations outlined therein. The principal investigator has the responsibility for any other administrative or regulatory approvals that may pertain to this research study, and for ensuring that the authorized research is carried out according to governing law. This approval is valid for the specified period provided there is no change to the approved protocol or consent process.

FIRST TIME REVIEW AND CONTINUING APPROVAL
The University of Saskatchewan Biomedical Research Ethics Board reviews above minimal studies at a full-board (face-to-face) meeting. If a protocol has been reviewed at a full board meeting, a subsequent study of the same protocol may be reviewed through the delegated review process. Any research classified as minimal risk is reviewed through the delegated (subcommittee) review process. The initial Certificate of Approval includes the approval period the REB has assigned to a study. The Status Report form must be submitted within one month prior to the assigned expiry date. The researcher shall indicate to the REB any specific requirements of the sponsoring organizations (e.g. requirement for full-board review and approval) for the continuing review process deemed necessary for that project. For more information visit http://research.usask.ca/for-researchers/ethics/index.php.

REB ATTESTATION
In respect to clinical trials, the University of Saskatchewan Research Ethics Board complies with the membership requirements for Research Ethics Boards defined in Part 4 of the Natural Health Products Regulations and Division 5 of the Food and Drug Regulations and carries out its functions in a manner consistent with Good Clinical Practices. Members of the Bio-REB who are named as investigators, do not participate in the discussion related to, nor vote on such studies when presented to the Bio-REB. This approval and the views of this REB have been documented in writing. The University of Saskatchewan Biomedical Research Ethics Board has been

Please send all correspondence to:

Research Ethics Office
University of Saskatchewan
Box 5000 RPO University
1607 110 Gymnasium Place
Saskatoon, SK Canada S7N 4J8

PRINCIPAL INVESTIGATOR	DEPARTMENT	Bio #
James A. Dosman	Canadian Centre for Health and Safety in Agriculture	15-69

approved by the Minister of Health, Province of Saskatchewan, to serve as a Research Ethics Board (REB) for research projects involving human subjects under section 29 of The Health Information Protection Act (HIPA).

Ildiko Badea, Chair
University of Saskatchewan
Biomedical Research Ethics Board

Please send all correspondence to:

Research Ethics Office
University of Saskatchewan
Box 5000 RPO University
1607 - 110 Gymnasium Place
Saskatoon, SK Canada S7N 4J8

Appendix D

Parental Consent/Child Assent to Participate in the study

PARENT-CHILD INFORMATION AND CONSENT/ASSENT FORMS
FIRST NATIONS LUNG HEALTH PROJECT
CLINICAL ASSESSMENT – CHILDREN'S EVALUATION

WHO IS DOING THE PROJECT?
The project is being led by Dr. James Dosman, Dr. Punam Pahwa, Dr. Sylvia Abonyi from University of Saskatchewan, Dr. Jo-Ann Episkenew from the Indigenous Peoples Health Research Centre, University of Regina, in cooperation with your Band Council, Health Centre, Elder from your community, and support from the school principal/board. The project is funded by the Canadian Institutes of Health Research.

WHY IS THIS PROJECT BEING DONE?
To learn more about conditions such as housing and breathing problems in all school-age children. What we learn will help your Band Council to make programs and policies to improve the lung health of children. The project will be conducted in the two communities [redacted].

WHAT IS INVOLVED?
After completing the questionnaire you will be invited to participate in the clinical assessment of the study. Height, weight, waist and blood pressure **measurements** will be taken. For a **breathing test** your child will blow out into a machine called a spirometer. For an **allergy test**, a nurse will put 8 small drops on the arm and the skin underneath each drop will be lightly scratched. All of these tests will take place in one meeting with your child lasting no more than 30 minutes. All these measurements will be collected at your school by a registered nurse (RN) or licensed practical nurse (LPN). Copy of the results from the testing will be mailed to you for your records.

ARE THERE POSSIBLE RISKS AND DISCOMFORTS?
The breathing test may cause mild, temporary discomfort such as dizziness or coughing for a few seconds after the test. The allergy skin prick test might make your arm itchy but that will go away after about an hour. We will keep you with us until the itchiness is gone and we can apply anti-itch cream if you would like. In the event that you become sick from having this test, we will provide you with the necessary medical treatment at no cost. There is no risk to the height, weight, waist and blood pressure measurements. A nurse is there to help if needed. You don't have to answer questions that you choose not to or feel uncomfortable to answer.

WHAT ARE THE BENEFITS OF PARTICIPATING?
There may be no direct benefit to your child from doing these tests. The information from this testing can be used in the future to benefit lung health of other First Nations children. The school will get funds that may be used for special needs of the classes.

WHAT HAPPENS IF MY CHILD DECIDES TO WITHDRAW?
Your child does not have to be in the study if they do not want to be. Your child may stop at any time. Even if your child agrees to participate, you may withdraw information gathered about your child until his/her data pooled with other children data. Withdrawing from this study will not affect your child's school or medical care.

HOW WILL WE BE INFORMED OF THE RESULTS?
A copy of the child's result will be mailed to parents/caregivers. Summary findings will be put in newsletters produced by the [redacted]

HOW WILL STUDY INFORMATION BE SHARED?
No personal or individual information will be shared – it will be strictly protected. Results may be published and presented at conferences but it will not be possible to identify individuals. All information (with your personal information removed) will be stored in a secure location at the University of Saskatchewan in the care of Dr. James Dosman and Dr. Punam Pahwa for at least five years.

WHO DO I CONTACT IF I HAVE QUESTIONS?
If you have any questions or want further information about this project before, during, or after participation, you can contact **Dr. James Dosman at (306) 966-7884** or **Ms. Kathleen McMullin at (306) 960-3238**, or the **Community Contact at (Health Director at (306) 467-4402)**, or you can call the Chair of the **University of Saskatchewan Biomedical Research Ethics Board at (306) 966-4053.** Out of town participants may call collect. This study has been reviewed and approved on ethical grounds by the University of Saskatchewan Research Ethics Board.

PARENTAL CONSENT/CHILD ASSENT TO PARTICIPATE

FIRST NATIONS LUNG HEALTH PROJECT CLINICAL ASSESSMENT – CHILDREN'S EVALUATION

- I understand the purpose and procedures and the possible risks and benefits of the project for my child
- I understand that the time to complete the tests will be about 30 minutes.
- I understand that my child is free to withdraw for any reason before, during, and after the testing.
- I understand that I may withdraw any information about my child who has participated until his/her data pooled with other children data.
- I give permission to use the information collected in a way that does not identify my child.
- I understand that by signing this document I do not waive any of my legal rights.
- I will be given a signed copy of this consent form.
- I agree that my child may participate in any or all of the following tests:

	Please check all or any:	
Blood Pressure Measurement	**Yes** ☐	**No** ☐
Breathing Test with height, weight and waist measurements	**Yes** ☐	**No** ☐
Allergy test on the skin	**Yes** ☐	**No** ☐

Name of child _____

Printed name of Parent or Caregiver:	Signature	Date
_____	_____	_____

I have read this paper or have had it read to me. I understand what I have to do in this study and I agree to take part in it.

Printed name of Child:	Signature	Date
_____	_____	_____

FOR ADMINSTRATIVE USE ONLY

Please check which statement applies (to be completed by the person administering the assent):

☐ The child is capable of reading and understanding the assent form and has signed the above documentation of assent to take part in this study.

☐ The child is not capable of reading the assent form, however, the information was explained verbally to the subject who has verbally given assent to take part in this study.

Printed name of Parent or Caregiver:	Signature	Date
_____	_____	_____

Appendix E

Software Implementation for Data Analysis

The data was stored in the format of .sav. The data was read from statistical software R version 3.2.2 for further analysis. The SAS version 9.4 was used to do the comparison between the means of the study variables described in Table 4.1. The corresponding R and SAS codes are given below:

```
1
2  # Update R packages
3  update.packages(ask = FALSE, dependencies = c('Suggests'))
4  library(foreign)
5  library(zoo)
6  library(nlme)
7  library(MASS)
8  library(mgcv)
9  library(gamlss.data)
10 library(gamlss.dist)
11 library(splines)
12 library(parallel)
13 library(gamlss)
14 library(rgl)
15 library(gamlss.dist)
16 library(scatterplot3d)
17 ##########################################################
18 # Read the original data
19 dat <- read.spss("FNChildren2013_RIFAT.sav", to.data.frame=TRUE)
20 ##########################################################
21 # Data construction
22 dat<-dat[which(dat$FVC!="NA"),]
23 dat<-dat[which(dat$FEV1!="NA"),]
24 dat<-dat[which(dat$FEV1FVC!="NA"),]
25
26 dat<-dat[which(dat$CHILDAGE!="NA"),]
27 dat<-dat[which(dat$c_HEIGHT!="NA"),]
28
29 dat<-dat[which(dat$DIAGASTHMA=='No'),]
30 dat<-dat[which(dat$CURRENTCONGESTION!='Yes'),]
31 dat<-dat[which(dat$BRONCHITIS=='No'),]
32 dat<-dat[which(dat$WHEEZE12MTH!='Yes'),]
33 dat<-dat[which(dat$CURRENTCOUGH=='No'),]
34 dat<-dat[which(dat$PNEUMONIA=='No'),]
35 dat<-dat[which(dat$SLEEPAPNEA=='No'),]
36 dat<-dat[which(dat$CROUP=='No'),]
37 dat<-dat[which(dat$SMOKEYEARS >= 1),]
38
39
```

```
## Data construction for boys
datm1<-dat[which(dat$c_SEX!="Female"),]
agem <- datm1$CHILDAGE
heightm <- datm1$c_HEIGHT
weightm <- datm1$c_WEIGHT
fvcm <- datm1$FVC
fev1m <- datm1$FEV1
fev1fvcm <- datm1$FEV1FVC
bmim <- datm1$c_BMI
abgirthm <- datm1$c_ABGIRTH
datm <- cbind(agem, heightm, fvcm, fev1m, fev1fvcm, weightm, bmim)
datm <- data.frame(datm)

## Data construction for girls
datf1<-dat[which(dat$c_SEX!="Male"),]
agef <- datf1$CHILDAGE
heightf <- datf1$c_HEIGHT
weightf <- datf1$c_WEIGHT
fvcf <- datf1$FVC
fev1f <- datf1$FEV1
fev1fvcf <- datf1$FEV1FVC
bmif <- datf1$c_BMI
datf <- cbind(bmif, agef, heightf, fvcf, fev1f, fev1fvcf, weightf)
datf <- data.frame(datf)

### Descriptive Statistics ###
## Means ##
mean(agem)
mean(agef)
mean(heightm)
mean(heightf)
mean(fvcm)
mean(fvcf)
mean(fev1m)
mean(fev1f)
mean(fev1fvcm)
mean(fev1fvcf)
## Standard Deviations ##
sd(agem)
sd(agef)
sd(heightm)
sd(heightf)
sd(fvcm)
sd(fvcf)
sd(fev1m)
sd(fev1f)
sd(fev1fvcm)
sd(fev1fvcf)
## Skewness ##
skewness(agem)
skewness(agef)
```

```
skewness(heightm)
skewness(heightf)
skewness(fvcm)
skewness(fvcf)
skewness(fev1m)
skewness(fev1f)
skewness(fev1fvcm)
skewness(fev1fvcf)

## Mean test in SAS
data meantest;
  input x1bar s1 n1 x2bar s2 n2;
var1 = (((n1-1)*s1*s1) + ((n2-1)*s2*s2))/(n1 + n2 - 2);
var2 = (1/n1) + (1/n2);
var = var1*var2;
sd = sqrt(var);
t = (x1bar - x2bar)/sd;
datalines;
11.04 2.67 53 9.97 2.98 76
151.74 15.84 53 146 15.98 76
50.01 16.63 53 47.21 19.11 76
21.24 4.28 53 21.33 4.88 76
3.14 1.02 53 2.68 0.93 76
2.72 0.93 53 2.37 0.83 76
86.60 5.26 53 88.50 5.23 76
run;

proc print data = meantest;
run;

############################################################
gamlss.fit<-function(y,height,age,p1,p2,p3){
if(p3>0){
fit0<-gamlss(y ~ log(height) + log(age) + poly(age/20,p1,raw=TRUE),
                    sigma.fo = ~ 1+log(age) + poly(age/20,p2,raw=TRUE),
                    nu.fo = ~ 1+log(age) + poly(age/20, p3,raw=TRUE),family =
                        BCCGo)
}
if(p3==0){
fit0<-gamlss(y ~ log(height) + log(age) + poly(age/20,p1,raw=TRUE),
                    sigma.fo = ~ 1+log(age) + poly(age/20,p2,raw=TRUE),
                    nu.fo = ~ 1+log(age),family = BCCGo)
}
if(p3<0){
fit0<-gamlss(y ~ log(height) + log(age) + poly(age/20,p1,raw=TRUE),
                    sigma.fo = ~ 1+log(age) + poly(age/20,p2,raw=TRUE),
                    nu.fo = ~ 1,family = BCCGo)
}
repeat{
if(fit0$converged=="TRUE") break
fit0<-refit(fit0)
```

```
}
return(fit0)
}

#################################################################
#################################################################
# BOYS: FVC
#################################################################
fit1.fvcm<-gamlss.fit(fvcm,heightm,agem,2,2,2)
fit2.fvcm<-gamlss.fit(fvcm,heightm,agem,3,3,2)
fit3.fvcm<-gamlss.fit(fvcm,heightm,agem,4,3,1)
fit4.fvcm<-gamlss.fit(fvcm,heightm,agem,4,3,2)
fit5.fvcm<-gamlss.fit(fvcm,heightm,agem,5,2,2)
AIC(fit1.fvcm,fit2.fvcm,fit3.fvcm,fit4.fvcm,fit5.fvcm)

            df       AIC
fit3.fvcm 15 22.01192
fit4.fvcm 16 23.83976
fit2.fvcm 15 27.94295
fit5.fvcm 16 32.28319
fit1.fvcm 13 34.10697

fit.fvcm<-fit3.fvcm
summary(fit.fvcm)
#########################################################
age<-seq(min(agem), max(agem), length = 200)
height <- seq(min(heightm), max(heightm), length = 200)

newdata01<-cbind(1,log(height),log(age),poly(age/20,4,raw=TRUE))
newdata02<-cbind(1,log(age),poly(age/20,3,raw=TRUE))
newdata03<-cbind(1,log(age),poly(age/20,1,raw=TRUE))

pred01.fvcm<-pred(fit.fvcm$mu.coefficients,newdata01)
pred02.fvcm<-pred(fit.fvcm$sigma.coefficients,newdata02)
pred03.fvcm<-pred1(fit.fvcm$nu.coefficients,newdata03)
med.fvcm0<-qBCCG(0.5,pred01.fvcm,pred02.fvcm,pred03.fvcm)


dd<-scatterplot3d(agem,heightm,fvcm,color="dodgerblue4",pch=20,xlab="Age (in
    years)",ylab="Height (in cm)",zlab="FVC (in liters)")
dd$points3d(age,height,med.fvcm0,type="l",col="red")
#################################################################
age1 <- rep(6,200)
age2 <- rep(9,200)
age3 <- rep(12,200)
age4 <- rep(15,200)
age5 <- rep(17,200)
age6 <- rep(17,200)

height <- seq(min(heightm), max(heightm), length = 200)

```

```
191 newdata1<-cbind(1,log(height),log(age1),poly(age1/20,4,raw=TRUE))
192 newdata2<-cbind(1,log(height),log(age2),poly(age2/20,4,raw=TRUE))
193 newdata3<-cbind(1,log(height),log(age3),poly(age3/20,4,raw=TRUE))
194 newdata4<-cbind(1,log(height),log(age4),poly(age4/20,4,raw=TRUE))
195 newdata5<-cbind(1,log(height),log(age5),poly(age5/20,4,raw=TRUE))
196 newdata6<-cbind(1,log(height),log(age6),poly(age6/20,4,raw=TRUE))
197
198 newdata11<-cbind(1,log(age1),poly(age1/20,3,raw=TRUE))
199 newdata12<-cbind(1,log(age2),poly(age2/20,3,raw=TRUE))
200 newdata13<-cbind(1,log(age3),poly(age3/20,3,raw=TRUE))
201 newdata14<-cbind(1,log(age4),poly(age4/20,3,raw=TRUE))
202 newdata15<-cbind(1,log(age5),poly(age5/20,3,raw=TRUE))
203 newdata16<-cbind(1,log(age6),poly(age6/20,3,raw=TRUE))
204
205 newdata21<-cbind(1,log(age1),poly(age1/20,1,raw=TRUE))
206 newdata22<-cbind(1,log(age2),poly(age2/20,1,raw=TRUE))
207 newdata23<-cbind(1,log(age3),poly(age3/20,1,raw=TRUE))
208 newdata24<-cbind(1,log(age4),poly(age4/20,1,raw=TRUE))
209 newdata25<-cbind(1,log(age5),poly(age5/20,1,raw=TRUE))
210 newdata26<-cbind(1,log(age6),poly(age6/20,1,raw=TRUE))
211
212 pred1.fvcm<-pred(fit.fvcm$mu.coefficients,newdata1)
213
214
215 pred2.fvcm<-pred(fit.fvcm$mu.coefficients,newdata2)
216 pred3.fvcm<-pred(fit.fvcm$mu.coefficients,newdata3)
217 pred4.fvcm<-pred(fit.fvcm$mu.coefficients,newdata4)
218 pred5.fvcm<-pred(fit.fvcm$mu.coefficients,newdata5)
219 pred6.fvcm<-pred(fit.fvcm$mu.coefficients,newdata6)
220
221 pred11.fvcm<-pred(fit.fvcm$sigma.coefficients,newdata11)
222 pred12.fvcm<-pred(fit.fvcm$sigma.coefficients,newdata12)
223 pred13.fvcm<-pred(fit.fvcm$sigma.coefficients,newdata13)
224 pred14.fvcm<-pred(fit.fvcm$sigma.coefficients,newdata14)
225 pred15.fvcm<-pred(fit.fvcm$sigma.coefficients,newdata15)
226 pred16.fvcm<-pred(fit.fvcm$sigma.coefficients,newdata16)
227
228 pred21.fvcm<-pred1(fit.fvcm$nu.coefficients,newdata21)
229 pred22.fvcm<-pred1(fit.fvcm$nu.coefficients,newdata22)
230 pred23.fvcm<-pred1(fit.fvcm$nu.coefficients,newdata23)
231 pred24.fvcm<-pred1(fit.fvcm$nu.coefficients,newdata24)
232 pred25.fvcm<-pred1(fit.fvcm$nu.coefficients,newdata25)
233 pred26.fvcm<-pred1(fit.fvcm$nu.coefficients,newdata26)
234
235 med.fvcm1<-qBCCG(0.5,pred1.fvcm,pred11.fvcm,pred21.fvcm)
236 med.fvcm2<-qBCCG(0.5,pred2.fvcm,pred12.fvcm,pred22.fvcm)
237 med.fvcm3<-qBCCG(0.5,pred3.fvcm,pred13.fvcm,pred23.fvcm)
238 med.fvcm4<-qBCCG(0.5,pred4.fvcm,pred14.fvcm,pred24.fvcm)
239 med.fvcm5<-qBCCG(0.5,pred5.fvcm,pred15.fvcm,pred25.fvcm)
240 med.fvcm6<-qBCCG(0.5,pred6.fvcm,pred16.fvcm,pred26.fvcm)
241
```

```
242
243 plot(height,med.fvcm1,col=1,type="l",xlab="Height (in cm)",
244  ylab="Height-age adjusted median reference FVC (in liters)",lty=1,lwd=1,
245   ylim=c(min(c(med.fvcm1,med.fvcm2,med.fvcm3,med.fvcm4,med.fvcm5,med.fvcm6)),
246       max(c(med.fvcm1,med.fvcm2,med.fvcm3,med.fvcm4,med.fvcm5,med.fvcm6))))
247 lines(height,med.fvcm2,col=2,lty=2,lwd=1)
248 lines(height,med.fvcm3,col=3,lty=3,lwd=1)
249 lines(height,med.fvcm4,col=4,lty=4,lwd=1)
250 lines(height,med.fvcm5,col=5,lty=5,lwd=1)
251 lines(height,med.fvcm6,col=6,lty=6,lwd=1)
252
253 legend("topleft",c("age 7","age 9","age 11","age 13","age 15","age 17"),
254  col=1:6,lty=1:6)
255 ##########################################################################
256 # Contour Plot
257 ##########################################################################
258 age<-seq(min(agem),max(agem),length=500)
259 height<-seq(min(heightm),max(heightm),length=500)
260
261 newdata<-expand.grid(height=height,age=age)
262
263 newdata1<-cbind(1,newdata,poly(newdata$age/20,4,raw=TRUE))
264 newdata2<-cbind(1,newdata$age,poly(newdata$age/20,3,raw=TRUE))
265 newdata3<-cbind(1,newdata$age,poly(newdata$age/20,1,raw=TRUE))
266
267 newdata1[,2]<-log(newdata1[,2])
268 newdata1[,3]<-log(newdata1[,3])
269
270 newdata2[,2]<-log(newdata2[,2])
271
272 newdata3[,2]<-log(newdata3[,2])
273
274
275 pp<-qBCCG(0.5,pred(fit.fvcm$mu.coefficients,newdata1),
276  pred(fit.fvcm$sigma.coefficients,newdata2),pred1(fit.fvcm$nu.coefficients,
       newdata3))
277
278 pp.mat<-matrix(pp,ncol=length(age))
279
280 contour(height,age,pp.mat,nlevels=20,xlab="Height (in cm)", ylab="Age (in
      years)")
281
282 ##########################################################################
283 age1 <- seq(6,17,length=200)
284 height <- seq(min(heightm), max(heightm), length = 200)
285
286 newdata1<-cbind(1,log(height),log(age1),poly(age1/20,4,raw=TRUE))
287 newdata2<-cbind(1,log(age1),poly(age1/20,3,raw=TRUE))
288 newdata3<-cbind(1,log(age1),poly(age1/20,1,raw=TRUE))
289
290
```

```
Mpred1.fvcm<-pred(fit.fvcm$mu.coefficients,newdata1)
Spred1.fvcm<-pred(fit.fvcm$sigma.coefficients,newdata2)
Lpred1.fvcm<-pred1(fit.fvcm$nu.coefficients,newdata3)


med.fvcm<-qBCCG(0.5,Mpred1.fvcm,Spred1.fvcm,Lpred1.fvcm)
LLN.fvcm<-qBCCG(0.05,Mpred1.fvcm,Spred1.fvcm,Lpred1.fvcm)

plot(agem,fvcm,xlab="Age (in years)",ylab="FVC (in liters)")

lines(age1,med.fvcm,col=2,lty=1)
lines(age1,LLN.fvcm,lty=2,col=3)

legend("topleft",c("Height-age adjusted median FVC","Height-age adjusted LLN
    for FVC"),col=2:3,lty=1:2)

##############################################################################
##############################################################################
# GIRLS: FVC
#####################################################################
fit1.fvcf<-gamlss.fit(fvcf,heightf,agef,2,2,1)
fit2.fvcf<-gamlss.fit(fvcf,heightf,agef,2,2,2)
fit3.fvcf<-gamlss.fit(fvcf,heightf,agef,3,3,1)
fit4.fvcf<-gamlss.fit(fvcf,heightf,agef,3,3,2)
fit5.fvcf<-gamlss.fit(fvcf,heightf,agef,5,3,1)
fit6.fvcf<-gamlss.fit(fvcf,heightf,agef,5,3,2)
fit7.fvcf<-gamlss.fit(fvcf,heightf,agef,5,4,1)
fit8.fvcf<-gamlss.fit(fvcf,heightf,agef,5,4,2)

           df       AIC
fit3.fvcf  14  42.69662
fit4.fvcf  15  44.44175
fit5.fvcf  16  46.60327
fit7.fvcf  17  47.80257
fit6.fvcf  17  48.32411
fit8.fvcf  18  49.60252
fit1.fvcf  12  50.23985
fit2.fvcf  13  52.20455

AIC(fit1.fvcf,fit2.fvcf,fit3.fvcf,fit4.fvcf,fit5.fvcf,fit6.fvcf,fit7.fvcf,fit8
    .fvcf)

fit.fvcf<-fit3.fvcf
summary(fit.fvcf)
#########################################################
age<-seq(min(agef), max(agef), length = 200)
height <- seq(min(heightf), max(heightf), length = 200)

newdata01<-cbind(1,log(height),log(age),poly(age/20,3,raw=TRUE))
newdata02<-cbind(1,log(age),poly(age/20,3,raw=TRUE))
newdata03<-cbind(1,log(age),poly(age/20,1,raw=TRUE))
```

```
340
341 pred01.fvcf<-pred(fit.fvcf$mu.coefficients,newdata01)
342 pred02.fvcf<-pred(fit.fvcf$sigma.coefficients,newdata02)
343 pred03.fvcf<-pred1(fit.fvcf$nu.coefficients,newdata03)
344 med.fvcf0<-qBCCG(0.5,pred01.fvcf,pred02.fvcf,pred03.fvcf)
345
346
347 dd<-scatterplot3d(agef,heightf,fvcf,color="dodgerblue4",pch=20,xlab="Age (in
      years)",ylab="Height (in cm)",zlab="FVC (in liters)")
348 dd$points3d(age,height,med.fvcf0,type="l")
349 ##########################################################################
350
351 age1 <- rep(7,200)
352 age2 <- rep(9,200)
353 age3 <- rep(11,200)
354 age4 <- rep(13,200)
355 age5 <- rep(15,200)
356 age6 <- rep(17,200)
357
358 height <- seq(min(heightf), max(heightf), length = 200)
359
360 newdata1<-cbind(1,log(height),log(age1),poly(age1/20,3,raw=TRUE))
361 newdata2<-cbind(1,log(height),log(age2),poly(age2/20,3,raw=TRUE))
362 newdata3<-cbind(1,log(height),log(age3),poly(age3/20,3,raw=TRUE))
363 newdata4<-cbind(1,log(height),log(age4),poly(age4/20,3,raw=TRUE))
364 newdata5<-cbind(1,log(height),log(age5),poly(age5/20,3,raw=TRUE))
365 newdata6<-cbind(1,log(height),log(age6),poly(age6/20,3,raw=TRUE))
366
367 newdata11<-cbind(1,log(age1),poly(age1/20,3,raw=TRUE))
368 newdata12<-cbind(1,log(age2),poly(age2/20,3,raw=TRUE))
369 newdata13<-cbind(1,log(age3),poly(age3/20,3,raw=TRUE))
370 newdata14<-cbind(1,log(age4),poly(age4/20,3,raw=TRUE))
371 newdata15<-cbind(1,log(age5),poly(age5/20,3,raw=TRUE))
372 newdata16<-cbind(1,log(age6),poly(age6/20,3,raw=TRUE))
373
374 newdata21<-cbind(1,log(age1),poly(age1/20,1,raw=TRUE))
375 newdata22<-cbind(1,log(age2),poly(age2/20,1,raw=TRUE))
376 newdata23<-cbind(1,log(age3),poly(age3/20,1,raw=TRUE))
377 newdata24<-cbind(1,log(age4),poly(age4/20,1,raw=TRUE))
378 newdata25<-cbind(1,log(age5),poly(age5/20,1,raw=TRUE))
379 newdata26<-cbind(1,log(age6),poly(age6/20,1,raw=TRUE))
380
381 pred1.fvcf<-pred(fit.fvcf$mu.coefficients,newdata1)
382 pred2.fvcf<-pred(fit.fvcf$mu.coefficients,newdata2)
383 pred3.fvcf<-pred(fit.fvcf$mu.coefficients,newdata3)
384 pred4.fvcf<-pred(fit.fvcf$mu.coefficients,newdata4)
385 pred5.fvcf<-pred(fit.fvcf$mu.coefficients,newdata5)
386 pred6.fvcf<-pred(fit.fvcf$mu.coefficients,newdata6)
387
388 pred11.fvcf<-pred(fit.fvcf$sigma.coefficients,newdata11)
389 pred12.fvcf<-pred(fit.fvcf$sigma.coefficients,newdata12)
```

```
pred13.fvcf<-pred(fit.fvcf$sigma.coefficients,newdata13)
pred14.fvcf<-pred(fit.fvcf$sigma.coefficients,newdata14)
pred15.fvcf<-pred(fit.fvcf$sigma.coefficients,newdata15)
pred16.fvcf<-pred(fit.fvcf$sigma.coefficients,newdata16)

pred21.fvcf<-pred1(fit.fvcf$nu.coefficients,newdata21)
pred22.fvcf<-pred1(fit.fvcf$nu.coefficients,newdata22)
pred23.fvcf<-pred1(fit.fvcf$nu.coefficients,newdata23)
pred24.fvcf<-pred1(fit.fvcf$nu.coefficients,newdata24)
pred25.fvcf<-pred1(fit.fvcf$nu.coefficients,newdata25)
pred26.fvcf<-pred1(fit.fvcf$nu.coefficients,newdata26)

med.fvcf1<-qBCCG(0.5,pred1.fvcf,pred11.fvcf,pred21.fvcf)
med.fvcf2<-qBCCG(0.5,pred2.fvcf,pred12.fvcf,pred22.fvcf)
med.fvcf3<-qBCCG(0.5,pred3.fvcf,pred13.fvcf,pred23.fvcf)
med.fvcf4<-qBCCG(0.5,pred4.fvcf,pred14.fvcf,pred24.fvcf)
med.fvcf5<-qBCCG(0.5,pred5.fvcf,pred15.fvcf,pred25.fvcf)
med.fvcf6<-qBCCG(0.5,pred6.fvcf,pred16.fvcf,pred26.fvcf)


plot(height,med.fvcf1,col=1,type="l",xlab="Height (in cm)",
  ylab="Height-age adjusted median reference FVC (in liters)",
    lty=1,lwd=1,ylim=c(min(c(med.fvcf1,med.fvcf2,med.fvcf3,med.fvcf4,med.fvcf5,
        med.fvcf6)),
      max(c(med.fvcf1,med.fvcf2,med.fvcf3,med.fvcf4,med.fvcf5,med.fvcf6))))
lines(height,med.fvcf2,col=2,lty=2,lwd=1)
lines(height,med.fvcf3,col=3,lty=3,lwd=1)
lines(height,med.fvcf4,col=4,lty=4,lwd=1)
lines(height,med.fvcf5,col=5,lty=5,lwd=1)
lines(height,med.fvcf6,col=6,lty=6,lwd=1)

legend("topleft",c("age 7","age 9","age 11","age 13","age 15","age 17"),
 col=1:6,lty=1:6)

#########################################################################
#########################################################################
# Contour Plot
#########################################################################
age<-seq(min(agef),max(agef),length=500)
height<-seq(min(heightf),max(heightf),length=500)

newdata<-expand.grid(height=height,age=age)

newdata1<-cbind(1,newdata,poly(newdata$age/20,3,raw=TRUE))
newdata2<-cbind(1,newdata$age,poly(newdata$age/20,3,raw=TRUE))
newdata3<-cbind(1,newdata$age,poly(newdata$age/20,1,raw=TRUE))

newdata1[,2]<-log(newdata1[,2])
newdata1[,3]<-log(newdata1[,3])

newdata2[,2]<-log(newdata2[,2])
```

```
440
441 newdata3[,2]<-log(newdata3[,2])
442
443
444 pp<-qBCCG(0.5,pred(fit.fvcf$mu.coefficients,newdata1),
445  pred(fit.fvcf$sigma.coefficients,newdata2),pred1(fit.fvcf$nu.coefficients,
       newdata3))
446
447 pp.mat<-matrix(pp,ncol=length(age))
448
449 contour(height,age,pp.mat,xlab="Height (in cm)",ylab="Age (in years)",nlevels
    =20)
450 ##########################################################################
451 age1 <- seq(6,17,length=200)
452 height <- seq(min(heightf), max(heightf), length = 200)
453
454 newdata1<-cbind(1,log(height),log(age1),poly(age1/20,3,raw=TRUE))
455 newdata2<-cbind(1,log(age1),poly(age1/20,3,raw=TRUE))
456 newdata3<-cbind(1,log(age1),poly(age1/20,1,raw=TRUE))
457
458
459 Mpred1.fvcf<-pred(fit.fvcf$mu.coefficients,newdata1)
460 Spred1.fvcf<-pred(fit.fvcf$sigma.coefficients,newdata2)
461 Lpred1.fvcf<-pred1(fit.fvcf$nu.coefficients,newdata3)
462
463
464 med.fvcf<-qBCCG(0.5,Mpred1.fvcf,Spred1.fvcf,Lpred1.fvcf)
465 LLN.fvcf<-qBCCG(0.05,Mpred1.fvcf,Spred1.fvcf,Lpred1.fvcf)
466
467
468 plot(agef,fvcf,xlab="Age (in years)",ylab="FVC (in liters)")
469 lines(age1,med.fvcf,col=2,lty=1)
470 lines(age1,LLN.fvcf,col=3,lty=2)
471
472 legend("topleft",c("Height-age adjusted median FVC","Height-age adjusted LLN
    for FVC"),col=2:3,lty=1:2)
473
474 ##########################################################
475 ##########################################################
476 # BOYS: FEV1
477 ##########################################################
478 ##########################################################
479 fit1.fevm<-gamlss.fit(fev1m,heightm,agem,2,2,1)
480 fit2.fevm<-gamlss.fit(fev1m,heightm,agem,2,2,2)
481 fit3.fevm<-gamlss.fit(fev1m,heightm,agem,3,3,1)
482 fit4.fevm<-gamlss.fit(fev1m,heightm,agem,3,3,2)
483 fit5.fevm<-gamlss.fit(fev1m,heightm,agem,4,3,1)
484 fit6.fevm<-gamlss.fit(fev1m,heightm,agem,5,2,2)
485
486 AIC(fit1.fevm,fit2.fevm,fit3.fevm,fit4.fevm,fit5.fevm,fit6.fevm)
487
```

```
           df       AIC
fit5.fevm  15  14.91136
fit4.fevm  15  18.80708
fit1.fevm  12  20.84756
fit3.fevm  14  23.55715
fit6.fevm  16  24.71780
fit2.fevm  13  26.00631

fit.fevm<-fit5.fevm
summary(fit.fvcm)
############################################################
age<-seq(min(agem), max(agem), length = 200)
height <- seq(min(heightm), max(heightm), length = 200)

newdata01<-cbind(1,log(height),log(age),poly(age/20,4,raw=TRUE))
newdata02<-cbind(1,log(age),poly(age/20,3,raw=TRUE))
newdata03<-cbind(1,log(age),poly(age/20,1,raw=TRUE))

pred01.fevm<-pred(fit.fevm$mu.coefficients,newdata01)
pred02.fevm<-pred(fit.fevm$sigma.coefficients,newdata02)
pred03.fevm<-pred1(fit.fevm$nu.coefficients,newdata03)
med.fevm0<-qBCCG(0.5,pred01.fevm,pred02.fevm,pred03.fevm)


dd<-scatterplot3d(agem,heightm,fev1m,color="dodgerblue4",pch=20,xlab="Age (in
    years)",ylab="Height (in cm)",
   zlab=expression('FEV '[1]* '  (in liters)'))
dd$points3d(age,height,med.fevm0,type="l")

############################################################
age1 <- rep(7,200)
age2 <- rep(9,200)
age3 <- rep(11,200)
age4 <- rep(13,200)
age5 <- rep(15,200)
age6 <- rep(17,200)

height <- seq(min(heightm), max(heightm), length = 200)

newdata1<-cbind(1,log(height),log(age1),poly(age1/20,4,raw=TRUE))
newdata2<-cbind(1,log(height),log(age2),poly(age2/20,4,raw=TRUE))
newdata3<-cbind(1,log(height),log(age3),poly(age3/20,4,raw=TRUE))
newdata4<-cbind(1,log(height),log(age4),poly(age4/20,4,raw=TRUE))
newdata5<-cbind(1,log(height),log(age5),poly(age5/20,4,raw=TRUE))
newdata6<-cbind(1,log(height),log(age6),poly(age6/20,4,raw=TRUE))

newdata11<-cbind(1,log(age1),poly(age1/20,3,raw=TRUE))
newdata12<-cbind(1,log(age2),poly(age2/20,3,raw=TRUE))
newdata13<-cbind(1,log(age3),poly(age3/20,3,raw=TRUE))
newdata14<-cbind(1,log(age4),poly(age4/20,3,raw=TRUE))
newdata15<-cbind(1,log(age5),poly(age5/20,3,raw=TRUE))
```

```
538 newdata16<-cbind(1,log(age6),poly(age6/20,3,raw=TRUE))
539
540 newdata21<-cbind(1,log(age1),poly(age1/20,1,raw=TRUE))
541 newdata22<-cbind(1,log(age2),poly(age2/20,1,raw=TRUE))
542 newdata23<-cbind(1,log(age3),poly(age3/20,1,raw=TRUE))
543 newdata24<-cbind(1,log(age4),poly(age4/20,1,raw=TRUE))
544 newdata25<-cbind(1,log(age5),poly(age5/20,1,raw=TRUE))
545 newdata26<-cbind(1,log(age6),poly(age6/20,1,raw=TRUE))
546
547 pred1.fevm<-pred(fit.fevm$mu.coefficients,newdata1)
548 pred2.fevm<-pred(fit.fevm$mu.coefficients,newdata2)
549 pred3.fevm<-pred(fit.fevm$mu.coefficients,newdata3)
550 pred4.fevm<-pred(fit.fevm$mu.coefficients,newdata4)
551 pred5.fevm<-pred(fit.fevm$mu.coefficients,newdata5)
552 pred6.fevm<-pred(fit.fevm$mu.coefficients,newdata6)
553
554 pred11.fevm<-pred(fit.fevm$sigma.coefficients,newdata11)
555 pred12.fevm<-pred(fit.fevm$sigma.coefficients,newdata12)
556 pred13.fevm<-pred(fit.fevm$sigma.coefficients,newdata13)
557 pred14.fevm<-pred(fit.fevm$sigma.coefficients,newdata14)
558 pred15.fevm<-pred(fit.fevm$sigma.coefficients,newdata15)
559 pred16.fevm<-pred(fit.fevm$sigma.coefficients,newdata16)
560
561 pred21.fevm<-pred1(fit.fevm$nu.coefficients,newdata21)
562 pred22.fevm<-pred1(fit.fevm$nu.coefficients,newdata22)
563 pred23.fevm<-pred1(fit.fevm$nu.coefficients,newdata23)
564 pred24.fevm<-pred1(fit.fevm$nu.coefficients,newdata24)
565 pred25.fevm<-pred1(fit.fevm$nu.coefficients,newdata25)
566 pred26.fevm<-pred1(fit.fevm$nu.coefficients,newdata26)
567
568 med.fevm1<-qBCCG(0.5,pred1.fevm,pred11.fevm,pred21.fevm)
569 med.fevm2<-qBCCG(0.5,pred2.fevm,pred12.fevm,pred22.fevm)
570 med.fevm3<-qBCCG(0.5,pred3.fevm,pred13.fevm,pred23.fevm)
571 med.fevm4<-qBCCG(0.5,pred4.fevm,pred14.fevm,pred24.fevm)
572 med.fevm5<-qBCCG(0.5,pred5.fevm,pred15.fevm,pred25.fevm)
573 med.fevm6<-qBCCG(0.5,pred6.fevm,pred16.fevm,pred26.fevm)
574
575
576 plot(height,med.fevm1,col=1,type="l",xlab="Height (in cm)",
577   ylab=expression('Height-age adjusted median reference FEV '[1]*' (in
        liters)'),lty=1,lwd=1,
578   ylim=c(1,5))
579 lines(height,med.fevm2,col=2,lty=2,lwd=1)
580 lines(height,med.fevm3,col=3,lty=3,lwd=1)
581 lines(height,med.fevm4,col=4,lty=4,lwd=1)
582 lines(height,med.fevm5,col=5,lty=5,lwd=1)
583 lines(height,med.fevm6,col=6,lty=6,lwd=1)
584
585 legend("topleft",c("age 7","age 9","age 11","age 13","age 15","age 17"),
586  col=1:6,lty=1:6)
587
```

```
588 ##########################################################################
589 ##########################################################################
590 # Contour Plot
591 ##########################################################################
592 age<-seq(min(agem),max(agem),length=500)
593 height<-seq(min(heightm),max(heightm),length=500)
594
595 newdata<-expand.grid(height=height,age=age)
596
597 newdata1<-cbind(1,newdata,poly(newdata$age/20,4,raw=TRUE))
598 newdata2<-cbind(1,newdata$age,poly(newdata$age/20,3,raw=TRUE))
599 newdata3<-cbind(1,newdata$age,poly(newdata$age/20,1,raw=TRUE))
600
601 newdata1[,2]<-log(newdata1[,2])
602 newdata1[,3]<-log(newdata1[,3])
603
604 newdata2[,2]<-log(newdata2[,2])
605
606 newdata3[,2]<-log(newdata3[,2])
607
608
609 pp<-qBCCG(0.5,pred(fit.fevm$mu.coefficients,newdata1),
610  pred(fit.fevm$sigma.coefficients,newdata2),pred1(fit.fevm$nu.coefficients,
       newdata3))
611
612 pp.mat<-matrix(pp,ncol=length(age))
613
614 contour(height,age,pp.mat,xlab="Height (in cm)",ylab="Age (in years)",nlevels
     =20)
615 ##########################################################################
616 age1 <- seq(6,17,length=200)
617 height <- seq(min(heightm), max(heightm), length = 200)
618
619 newdata1<-cbind(1,log(height),log(age1),poly(age1/20,4,raw=TRUE))
620 newdata2<-cbind(1,log(age1),poly(age1/20,3,raw=TRUE))
621 newdata3<-cbind(1,log(age1),poly(age1/20,1,raw=TRUE))
622
623
624 Mpred1.fevm<-pred(fit.fevm$mu.coefficients,newdata1)
625 Spred1.fevm<-pred(fit.fevm$sigma.coefficients,newdata2)
626 Lpred1.fevm<-pred1(fit.fevm$nu.coefficients,newdata3)
627
628 med.fevm<-qBCCG(0.5,Mpred1.fevm,Spred1.fevm,Lpred1.fevm)
629 LLN.fevm<-qBCCG(0.05,Mpred1.fevm,Spred1.fevm,Lpred1.fevm)
630
631
632
633 plot(agem,fev1m,ylim=c(1,5),xlab="Age (in years)",
634   ylab=expression('FEV '[1]*' (in liters)'))
635 lines(age1,med.fevm,col=2,lty=1)
636 lines(age1,LLN.fevm,col=3,lty=2)
```

```
637
638 legend("topleft",c(expression('Height-age adjusted median FEV '[1]),
639   expression('Height-age adjusted LLN for FEV '[1])),col=2:3,lty=1:2)
640
641 ##############################################################################
642 ##############################################################################
643 ##############################################################################
644 # GIRLS: FEV1
645 ###################################################################
646 fit1.fevf<-gamlss.fit(fev1f,heightf,agef,2,2,1)
647 fit2.fevf<-gamlss.fit(fev1f,heightf,agef,2,2,2)
648 fit3.fevf<-gamlss.fit(fev1f,heightf,agef,3,3,1)
649 fit4.fevf<-gamlss.fit(fev1f,heightf,agef,3,3,2)
650 fit5.fevf<-gamlss.fit(fev1f,heightf,agef,5,3,1)
651 fit6.fevf<-gamlss.fit(fev1f,heightf,agef,5,3,2)
652 fit7.fevf<-gamlss.fit(fev1f,heightf,agef,5,4,1)
653 fit8.fevf<-gamlss.fit(fev1f,heightf,agef,5,4,2)
654
655
656 AIC(fit1.fevf,fit2.fevf,fit3.fevf,fit4.fevf,fit5.fevf,fit6.fevf,fit7.fevf,fit8
      .fevf)
657
658           df      AIC
659 fit2.fevf 13 24.38679
660 fit3.fevf 14 24.52449
661 fit4.fevf 15 26.52100
662 fit7.fevf 17 28.12082
663 fit5.fevf 16 28.18959
664 fit1.fevf 12 29.10914
665 fit8.fevf 18 30.12164
666 fit6.fevf 17 30.18967
667
668 fit.fevf<-fit3.fevf
669 summary(fit.fevf)
670 ##########################################################
671 age<-seq(min(agef), max(agef), length = 200)
672 height <- seq(min(heightf), max(heightf), length = 200)
673
674 newdata01<-cbind(1,log(height),log(age),poly(age/20,3,raw=TRUE))
675 newdata02<-cbind(1,log(age),poly(age/20,3,raw=TRUE))
676 newdata03<-cbind(1,log(age),poly(age/20,1,raw=TRUE))
677
678 pred01.fevf<-pred(fit.fevf$mu.coefficients,newdata01)
679 pred02.fevf<-pred(fit.fevf$sigma.coefficients,newdata02)
680 pred03.fevf<-pred1(fit.fevf$nu.coefficients,newdata03)
681 med.fevf0<-qBCCG(0.5,pred01.fevf,pred02.fevf,pred03.fevf)
682
683
684 dd<-scatterplot3d(agef,heightf,fev1f,color="dodgerblue4",pch=20,xlab="Age (in
      years)",ylab="Height (in cm)",
685    zlab=expression('FEV '[1]* '  (in liters)'))
```

```
686 dd$points3d(age,height,med.fevf0,type="l")
687 ##########################################################################
688 age1 <- rep(7,200)
689 age2 <- rep(9,200)
690 age3 <- rep(11,200)
691 age4 <- rep(13,200)
692 age5 <- rep(15,200)
693 age6 <- rep(17,200)
694
695 height <- seq(min(heightf), max(heightf), length = 200)
696
697 newdata1<-cbind(1,log(height),log(age1),poly(age1/20,3,raw=TRUE))
698 newdata2<-cbind(1,log(height),log(age2),poly(age2/20,3,raw=TRUE))
699 newdata3<-cbind(1,log(height),log(age3),poly(age3/20,3,raw=TRUE))
700 newdata4<-cbind(1,log(height),log(age4),poly(age4/20,3,raw=TRUE))
701 newdata5<-cbind(1,log(height),log(age5),poly(age5/20,3,raw=TRUE))
702 newdata6<-cbind(1,log(height),log(age6),poly(age6/20,3,raw=TRUE))
703
704 newdata11<-cbind(1,log(age1),poly(age1/20,3,raw=TRUE))
705 newdata12<-cbind(1,log(age2),poly(age2/20,3,raw=TRUE))
706 newdata13<-cbind(1,log(age3),poly(age3/20,3,raw=TRUE))
707 newdata14<-cbind(1,log(age4),poly(age4/20,3,raw=TRUE))
708 newdata15<-cbind(1,log(age5),poly(age5/20,3,raw=TRUE))
709 newdata16<-cbind(1,log(age6),poly(age6/20,3,raw=TRUE))
710
711 newdata21<-cbind(1,log(age1),poly(age1/20,1,raw=TRUE))
712 newdata22<-cbind(1,log(age2),poly(age2/20,1,raw=TRUE))
713 newdata23<-cbind(1,log(age3),poly(age3/20,1,raw=TRUE))
714 newdata24<-cbind(1,log(age4),poly(age4/20,1,raw=TRUE))
715 newdata25<-cbind(1,log(age5),poly(age5/20,1,raw=TRUE))
716 newdata26<-cbind(1,log(age6),poly(age6/20,1,raw=TRUE))
717
718 pred1.fevf<-pred(fit.fevf$mu.coefficients,newdata1)
719 pred2.fevf<-pred(fit.fevf$mu.coefficients,newdata2)
720 pred3.fevf<-pred(fit.fevf$mu.coefficients,newdata3)
721 pred4.fevf<-pred(fit.fevf$mu.coefficients,newdata4)
722 pred5.fevf<-pred(fit.fevf$mu.coefficients,newdata5)
723 pred6.fevf<-pred(fit.fevf$mu.coefficients,newdata6)
724
725 pred11.fevf<-pred(fit.fevf$sigma.coefficients,newdata11)
726 pred12.fevf<-pred(fit.fevf$sigma.coefficients,newdata12)
727 pred13.fevf<-pred(fit.fevf$sigma.coefficients,newdata13)
728 pred14.fevf<-pred(fit.fevf$sigma.coefficients,newdata14)
729 pred15.fevf<-pred(fit.fevf$sigma.coefficients,newdata15)
730 pred16.fevf<-pred(fit.fevf$sigma.coefficients,newdata16)
731
732 pred21.fevf<-pred1(fit.fevf$nu.coefficients,newdata21)
733 pred22.fevf<-pred1(fit.fevf$nu.coefficients,newdata22)
734 pred23.fevf<-pred1(fit.fevf$nu.coefficients,newdata23)
735 pred24.fevf<-pred1(fit.fevf$nu.coefficients,newdata24)
736 pred25.fevf<-pred1(fit.fevf$nu.coefficients,newdata25)
```

```
pred26.fevf<-pred1(fit.fevf$nu.coefficients,newdata26)

med.fevf1<-qBCCG(0.5,pred1.fevf,pred11.fevf,pred21.fevf)
med.fevf2<-qBCCG(0.5,pred2.fevf,pred12.fevf,pred22.fevf)
med.fevf3<-qBCCG(0.5,pred3.fevf,pred13.fevf,pred23.fevf)
med.fevf4<-qBCCG(0.5,pred4.fevf,pred14.fevf,pred24.fevf)
med.fevf5<-qBCCG(0.5,pred5.fevf,pred15.fevf,pred25.fevf)
med.fevf6<-qBCCG(0.5,pred6.fevf,pred16.fevf,pred26.fevf)


plot(height,med.fevf1,col=1,type="l",xlab="Height (in cm)",
  ylab=expression('Height-age adjusted median reference FEV '[1]*' (in
      liters)'),lty=1,lwd=1,
  ylim=c(1,5))
lines(height,med.fevf2,col=2,lty=2,lwd=1)
lines(height,med.fevf3,col=3,lty=3,lwd=1)
lines(height,med.fevf4,col=4,lty=4,lwd=1)
lines(height,med.fevf5,col=5,lty=5,lwd=1)
lines(height,med.fevf6,col=6,lty=6,lwd=1)

legend("topleft",c("age 7","age 9","age 11","age 13","age 15","age 17"),
 col=1:6,lty=1:6)

#########################################################################
#########################################################################
# Contour Plot
#########################################################################
age<-seq(min(agef),max(agef),length=500)
height<-seq(min(heightf),max(heightf),length=500)

newdata<-expand.grid(height=height,age=age)

newdata1<-cbind(1,newdata,poly(newdata$age/20,3,raw=TRUE))
newdata2<-cbind(1,newdata$age,poly(newdata$age/20,3,raw=TRUE))
newdata3<-cbind(1,newdata$age,poly(newdata$age/20,1,raw=TRUE))

newdata1[,2]<-log(newdata1[,2])
newdata1[,3]<-log(newdata1[,3])

newdata2[,2]<-log(newdata2[,2])

newdata3[,2]<-log(newdata3[,2])


pp<-qBCCG(0.5,pred(fit.fevf$mu.coefficients,newdata1),
 pred(fit.fevf$sigma.coefficients,newdata2),pred1(fit.fevf$nu.coefficients,
     newdata3))

pp.mat<-matrix(pp,ncol=length(age))

```

```
785 contour(height,age,pp.mat,xlab="Height (in cm)",ylab="Age (in years)",nlevels
    =20)
786 ##########################################################################
787 age1 <- seq(6,17,length=200)
788 height <- seq(min(heightf), max(heightf), length = 200)
789
790 newdata1<-cbind(1,log(height),log(age1),poly(age1/20,3,raw=TRUE))
791 newdata2<-cbind(1,log(age1),poly(age1/20,3,raw=TRUE))
792 newdata3<-cbind(1,log(age1),poly(age1/20,1,raw=TRUE))
793
794
795 Mpred1.fevf<-pred(fit.fevf$mu.coefficients,newdata1)
796 Spred1.fevf<-pred(fit.fevf$sigma.coefficients,newdata2)
797 Lpred1.fevf<-pred1(fit.fevf$nu.coefficients,newdata3)
798
799 med.fevf<-qBCCG(0.5,Mpred1.fevf,Spred1.fevf,Lpred1.fevf)
800 LLN.fevf<-qBCCG(0.05,Mpred1.fevf,Spred1.fevf,Lpred1.fevf)
801
802
803
804 plot(agef,fev1f,ylim=c(1,5),xlab="Age (in years)",
805   ylab=expression('FEV '[1]*' (in liters)'))
806 lines(age1,med.fevf,col=2,lty=1)
807 lines(age1,LLN.fevf,col=3,lty=2)
808
809 legend("topleft",c(expression('Height-age adjusted median FEV '[1]),
810   expression('Height-age adjusted LLN for FEV '[1])),col=2:3,lty=1:2)
811
812 ##########################################################################
813 ##########################################################################
814 # BOYS: FEV1/FVC
815 ##########################################################################
816 fit1.ffm<-gamlss.fit(fev1fvcm,heightm,agem,2,2,0)
817 fit2.ffm<-gamlss.fit(fev1fvcm,heightm,agem,2,2,1)
818 fit3.ffm<-gamlss.fit(fev1fvcm,heightm,agem,2,2,2)
819 fit4.ffm<-gamlss.fit(fev1fvcm,heightm,agem,3,2,0)
820 fit5.ffm<-gamlss.fit(fev1fvcm,heightm,agem,3,2,1)
821 fit6.ffm<-gamlss.fit(fev1fvcm,heightm,agem,3,2,2)
822 fit7.ffm<-gamlss.fit(fev1fvcm,heightm,agem,3,3,2)
823 fit8.ffm<-gamlss.fit(fev1fvcm,heightm,agem,4,2,1)
824 fit9.ffm<-gamlss.fit(fev1fvcm,heightm,agem,4,3,2)
825 fit10.ffm<-gamlss.fit(fev1fvcm,heightm,agem,5,2,1)
826 fit11.ffm<-gamlss.fit(fev1fvcm,heightm,agem,5,3,2)
827 fit12.ffm<-gamlss.fit(fev1fvcm,heightm,agem,2,2,-1)
828 fit13.ffm<-gamlss.fit(fev1fvcm,heightm,agem,3,2,-1)
829 fit14.ffm<-gamlss.fit(fev1fvcm,heightm,agem,3,3,-1)
830 fit15.ffm<-gamlss.fit(fev1fvcm,heightm,agem,6,2,1)
831 fit16.ffm<-gamlss.fit(fev1fvcm,heightm,agem,5,3,5)
832
833 AIC(fit1.ffm,fit2.ffm,fit3.ffm,fit4.ffm,fit5.ffm,fit6.ffm,fit7.ffm,fit8.ffm,
      fit9.ffm,fit10.ffm,
```

```
  fit11.ffm, fit12.ffm, fit13.ffm, fit14.ffm, fit15.ffm, fit16.ffm)

           df       AIC
fit7.ffm   15 316.5359
fit9.ffm   16 316.8550
fit3.ffm   13 316.9713
fit11.ffm  17 321.2549
fit14.ffm  12 327.5151
fit6.ffm   14 328.1254
fit8.ffm   14 329.2638
fit4.ffm   12 330.1719
fit2.ffm   12 330.4626
fit16.ffm  20 330.4806
fit12.ffm  10 331.1448
fit5.ffm   13 332.0017
fit13.ffm  11 332.3289
fit1.ffm   11 332.9067
fit10.ffm  15 333.6785
fit15.ffm  16 335.5908

fit.ffm<-fit7.ffm
summary(fit.ffm)

#########################################################
age<-seq(min(agem), max(agem), length = 200)
height <- seq(min(heightm), max(heightm), length = 200)

newdata01<-cbind(1,log(height),log(age),poly(age/20,3,raw=TRUE))
newdata02<-cbind(1,log(age),poly(age/20,3,raw=TRUE))
newdata03<-cbind(1,log(age),poly(age/20,2,raw=TRUE))

pred01.ffm<-pred(fit.ffm$mu.coefficients,newdata01)
pred02.ffm<-pred(fit.ffm$sigma.coefficients,newdata02)
pred03.ffm<-pred1(fit.ffm$nu.coefficients,newdata03)

med.ffm0<-qBCCG(0.5,pred01.ffm,pred02.ffm,pred03.ffm)


dd<-scatterplot3d(agem,heightm,fev1fvcm,color="dodgerblue4",pch=20,xlab="Age (
    in years)",ylab="Height (in cm)",
   zlab=expression('FEV '[1]*' / FVC'))
dd$points3d(age,height,med.ffm0,type="l")
#################################################################
age1 <- rep(7,200)
age2 <- rep(9,200)
age3 <- rep(11,200)
age4 <- rep(13,200)
age5 <- rep(15,200)
age6 <- rep(17,200)

height <- seq(min(heightm), max(heightm), length = 200)
```

```
884
885 newdata1<-cbind(1,log(height),log(age1),poly(age1/20,3,raw=TRUE))
886 newdata2<-cbind(1,log(height),log(age2),poly(age2/20,3,raw=TRUE))
887 newdata3<-cbind(1,log(height),log(age3),poly(age3/20,3,raw=TRUE))
888 newdata4<-cbind(1,log(height),log(age4),poly(age4/20,3,raw=TRUE))
889 newdata5<-cbind(1,log(height),log(age5),poly(age5/20,3,raw=TRUE))
890 newdata6<-cbind(1,log(height),log(age6),poly(age6/20,3,raw=TRUE))
891
892 newdata11<-cbind(1,log(age1),poly(age1/20,3,raw=TRUE))
893 newdata12<-cbind(1,log(age2),poly(age2/20,3,raw=TRUE))
894 newdata13<-cbind(1,log(age3),poly(age3/20,3,raw=TRUE))
895 newdata14<-cbind(1,log(age4),poly(age4/20,3,raw=TRUE))
896 newdata15<-cbind(1,log(age5),poly(age5/20,3,raw=TRUE))
897 newdata16<-cbind(1,log(age6),poly(age6/20,3,raw=TRUE))
898
899 newdata21<-cbind(1,log(age1),poly(age1/20,2,raw=TRUE))
900 newdata22<-cbind(1,log(age2),poly(age2/20,2,raw=TRUE))
901 newdata23<-cbind(1,log(age3),poly(age3/20,2,raw=TRUE))
902 newdata24<-cbind(1,log(age4),poly(age4/20,2,raw=TRUE))
903 newdata25<-cbind(1,log(age5),poly(age5/20,2,raw=TRUE))
904 newdata26<-cbind(1,log(age6),poly(age6/20,2,raw=TRUE))
905
906 pred1.ffm<-pred(fit.ffm$mu.coefficients,newdata1)
907 pred2.ffm<-pred(fit.ffm$mu.coefficients,newdata2)
908 pred3.ffm<-pred(fit.ffm$mu.coefficients,newdata3)
909 pred4.ffm<-pred(fit.ffm$mu.coefficients,newdata4)
910 pred5.ffm<-pred(fit.ffm$mu.coefficients,newdata5)
911 pred6.ffm<-pred(fit.ffm$mu.coefficients,newdata6)
912
913 pred11.ffm<-pred(fit.ffm$sigma.coefficients,newdata11)
914 pred12.ffm<-pred(fit.ffm$sigma.coefficients,newdata12)
915 pred13.ffm<-pred(fit.ffm$sigma.coefficients,newdata13)
916 pred14.ffm<-pred(fit.ffm$sigma.coefficients,newdata14)
917 pred15.ffm<-pred(fit.ffm$sigma.coefficients,newdata15)
918 pred16.ffm<-pred(fit.ffm$sigma.coefficients,newdata16)
919
920 pred21.ffm<-pred1(fit.ffm$nu.coefficients,newdata21)
921 pred22.ffm<-pred1(fit.ffm$nu.coefficients,newdata22)
922 pred23.ffm<-pred1(fit.ffm$nu.coefficients,newdata23)
923 pred24.ffm<-pred1(fit.ffm$nu.coefficients,newdata24)
924 pred25.ffm<-pred1(fit.ffm$nu.coefficients,newdata25)
925 pred26.ffm<-pred1(fit.ffm$nu.coefficients,newdata26)
926
927 med.ffm1<-qBCCG(0.5,pred1.ffm,pred11.ffm,pred21.ffm)
928 med.ffm2<-qBCCG(0.5,pred2.ffm,pred12.ffm,pred22.ffm)
929 med.ffm3<-qBCCG(0.5,pred3.ffm,pred13.ffm,pred23.ffm)
930 med.ffm4<-qBCCG(0.5,pred4.ffm,pred14.ffm,pred24.ffm)
931 med.ffm5<-qBCCG(0.5,pred5.ffm,pred15.ffm,pred25.ffm)
932 med.ffm6<-qBCCG(0.5,pred6.ffm,pred16.ffm,pred26.ffm)
933
934
```

```
plot(height,med.ffm1,col=1,type="l",xlab="Height (in cm)",
  ylab=expression('Height-age adjusted median reference FEV '[1]* ' / FVC'),
      lty=1,lwd=1,
  ylim=c(75,102))
lines(height,med.ffm2,col=2,lty=2,lwd=1)
lines(height,med.ffm3,col=3,lty=3,lwd=1)
lines(height,med.ffm4,col=4,lty=4,lwd=1)
lines(height,med.ffm5,col=5,lty=5,lwd=1)
lines(height,med.ffm6,col=6,lty=6,lwd=1)

legend("topright",c("age 7","age 9","age 11","age 13","age 15","age 17"),
 col=1:6,lty=1:6)

#########################################################################
#########################################################################
# Contour Plot
#########################################################################
age<-seq(min(agem),max(agem),length=500)
height<-seq(min(heightm),max(heightm),length=500)

newdata<-expand.grid(height=height,age=age)

newdata1<-cbind(1,newdata,poly(newdata$age/20,3,raw=TRUE))
newdata2<-cbind(1,newdata$age,poly(newdata$age/20,3,raw=TRUE))
newdata3<-cbind(1,newdata$age,poly(newdata$age/20,2,raw=TRUE))

newdata1[,2]<-log(newdata1[,2])
newdata1[,3]<-log(newdata1[,3])

newdata2[,2]<-log(newdata2[,2])

newdata3[,2]<-log(newdata3[,2])


pp<-qBCCG(0.5,pred(fit.ffm$mu.coefficients,newdata1),
 pred(fit.ffm$sigma.coefficients,newdata2),pred1(fit.ffm$nu.coefficients,
     newdata3))

pp.mat<-matrix(pp,ncol=length(age))

contour(height,age,pp.mat,xlab="Height (in cm)",ylab="Age (in years)",nlevels
    =10)
#########################################################################
age1 <- seq(6,17,length=200)
height <- seq(min(heightm), max(heightm), length = 200)

newdata1<-cbind(1,log(height),log(age1),poly(age1/20,3,raw=TRUE))
newdata2<-cbind(1,log(age1),poly(age1/20,3,raw=TRUE))
newdata3<-cbind(1,log(age1),poly(age1/20,2,raw=TRUE))


```

```
Mpred1.ffm<-pred(fit.ffm$mu.coefficients,newdata1)
Spred1.ffm<-pred(fit.ffm$sigma.coefficients,newdata2)
Lpred1.ffm<-pred1(fit.ffm$nu.coefficients,newdata3)

med.ffm<-qBCCG(0.5,Mpred1.ffm,Spred1.ffm,Lpred1.ffm)
LLN.ffm<-qBCCG(0.05,Mpred1.ffm,Spred1.ffm,Lpred1.ffm)



plot(agem,fev1fvcm,ylim=c(70,105),xlab="Age (in years)",
  ylab=expression('FEV '[1]*' / FVC'))
lines(age1,med.ffm,col=2,lty=1)
lines(age1,LLN.ffm,col=3,lty=2)

legend("topleft",c(expression('Height-age adjusted median FEV '[1]*' / FVC'),
  expression('Height-age adjusted LLN for FEV '[1]*' / FVC')),col=2:3,lty
    =1:2)

##############################################################################
##############################################################################
# GIRLS: FEV1/FVC
##############################################################################
fit1.fff<-gamlss.fit(fev1fvcf,heightf,agef,2,2,0)
fit2.fff<-gamlss.fit(fev1fvcf,heightf,agef,2,2,1)
fit3.fff<-gamlss.fit(fev1fvcf,heightf,agef,2,2,2)
fit4.fff<-gamlss.fit(fev1fvcf,heightf,agef,3,2,0)
fit5.fff<-gamlss.fit(fev1fvcf,heightf,agef,3,2,1)
fit6.fff<-gamlss.fit(fev1fvcf,heightf,agef,3,2,2)
fit7.fff<-gamlss.fit(fev1fvcf,heightf,agef,3,3,2)
fit8.fff<-gamlss.fit(fev1fvcf,heightf,agef,4,2,1)
fit9.fff<-gamlss.fit(fev1fvcf,heightf,agef,4,3,2)
fit10.fff<-gamlss.fit(fev1fvcf,heightf,agef,5,2,1)
fit11.fff<-gamlss.fit(fev1fvcf,heightf,agef,5,3,2)
fit12.fff<-gamlss.fit(fev1fvcf,heightf,agef,2,2,-1)
fit13.fff<-gamlss.fit(fev1fvcf,heightf,agef,3,2,-1)
fit14.fff<-gamlss.fit(fev1fvcf,heightf,agef,3,3,-1)
fit15.fff<-gamlss.fit(fev1fvcf,heightf,agef,6,2,1)
fit16.fff<-gamlss.fit(fev1fvcf,heightf,agef,5,3,5)
fit17.fff<-gamlss.fit(fev1fvcf,heightf,agef,3,2,4)

AIC(fit1.fff,fit2.fff,fit3.fff,fit4.fff,fit5.fff,fit6.fff,fit7.fff,fit8.fff,
    fit9.fff,fit10.fff,
  fit11.fff,fit12.fff,fit13.fff,fit14.fff,fit15.fff,fit16.fff,fit17.fff)

           df       AIC
fit13.fff  11  464.3573
fit4.fff   12  466.1543
fit14.fff  12  466.3556
fit12.fff  10  467.3180
fit5.fff   13  467.9000
fit8.fff   14  469.1194
```

```
fit1.fff   11  469.1888
fit6.fff   14  469.8260
fit10.fff  15  470.8065
fit15.fff  16  471.1551
fit2.fff   12  471.1884
fit7.fff   15  471.8128
fit9.fff   16  472.9948
fit17.fff  16  473.0401
fit3.fff   13  473.1880
fit11.fff  17  473.8371
fit16.fff  20  479.2822

fit.fff<-fit5.fff
summary(fit.fff)

#########################################################
age<-seq(min(agef), max(agef), length = 200)
height <- seq(min(heightf), max(heightf), length = 200)

newdata01<-cbind(1,log(height),log(age),poly(age/20,3,raw=TRUE))
newdata02<-cbind(1,log(age),poly(age/20,2,raw=TRUE))
newdata03<-cbind(1,log(age),poly(age/20,1,raw=TRUE))

pred01.fff<-pred(fit.fff$mu.coefficients,newdata01)
pred02.fff<-pred(fit.fff$sigma.coefficients,newdata02)
pred03.fff<-pred1(fit.fff$nu.coefficients,newdata03)

med.fff0<-qBCCG(0.5,pred01.fff,pred02.fff,pred03.fff)

dd<-scatterplot3d(agef,heightf,fev1fvcf,color="dodgerblue4",pch=20,xlab="Age (
    in years)",ylab="Height (in cm)",
   zlab=expression('FEV '[1]*' / FVC'))
dd$points3d(age,height,med.fff0,type="l")
#####################################################################
age1 <- rep(7,200)
age2 <- rep(9,200)
age3 <- rep(11,200)
age4 <- rep(13,200)
age5 <- rep(15,200)
age6 <- rep(17,200)

height <- seq(min(heightf), max(heightf), length = 200)

newdata1<-cbind(1,log(height),log(age1),poly(age1/20,3,raw=TRUE))
newdata2<-cbind(1,log(height),log(age2),poly(age2/20,3,raw=TRUE))
newdata3<-cbind(1,log(height),log(age3),poly(age3/20,3,raw=TRUE))
newdata4<-cbind(1,log(height),log(age4),poly(age4/20,3,raw=TRUE))
newdata5<-cbind(1,log(height),log(age5),poly(age5/20,3,raw=TRUE))
newdata6<-cbind(1,log(height),log(age6),poly(age6/20,3,raw=TRUE))

newdata11<-cbind(1,log(age1),poly(age1/20,2,raw=TRUE))
```

```
1082 newdata12<-cbind(1,log(age2),poly(age2/20,2,raw=TRUE))
1083 newdata13<-cbind(1,log(age3),poly(age3/20,2,raw=TRUE))
1084 newdata14<-cbind(1,log(age4),poly(age4/20,2,raw=TRUE))
1085 newdata15<-cbind(1,log(age5),poly(age5/20,2,raw=TRUE))
1086 newdata16<-cbind(1,log(age6),poly(age6/20,2,raw=TRUE))
1087
1088 newdata21<-cbind(1,log(age1),poly(age1/20,1,raw=TRUE))
1089 newdata22<-cbind(1,log(age2),poly(age2/20,1,raw=TRUE))
1090 newdata23<-cbind(1,log(age3),poly(age3/20,1,raw=TRUE))
1091 newdata24<-cbind(1,log(age4),poly(age4/20,1,raw=TRUE))
1092 newdata25<-cbind(1,log(age5),poly(age5/20,1,raw=TRUE))
1093 newdata26<-cbind(1,log(age6),poly(age6/20,1,raw=TRUE))
1094
1095 pred1.fff<-pred(fit.fff$mu.coefficients,newdata1)
1096 pred2.fff<-pred(fit.fff$mu.coefficients,newdata2)
1097 pred3.fff<-pred(fit.fff$mu.coefficients,newdata3)
1098 pred4.fff<-pred(fit.fff$mu.coefficients,newdata4)
1099 pred5.fff<-pred(fit.fff$mu.coefficients,newdata5)
1100 pred6.fff<-pred(fit.fff$mu.coefficients,newdata6)
1101
1102 pred11.fff<-pred(fit.fff$sigma.coefficients,newdata11)
1103 pred12.fff<-pred(fit.fff$sigma.coefficients,newdata12)
1104 pred13.fff<-pred(fit.fff$sigma.coefficients,newdata13)
1105 pred14.fff<-pred(fit.fff$sigma.coefficients,newdata14)
1106 pred15.fff<-pred(fit.fff$sigma.coefficients,newdata15)
1107 pred16.fff<-pred(fit.fff$sigma.coefficients,newdata16)
1108
1109 pred21.fff<-pred1(fit.fff$nu.coefficients,newdata21)
1110 pred22.fff<-pred1(fit.fff$nu.coefficients,newdata22)
1111 pred23.fff<-pred1(fit.fff$nu.coefficients,newdata23)
1112 pred24.fff<-pred1(fit.fff$nu.coefficients,newdata24)
1113 pred25.fff<-pred1(fit.fff$nu.coefficients,newdata25)
1114 pred26.fff<-pred1(fit.fff$nu.coefficients,newdata26)
1115
1116 med.fff1<-qBCCG(0.5,pred1.fff,pred11.fff,pred21.fff)
1117 med.fff2<-qBCCG(0.5,pred2.fff,pred12.fff,pred22.fff)
1118 med.fff3<-qBCCG(0.5,pred3.fff,pred13.fff,pred23.fff)
1119 med.fff4<-qBCCG(0.5,pred4.fff,pred14.fff,pred24.fff)
1120 med.fff5<-qBCCG(0.5,pred5.fff,pred15.fff,pred25.fff)
1121 med.fff6<-qBCCG(0.5,pred6.fff,pred16.fff,pred26.fff)
1122
1123
1124 plot(height,med.fff1,col=1,type="l",xlab="Height (in cm)",
1125   ylab=expression('Height-age adjusted median reference FEV '[1]* ' / FVC'),
        lty=1,lwd=1,
1126   ylim=c(75,102))
1127 lines(height,med.fff2,col=2,lty=2,lwd=1)
1128 lines(height,med.fff3,col=3,lty=3,lwd=1)
1129 lines(height,med.fff4,col=4,lty=4,lwd=1)
1130 lines(height,med.fff5,col=5,lty=5,lwd=1)
1131 lines(height,med.fff6,col=6,lty=6,lwd=1)
```

```
legend("topright",c("age 7","age 9","age 11","age 13","age 15","age 17"),
 col=1:6,lty=1:6)
#############################################################################
#############################################################################
# Contour Plot
#############################################################################
age<-seq(min(agef),max(agef),length=500)
height<-seq(min(heightf),max(heightf),length=500)

newdata<-expand.grid(height=height,age=age)

newdata1<-cbind(1,newdata,poly(newdata$age/20,3,raw=TRUE))
newdata2<-cbind(1,newdata$age,poly(newdata$age/20,2,raw=TRUE))
newdata3<-cbind(1,newdata$age,poly(newdata$age/20,1,raw=TRUE))

newdata1[,2]<-log(newdata1[,2])
newdata1[,3]<-log(newdata1[,3])
newdata2[,2]<-log(newdata2[,2])
newdata3[,2]<-log(newdata3[,2])

pp<-qBCCG(0.5,pred(fit.fff$mu.coefficients,newdata1),
 pred(fit.fff$sigma.coefficients,newdata2),pred1(fit.fff$nu.coefficients,
     newdata3))

pp.mat<-matrix(pp,ncol=length(age))

contour(height,age,pp.mat,xlab="Height (in cm)",ylab="Age (in years)",nlevels
    =10)
#############################################################################
age1 <- seq(6,17,length=200)
height <- seq(min(heightf), max(heightf), length = 200)

newdata1<-cbind(1,log(height),log(age1),poly(age1/20,3,raw=TRUE))
newdata2<-cbind(1,log(age1),poly(age1/20,2,raw=TRUE))
newdata3<-cbind(1,log(age1),poly(age1/20,1,raw=TRUE))

Mpred1.fff<-pred(fit.fff$mu.coefficients,newdata1)
Spred1.fff<-pred(fit.fff$sigma.coefficients,newdata2)
Lpred1.fff<-pred1(fit.fff$nu.coefficients,newdata3)

med.fff<-qBCCG(0.5,Mpred1.fff,Spred1.fff,Lpred1.fff)
LLN.fff<-qBCCG(0.05,Mpred1.fff,Spred1.fff,Lpred1.fff)

plot(agef,fev1fvcf,xlab="Age (in years)",
  ylab=expression('FEV '[1]* ' / FVC'),ylim=c(70,105))
lines(age1,med.fff,col=2,lty=1)
lines(age1,LLN.fff,col=3,lty=2)

legend("topleft",c(expression('Height-age adjusted median FEV '[1]* ' / FVC'),
  expression('Height-age adjusted LLN for FEV '[1]* ' / FVC')),col=2:3,lty
     =1:2)
```

```
1180
1181 ##########################################################################
1182 ## Comparison between the curves of males and females ####################
1183 ##########################################################################
1184 age1 <- seq(6,17,length=200)
1185 height <- seq(min(heightm), max(heightm), length = 200)
1186
1187 newdata1<-cbind(1,log(height),log(age1),poly(age1/20,4,raw=TRUE))
1188 newdata2<-cbind(1,log(age1),poly(age1/20,3,raw=TRUE))
1189 newdata3<-cbind(1,log(age1),poly(age1/20,1,raw=TRUE))
1190
1191 Mpred1.fvcm<-pred(fit.fvcm$mu.coefficients,newdata1)
1192 Spred1.fvcm<-pred(fit.fvcm$sigma.coefficients,newdata2)
1193 Lpred1.fvcm<-pred1(fit.fvcm$nu.coefficients,newdata3)
1194
1195 med.fvcm<-qBCCG(0.5,Mpred1.fvcm,Spred1.fvcm,Lpred1.fvcm)
1196 LLN.fvcm<-qBCCG(0.05,Mpred1.fvcm,Spred1.fvcm,Lpred1.fvcm)
1197
1198 plot(age1,med.fvcm,type="l",col=1,lty=1,xlab="Age (in years)",
1199  ylab="Height-age adjusted median reference values (in liters)",ylim=c(1,6))
1200 lines(age1,med.fevm,col=1,lty=2)
1201 lines(age1,med.fvcf,col=2,lty=1)
1202 lines(age1,med.fevf,col=2,lty=2)
1203
1204 legend("topleft",c("Male FVC",expression('Male FEV '[1]),
1205     "Female FVC",expression('Female FEV '[1]) ),col=c(1,1,2,2),lty=c(1,2,1,2))
1206
1207 ########################################################################
1208 #### Comaprisons with Other Study (Hankinson et. al, 1999) ####
1209 ########################################################################
1210 hankinson.fvcm <- -0.2584 - 0.20415*age + 0.010133*age*age + 0.00018642*height
        *height
1211 hankinson.fev1m <- -0.7453 - 0.04106*age + 0.004477*age*age + 0.00014098*
        height*height
1212 hankinson.fev1fvcm <- 88.066 - 0.2066*age
1213
1214 par(mar=c(5,5,4,1))
1215 dd<-scatterplot3d(agem,heightm,fvcm,color="dodgerblue4",pch=20,xlab="Age (in
        years)",ylab="Height (in cm)",zlab="FVC (in liters)",cex.lab=1.2, cex.axis
        =1.5, cex.main=1.5, cex.sub=1.5)
1216 dd$points3d(age,height,med.fvcm0,type="l",lwd=2)
1217 dd$points3d(age,height,hankinson.fvcm,type="l",col="red",lwd=2)
1218
1219 dd<-scatterplot3d(agem,heightm,fev1m,color="dodgerblue4",pch=20,xlab="Age (in
        years)",ylab="Height (in cm)", zlab=expression('FEV '[1]* ' (in liters)')
        )
1220 dd$points3d(age,height,med.fevm0,type="l", col = "red")
1221 dd$points3d(age,height,hankinson.fev1m,type="l", col = "chartreuse4")
1222
1223 dd<-scatterplot3d(agem,heightm,fev1fvcm,color="dodgerblue4",pch=20,xlab="Age (
        in years)",ylab="Height (in cm)", zlab=expression('FEV '[1]*' / FVC'))
```

```
dd$points3d(age,height,med.ffm0,type="l", col = "red")
dd$points3d(age,height,hankinson.fev1fvcm,type="l", col = "chartreuse4")

hankinson.fvcf <- -1.2082 + 0.05916*age + 0.00014815*height*height
hankinson.fev1f <- -0.8710 + 0.06537*age + 0.00011496*height*height
hankinson.fev1fvcf <- 90.809 - 0.2125*age

par(mar=c(5,5,4,1))
dd<-scatterplot3d(agef,heightf,fvcf,color="dodgerblue4",pch=20,xlab="Age (in
    years)",ylab="Height (in cm)",zlab="FVC (in liters)",cex.lab=1.2, cex.axis
    =1.5, cex.main=1.5, cex.sub=1.5)
dd$points3d(age,height,med.fvcf0,type="l", lwd=2)
dd$points3d(age,height,hankinson.fvcf,type="l", col = "red",lwd=2)


dd<-scatterplot3d(agef,heightf,fev1f,color="dodgerblue4",pch=20,xlab="Age (in
    years)",ylab="Height (in cm)", zlab=expression('FEV '[1]* ' (in liters)')
    )
dd$points3d(age,height,med.fevf0,type="l", col = "red")
dd$points3d(age,height,hankinson.fev1f,type="l", col = "chartreuse4")

dd<-scatterplot3d(agef,heightf,fev1fvcf,color="dodgerblue4",pch=20,xlab="Age (
    in years)",ylab="Height (in cm)", zlab=expression('FEV '[1]*' / FVC'))
dd$points3d(age,height,med.fff0,type="l", col = "red")
dd$points3d(age,height,hankinson.fev1fvcf,type="l", col = "chartreuse4")
```