Aida Mashkouri Najafi

Spirometry in healthy children from birth to school age

MASTER THESIS

written to obtain the academic degree of a Master of Science (MSc)

Operations Research and Statistics

Graz University of Technology

Supervisors:
Univ.-Prof. Dipl.-Ing. Dr. Ernst Stadlober
Institute of Statistics
Dr. Sooky Loom
Portex Unit. UCL Institute of Child Health

Graz, June 2013

Eidesstattliche Erklärung

Ich erkläre an Eides statt, dass ich die vorliegende Arbeit selbständig verfasst, andere als die angegebenen Quellen/Hilfsmittel nicht benutzt, und die den benutzten Quellen wörtlich und inhaltlich entnommenen Stellen als solche kenntlich gemacht habe.

Graz, am ....
(Unterschrift)

Statutory Declaration

I declare that I have authored this thesis independently, that I have not used other than the declared sources/resources, and that I have explicitly marked all material which has been quoted either literally or by content from the used sources.

....
date (signature)

Acknowledgement

This dissertation would not have been possible without the guidance and the help of several individuals who in one way or another contributed and extended their valuable assistance in the preparation and completion of this study. First and foremost, I would like to express my gratitude to my supervisor, Prof. Ernst Stadlober. It has been a great honour to be his student. His invaluable scientific supervision on my research projects and his kindness and support across these years helped me to be here that I am now.

Next I want to thank Prof. Janet Stock for giving me this great opportunity to do my master thesis at porthex unit, Institute of Child health at UCL. I would like to express my appreciation to Dr. Sooky Lum for her enormous help and support at medical aspects of this research. I am grateful of her generous support and patience that helped me to finish this research. I also want to add, that it was a great pleasure working with Dr. Angie Wade, her perspective on statistical techniques was invaluable. The most special thanks goes to my best partner and friend, my husband. Amir, you gave me your unconditional support and love through all this long process. Without you helping me always to see the other full side of the glass I would not be able to be here. I want to also thank my parents, Mohammad and Azar, who have supported me unconditionally over the years in all aspects of my life. You taught me to be strong and go forward with all the challenges on my way. Alaleh, simply thanks for all your support little sister.

Abstract

Longitudinal studies refer to investigations where measurements are collected at multiple follow-ups and it is always a matter of interest to access the change in these measurements. Our research has been focused on studying the effect of growth on Forced Expiratory Flows and Volumes outcomes such as $FEV_{0.5}$, $FEV_{0.75}$, FEV_1, FEF_{25-75} in healthy children. It is studied that to what extent these values track within and between each individual, i.e. is it possible to predict FEV_1 at school age from $FEV_{0.5}$ measures in infancy or childhood? Healthy infants between 3 weeks to 2.5 years were recruited applying the $RVRTC$ technique and 3 to 12 years children were recruited in other research units of Great Britain to collect the spirometry measurements. Results from these healthy controls were expressed as z-scores using published prediction equations of Jones et al. (2000) and Stanojevic et al. (2008). To address change in FEF lung function parameters between between different growth stages, the Bland-Altman technique is applied. These were further explored using multilevel regression modelling (**MLwiN** version 2.12 and **R** version 2.15) to study the within and between subjects variability for various important respiratory parameters in growth.

Zusammenfassung

Longitudinale Studien sind Untersuchungen bei denen Messungen an aufeinander folgenden Zeitpunkten erfolgen, um Änderungen im Zeitverlauf zu studieren. In unserer Arbeit wird der Einfluss des Wachstums von gesunden Kindern auf forcierte Atemflüsse und Volumina wie $FEV_{0.5}$, $FEV_{0.75}$, FEV_1 und FEV_{25-75} studiert. Es wird untersucht in welchem Ausmaß diese Werte innerhalb der Individuen und zwischen den Individuen variieren. D.h. ist es zum Beispiel möglich die Einsekundenkapazität FEV_1 im Schulalter auf Grund von $FEV_{0.5}$-Messungen im Säuglings– und Kleinkindalter zu prognostizieren? Dafür wurden gesunde Babies im Alter zwischen 3 Wochen und 2.5 Jahren ausgesucht und an ihnen Messungen mit Hilfe der $RVRTC$–Technik durchgeführt. Spirometrische Messungen von Kindern im Alter von 3 bis 12 Jahren wurden an verschiedenen Forschungseinrichtungen in Großbritannien vorgenommen. Die Resultate dieser Messungen werden als z–Scores der von Jones et al. (2000) und Stanojevic et al. (2008) publizierten Vorhersagegleichungen ausgedrückt. Zur Erfassung der Unterschiede von FEV-Lungenfunktionsparametern zwischen verschiedenen Wachstumsstufen wird die Bland-Altman Technik angewandt. Die Variabilität von wichtigen respiratorischen Parametern in Abhängigkeit vom Wachstum innerhalb und zwischen Individuen wird durch einen mehrstufigen Regressionsansatz (**MLwiN** version 2.12 und **R** version 2.15) modelliert.

Abbreviations

Abbreviation	Definition
PFT	Pulmonary Function Test
$RVRTC$	Raised Volume Rapid Thoracic Compression
FEV_1	Forced Expiratory Volume in 1 Second (l)
$FEV_{0.5}$	Forced Expiratory Volume in 0.5 Second (l)
$FEV_{0.75}$	Forced Expiratory Volume in 0.75 Second (l)
FVC	Forced Vital Capacity (l)
FEF	Forced Expiratory Flow at specific lung volumes (l)
FEF_{25-75}	Forced Expiratory Flow at volumes 25–75%
CF	Cystic Fibrosis
LLN	Lower Limit of Normal
ULN	Upper Limit of Normal
CV	Coefficient of Variation (between-subject variability)
PSA	Preschool age
SA	School age
ICH	Institute of child health

Contents

List of Figures

List of Tables

Chapter 1

Introduction

Globally the respiratory diseases causing long term illnesses, which decreases the quality of life of patients and also brings a lot of financial pressure on families and health care systems. Respiratory diseases in early childhood such as Asthma and Cystic Fibrosis can leave critical problems and permanent damages in the respiratory system later in life [1].

1.1 Diagnosis of respiratory diseases in children

A complete understanding of the normal growth and development of the respiratory system is necessary in order to distinguish the effects of the disease from those of normal physiology. Besides, availability of longitudinal observations from birth to adulthood can provide insight into how the disease progresses throughout different stages of growth and development.

Pulmonary function tests (PFTs) are an integral part of the clinical assessment of respiratory diseases, providing an objective means of measuring the airways and the function of the respiratory system [1].

The aim of this chapter is to describe briefly the forced expiratory flow volume curves and different methods and techniques of measuring these values from infancy to school-age. Then the normative values, interpretation of results being expressed as z-scores and the potential applications of these techniques for the various age groups will be discussed.

1.2 Pulmonary function tests ($PFTs$)

Pulmonary function testing has been a major step forward in assessing the functional status of the lungs and is a valuable tool for evaluating the respiratory system, representing an important supplement to the patient history and various lung imaging studies [1].

1.2.1 Pulmonary function tests in infants

Since infants are not able to perform forced expiratory manoeuvres, the partial forced expiratory flow-volume ($PEFV$) curves can be produced by wrapping an inflatable jacket around infant's chest and abdomen, while allowing them to breath through a face mask

and flow meter. This technique is referred to Tidal Rapid Thoraco-abdominal Compression (RTC). After at least five regular tidal breaths, an initial jacket pressure (P_j) of 2-3 kPa is usually applied at the end of inspiration, which provides a pressure around the chest and abdomen to force expiration. This trial will regenerate with incremental pressure of 1-2 kPa until a maximum P_j has reached at which the highest flow is obtained [1].

The raised volume RTC ($RVRTC$) technique is an adaptation of the tidal RTC technique, where the infant's lung are passively inflated toward total lung capacity (TLC) using a preset pressure before applying RTC. This enables full forced expiratory manoeuvres to be obtained in infants as in adults [1].

1.2.2 Pulmonary function tests in preschool and school age children

Preschool children (aged 3 to 6 years) are the challenging cases. These children are generally too old to sedate for pulmonary function testing (PFT), as is done with infants, and measurement of lung function under anaesthesia is neither ethically acceptable nor physiologically relevant to clinical management. Children in this age group are not able to voluntarily perform many of the physiological manoeuvres required for the pulmonary function tests used in older children and adults. They have a short attention span and are easily distracted. Due to these issues, the children need to be engaged and encouraged by the operator to participate in the test [2]. There has been a substantial increase in the number of studies designed to evaluate and apply PFTs to preschool children. American Thoracic Society (ATS)/ European Respiratory Society (ERS) paediatric pulmonary function test task force published recommendations for PFTs for children aged 3–6 [2].

School age children (more than 6 years old) can perform many of the tests available to adults and the majority of these children can perform these tests to meet international quality standards.

1.3 The use of reference ranges to interpret results

The interpretation of many medical observations, including PFTs, relies on the availability of normative reference data. To recognize the effects of disease from normal variability in the population and also to interpret the medical measurements, normative reference data are required. The main reason for having normative reference data is that it can be used to represent the range of values expected in a healthy population. As a result, it is important to choose "healthy" subjects to make the reference population that defines normal.

Since the human respiratory system is growing in childhood, the reference data in this area depends on growth, therefore the height and weight could be the two main indicators of reference range. The reference data should have a uniform distribution across the height or weight range studied [3].

1.4 Interpreting reference data

To interpret results appropriately, the reference sample should be large enough to ensure that the extreme limits of "normal" can be estimated with reasonable precision. In order to be able to interpret the reference ranges, the assumption that 95% of the values will fall within approximately two standard deviations ($mean \pm 2sd$) is needed. Values outside this range are not necessarily abnormal, these observations are unusual as they lie outside of the reference range within approximately 95% of the population lies.

In respiratory medicine, the percent predicted [(observed/predicted)*100] is the most commonly used term to describe an individual spirometric observation, in relation to the reference population, however percent predicted does not consider the variability of values around the mean.

A better approach to reporting lung function rather than percent predicted is to express the results as z-scores (also commonly known as Standard Deviation Score (SDS)). The z-score is a mathematical combination of the percent predicted and between subject variability to give a single number that accounts for age, sex, and height related lung function variability expected within comparable healthy individuals.

In terms of representing a healthy population, 95% of population should fall into approximately ± 1.96 z-scores. z-scores are useful for tracking changes in lung function with growth or treatment, as they allow comparison of lung function results obtained with different techniques.

The SDS or z-score of a child's measurement y is calculated from the L(skewness), M(Median) and S(coefficient of variation) curves, using values appropriate for the child's sex, age and height [4]. The following two equations are relevant depending on the value of L:

$$z = \frac{(y/M)^L - 1}{L \times S}, L \neq 0, \tag{1.1}$$

$$z = \frac{\log(y/M)}{S}, L = 0. \tag{1.2}$$

The LMS method provides a way of obtaining normalized growth centile standards, and deals quite generally with skewness in the distribution of the measurement like age and height.

It assumes that the data can be normalized by using a power transformation, which shrinks one tail and stretches the other therefore removing the skewness. The optimal power to obtain normality is calculated for each of a series of age groups and the trend summarized by a smooth (L) curve. Trends in the mean (M) and coefficient of variation (S) are similarly smoothed. The resulting L, M and S curves contain the information to draw any centile curve, and to convert measurements (even extreme values) into exact Standard Deviation Scores [4].

Figure 1.1 illustrates the relationship between centiles and z-scores.

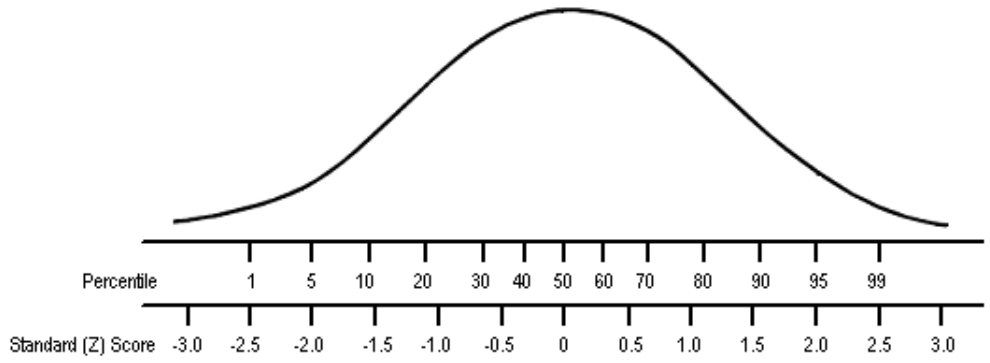


Figure 1.1: z-score centile comparison plot.

Chapter 2

Longitudinal Data Analysis

2.1 Basic Concepts of Longitudinal Data Analysis

This chapter covers some of the main characteristics of longitudinal data and the models designed for these family of data [5]. With longitudinal data, one can measure and compare the outcomes for each individual over time. This property provides the ability of studying the changes over time within subjects and changes over time between groups. Statistical models for this type of data estimates both individual-level regression parameters and population-level regression parameters.

In longitudinal data the set of measurements on each subject are correlated and measurements on the same subject close in time tend to be more highly correlated than measurements far apart in time and also the variance of longitudinal often changes over time. These potential patterns of correlation and variation may combine to produce a complicated covariance structure. This covariance structure must be taken into account to draw valid statistical inferences. Therefore, standard regression and $ANOVA$ models may not be the appropriate choice since it will produce invalid results because two of the parametric assumptions (independent observations and equal variances) may not be valid [6].

2.2 Features of Longitudinal Data

As mentioned in the previous section, in a longitudinal study participants are measured repeatedly at different occasions or times. If the number and timing of the repeated measurements are the same for all individuals, regardless of whether the occasions of the measurements are equally or unequally distributed throughout the duration of the study, this studies refers as being "*balanced*" over time. On the other hand it is an almost inescapable feature of longitudinal studies in the medical sciences, especially in those were the repeated measurements extend over relatively long duration, that some individuals will miss their scheduled visit or date of observation, or even the test occasions were different between the subjects. Therefore the sequence of observation times is no longer equal in all individuals. This case refers to the data as being "*unbalanced*" over time [5].

2.2.1 Dependence and Correlation

Let Y_{ij} indicate the response variable of the i^{th} $(i = 1, \ldots, N)$ individual at its j^{th} $(j = 1, \ldots, n)$ measurement occasion. The mean or expectation of each response Y_{ij} is denoted by $\mu_{ij} = \mathrm{E}(Y_{ij})$. A double letter subscript is used for mean, in order to allow the changes in mean response to vary between individuals as a function of individual level covariates.

In longitudinal studies, the response of an individual in one occasion is very likely to be predictive of the response of the same individual at a future test occasion, therefore the assumption of independent observations in linear regression models is not appropriate anymore. The conditional variance of Y_{ij} is defined as,

$$\sigma_j^2 = E(Y_{ij} - E(Y_{ij}))^2 = E(Y_{ij} - \mu_{ij})^2,$$

where μ_{ij} indicates a measure of the location of the center of the distribution of Y_{ij}, while the variance provides us a measure of spread of the value Y_{ij} around their conditional mean. Because of dependency between response variables in a longitudinal study, the conditional covariance between the responses at two different test occasions, say Y_{ij}, Y_{ik}, is defined by

$$\sigma_{jk} = E\left((Y_{ij} - \mu_{ij})(Y_{ik} - \mu_{ik})\right) \ .$$

Degree of dependence between the Y_{ij} and Y_{ik} and also their units of measurements defines the significance of the covariance. However this is quit difficult to interpret without comparison to the variability of the two variables. The sign of the covariance shows the positive or negative dependence between the two variables.

Correlation between Y_{ij} and Y_{ik} indicates the linear dependence between the two. This is denoted by:

$$\rho_{jk} = \frac{E\left((Y_{ij} - \mu_{ij})(Y_{ik} - \mu_{ik})\right)}{\sigma_j \sigma_k},$$

where σ_j, σ_k are the conditional standard deviations of Y_{ij} and Y_{ik}, respectively. In longitudinal study, the repeated measures on the same individual are predicted to be positively correlated. When n repeated measures are collected into a vector $\mathbf{Y_i}$, the variance-covariance matrix can be defined as

$$Cov\begin{pmatrix} Y_{i1} \\ \vdots \\ Y_{in} \end{pmatrix} = \begin{pmatrix} Var(Y_{i1}) & Cov(Y_{i1}, Y_{i2}) & \ldots & Cov(Y_{i1}, Y_{in}) \\ Cov(Y_{i2}, Y_{i1}) & Var(Y_{i2}) & \ldots & Cov(Y_{i2}, Y_{in}) \\ \vdots & \vdots & \ddots & \vdots \\ Cov(Y_{in}, Y_{i1}) & Cov(Y_{in}, Y_{i2}) & \ldots & Var(Y_{in}) \end{pmatrix} =$$

$$\begin{pmatrix} \sigma_1^2 & \sigma_{12} & \ldots & \sigma_{1n} \\ \sigma_{21} & \sigma_2^2 & \ldots & \sigma_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ \sigma_{n1} & \sigma_{n2} & \ldots & \sigma_n^2 \end{pmatrix}$$

2.2.2 Between-Individual Heterogeneity

In clinical trials, observing heterogeneity among individuals is a norm. Some individuals consistently respond higher or lower than average. Thus, the variability between individuals can be seen as a source of the positive correlation of the repeated measurements. It means for individuals with a high response value at one test occasion, it is expected to have a high value in the following time points. In many cases, these personal characteristics may be unobservable, leading to unexplained heterogeneity in the population. Modelling this unobserved heterogeneity in terms of variance components that describe subject-level effects is one way to provide the correlation of the repeated responses over time and to better describe individual differences in the statistical characterization of the observed data. Between-individual variability can be answered by the introduction of individual-specific "random effects" (e.g. randomly varying intercept and slopes.) in statistical models for longitudinal data.

2.2.3 Within-Individual Variation

Many health outcomes of interest in the clinical laboratory can vary over an individual's lifetime, simply because of natural biological factors (within-individual biological variability). A sequence of repeated measures on any particular individual will vary around some homoeostatic set points in a random manner. Many of these variables can be thought of as realization of some biological process or combination of biological processes operating within the individual that vary over time. This variability is sometimes referred to as the inherent within-individual biological variability [7].

2.2.4 Errors of Measurements

In addition to heterogeneity in the population that leads to subject-specific deviations from the overall response pattern, there are also often short-term correlated errors of measurement. This describes the unexplained variability in the response variable, caused by the measurement process. For example if simultaneously two measurements taken from a subject but divided to smaller samples to get analysed separately such as blood test, the resulting values will be different with the amount of random error. This can clearly interpret the reason to the unexplained bias from 1 in correlation of two closely measured variables. Generally, the greater the measurement error, the weaker the correlation between repeated observations.

2.3 A Model for Longitudinal Data

Often longitudinal data are highly unbalanced. Due to this unbalanced characteristics, many longitudinal data sets cannot be analysed using multivariate regression techniques. However in many cases subject specific profiles can be estimated by linear regression functions. First, the vector of repeated measurements for each subject can be summarized by a

vector of a relatively small number of estimated subject-specific regression coefficients. In a second stage, multivariate regression techniques can be used to relate these estimates to known covariates such as treatment, baseline characteristic and so forth. This so-called two stage analysis [5] will be introduced in this section. Afterwards in section 2.5 the general linear mixed model as a result of combining the two stage model into one single statistical model will be over viewed.

2.4 A Two-Stage Analysis

Stage 1

For the i^{th} individual, let the random variable Y_{ij} represent the response of interest for $i = 1, \cdots, N$, $j = 1, \ldots, n_i$. Let Y_{ij} be measured at time t_{ij} and for the i^{th} subject, let $\mathbf{Y_i} = (Y_{i1}, Y_{i2}, \ldots, Y_{in_i})$ be the vector representation of all repeated measurements.

Assume that Z_i is a $(n_i \times q)$ matrix of known covariates which describes how response variables for i^{th} subject change over time. In the first stage, we assume that $\mathbf{Y_i}$ satisfies the linear regression model [5]

$$\boldsymbol{Y_i} = Z_i \boldsymbol{\alpha_i} + \boldsymbol{\epsilon_i}, \tag{2.1}$$

where, the q-dimensional vector $\boldsymbol{\alpha_i} = (\alpha_{i1}, \ldots, \alpha_{iq})'$ is the set of unknown subject-specific regression coefficients, and the residual components $\boldsymbol{\epsilon_i} = (\epsilon_{i1}, \ldots, \epsilon_{in_i})'$ is a n_i-dimensional vector for $j = 1, \cdots, n_i$. Let I_{ni} to be the n_i-dimensional identity matrix. We often assume that residuals are independent and normally distributed with mean zero, and covariance matrix $\sigma^2 I_{n_i}$ [5].

Stage 2

After the first step, we use a multivariate regression model to analyse the observed variability between the subjects, with respect to their subject-specific regression coefficients $\boldsymbol{\alpha}_i$ as

$$\boldsymbol{\alpha_i} = K_i \boldsymbol{\beta} + \mathbf{b_i}, \tag{2.2}$$

where K_i is a $(q \times p)$ matrix of known covariates, and the p-dimensional vector $\boldsymbol{\beta} = (\beta_1, \ldots, \beta_p)'$ is a the unknown regression parameters and the q-dimensional vector of correlated random effects is $\mathbf{b_i} = (b_{i1}, \ldots, b_{iq})$ [5].

In order to fit these models in equation (2.2), we first estimate the 1st stage (2.1) and obtain $\boldsymbol{\alpha}_i$ values. Then we come back to estimate $\boldsymbol{\beta}$ for (2.2).

2.5 The General Linear Mixed Effect Model

If we combine the models from the previous stages by replacing the $\boldsymbol{\alpha}_i$ in (2.1) by (2.2), we have [5]

$$\boldsymbol{Y_i} \;=\; Z_i(K_i\boldsymbol{\beta} + \boldsymbol{b_i}) + \boldsymbol{\epsilon_i}$$

$$\begin{aligned} &= (Z_i K_i)\boldsymbol{\beta} + Z_i \boldsymbol{b_i} + \boldsymbol{\epsilon_i} \\ &= X_i \boldsymbol{\beta} + Z_i \boldsymbol{b_i} + \boldsymbol{\epsilon_i}. \end{aligned} \tag{2.3}$$

We call this model (2.3) a *linear mixed effect model with fixed effects* [5]. Note that $\boldsymbol{\beta}$ relates to the population as a whole and is the same for all subjects while $\mathbf{b_i}$ is a specific to each subject and is assumed to be random (and hence, we often call this random effects). In this equation $X_i = Z_i K_i$ is the $(n_i \times p)$ matrix of known covariates.

2.5.1 Random Effects Covariance Structure

In general a linear mixed effects model is any model which satisfies [8]:

$$\begin{cases} \mathbf{Y_i} = X_i \boldsymbol{\beta} + Z_i \boldsymbol{b}_i + \boldsymbol{\epsilon}_i, \\ \boldsymbol{b}_i \sim N(\mathbf{0}, D), \\ \boldsymbol{\epsilon}_i \sim N(\mathbf{0}, \Sigma_i), \\ \boldsymbol{b}_1, \ldots, \boldsymbol{b}_N, \boldsymbol{\epsilon}_1, \ldots, \boldsymbol{\epsilon}_N \end{cases} \tag{2.4}$$

This equation reflects the assumptions on the covariance structure of random effect $\mathbf{b_i}$ and the measurement error $\boldsymbol{\epsilon}_i$. It assumes that random effects $\mathbf{b_i}$ are independent and normal distributed with $E(\boldsymbol{b}_i) = \mathbf{0}$ and covariance matrix $Cov(\boldsymbol{b_i}) = D$. D is a positive definite $(q \times q)$ covariance matrix with (i, j) element $d_{ij} = d_{ji}$. The error components $\epsilon_\mathbf{i}$ are also normal distributed with $E(\boldsymbol{\epsilon}_i) = \mathbf{0}$ and $Cov(\boldsymbol{\epsilon}_i) = \Sigma_i$ where Σ_i is a $(n_i \times n_i)$ covariance matrix depending only on i through its dimension, i.e. the set of unknown parameters in Σ_i will not depend upon i. In some cases the assumption on error parameters is limited to

$$\boldsymbol{\epsilon_i} \sim N(\mathbf{0}, \sigma^2 I_{n_i}), \tag{2.5}$$

with I_{n_i} $(n_i \times n_i)$ identity matrix. The resulting model is referred to as a *conditional independence model*. Therefore, it is assumed that the n_i observations of the i^{th} individual given $\mathbf{b_i}$ and $\boldsymbol{\beta}$ are independent from each other and have the same variance σ^2. In this case correlation is generated only by the random effects.

The general form of the model (2.3) can be written as

$$\boldsymbol{Y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{Z}\boldsymbol{b} + \boldsymbol{\epsilon} \tag{2.6}$$

where $\boldsymbol{Y} = (\boldsymbol{Y_1}, \ldots, \boldsymbol{Y_N})'$ and the design matrix of fix and random effects are:

$$\mathbf{X} = \begin{pmatrix} X_1 \\ \vdots \\ X_N \end{pmatrix} \qquad \mathbf{Z} = \begin{pmatrix} Z_1 & & \mathbf{0} \\ & \ddots & \\ \mathbf{0} & & Z_N \end{pmatrix}.$$

Random effects and residual components have normal distribution of the form

$$\begin{pmatrix} \boldsymbol{b} \\ \boldsymbol{\epsilon} \end{pmatrix} \sim N\left(\begin{pmatrix} \mathbf{0} \\ \mathbf{0} \end{pmatrix}, \begin{pmatrix} \mathbf{D} & \mathbf{0} \\ \mathbf{0} & \Sigma \end{pmatrix}\right). \tag{2.7}$$

The covariance matrix $\mathbf{D}$ is a diagonal matrix with dimension $(N * q \times N * q)$ and Σ is also a diagonal matrix of $(N \times N)$-dimension with the form

$$\mathbf{D} = \text{diag}(D, \ldots, D, \ldots, D) \qquad \Sigma = \text{diag}(\Sigma_1, \ldots, \Sigma_i, \ldots, \Sigma_N).$$

Inference for linear mixed models is usually based on maximum likelihood or restricted maximum likelihood estimation under the marginal model for $\mathbf{Y_i}$, i.e., the multivariate normal model with mean $X_i\boldsymbol{\beta}$, and covariance $V_i = Z_i D Z_i' + \Sigma_i$ [5, 8]. Thus, the two different views on the linear mixed model, the conditional and marginal model can be combined. The conditional model is specified by

$$\boldsymbol{Y_i}|\boldsymbol{b}_i \sim N(X_i\boldsymbol{\beta} + Z_i\boldsymbol{b}_i, \Sigma_i).$$

It follows from (2.4) that, conditional on the random effect $\mathbf{b_i}$, $\mathbf{Y_i}$ is normally distributed with mean vector $X_i\boldsymbol{\beta} + Z_i\mathbf{b_i}$ and with covariance matrix Σ_i,

$$E(\boldsymbol{Y_i}|\boldsymbol{b_i}) = X_i\boldsymbol{\beta} + Z_i\boldsymbol{b_i}, \qquad Cov(\boldsymbol{Y_i}|\boldsymbol{b_i}) = Cov(\boldsymbol{\epsilon_i}) = \Sigma_i.$$

Further, $\mathbf{b_i}$ is assumed to be normally distributed with mean vector $\mathbf{0}$ and covariance matrix D.

Let $f(\mathbf{Y_i}|\mathbf{b_i})$ and $f(\mathbf{b_i})$ be the corresponding density functions. The marginal density function of $\mathbf{Y_i}$ is then given by

$$f(\mathbf{Y_i}) = \int \mathbf{f}(\mathbf{Y_i}|\mathbf{b_i})\mathbf{f}(\mathbf{b_i})\mathbf{db_i}, \tag{2.8}$$

which can easily be shown to be the density function of an n_i-dimensional normal distribution with mean vector $X_i\boldsymbol{\beta}$ and with covariance matrix $V_i = Z_i D Z_i' + \Sigma_i$.

$$\boldsymbol{Y_i}|\boldsymbol{b}_i \sim N(X_i\boldsymbol{\beta}, Z_i\mathbf{D}Z_i' + \Sigma_i),$$

The marginal mean includes the information about the whole population and not about each individual. However from the marginal covariance of $\mathbf{Y_i}$ it can be seen that the inclusion of random effects induced additional correlation in the model.

2.6 Estimation and Statistical Inference

Different estimation methods for the parameters in model (2.6) have been proposed over the years [9], but the most commonly used methods today are maximum likelihood (ML)

and restricted maximum likelihood (REML) [10].

The marginal model is used in a likelihood based approach. The estimation of the fixed effect parameter $\boldsymbol{\beta}$ can be performed independently from random effects $\boldsymbol{b_i}$. Inferences based on the marginal model do not explicitly assume the presence of random effects representing the natural heterogeneity between subjects. Although usually the most interest is in estimating the parameters of the marginal model (fixed effects $\boldsymbol{\beta}$ and the variance components in D and in all Σ_i), it is also useful to calculate the estimates for the random effects $\boldsymbol{b_i}$, since they reflect how much the subject-specific profile deviates from the overall average profile. These estimations are needed whenever interest is in prediction of subject specific evolutions. This section will cover the inference for parameters in the marginal distribution and in the next sections the estimation method for hierarchical models will be investigated.

As discussed previously, the general linear mixed model (2.4) implies the marginal model

$$\mathbf{Y_i} \sim N(X_i\boldsymbol{\beta}, Z_iDZ_i' + \Sigma_i). \tag{2.9}$$

Let us define $\boldsymbol{\alpha}$ to be the vector of all variance and covariance parameters in $Z_iDZ_i' + \Sigma_i$, it means $\boldsymbol{\alpha}$ consists of the $q(q+1)/2$ different elements in D and all parameters in Σ_i. Then if $\boldsymbol{\theta} = (\boldsymbol{\beta}', \boldsymbol{\alpha}')'$ denotes the s-dimensional vector of all parameters in the marginal model for $\mathbf{Y_i}$, and let $\boldsymbol{\Theta} = \Theta_\beta \times \Theta_\alpha$ denote the parameter space for $\boldsymbol{\theta}$, with Θ_β and Θ_α the parameter spaces for the fixed effects and for the variance components respectively. Note $\Theta_\beta = \mathbb{R}^p$, and Θ_α equals the set of values for $\boldsymbol{\alpha}$ such that D and all Σ_i are positive (semi-)definite [5]. The maximum likelihood method can be used to estimate the fixed effect, and the restricted maximum likelihood method provides an estimate of the covariance parameters in the model.

2.6.1 Fixed Effects Estimation with Known Covariance Parameter

The classical approach of inference is based on estimators obtained from maximizing the marginal likelihood function with respect to $\boldsymbol{\theta}$. Considering $E(\boldsymbol{Y}_i) = X_i\boldsymbol{\beta}$ and $Cov(\boldsymbol{Y}_i) = V_i = Z_iDZ_i' + \Sigma_i$, one can write

$$\begin{aligned} L_{ML}(\boldsymbol{\theta}) &= \prod_{i=1}^{N} \Big\{ (2\pi)^{-n_i/2} |V_i(\boldsymbol{\alpha})|^{1/2} \exp\big(-\frac{1}{2}(\mathbf{Y_i} - X_i\boldsymbol{\beta})' V_i^{-1}(\boldsymbol{\alpha})(\mathbf{Y_i} - X_i\boldsymbol{\beta})\big) \Big\} \\ &= (2\pi)^{-\frac{N}{2}} |V(\boldsymbol{\alpha})|^{-\frac{1}{2}} \exp\{-\frac{1}{2}(\boldsymbol{y} - X\boldsymbol{\beta})' V^{-1}(\boldsymbol{\alpha})(\boldsymbol{y} - X\boldsymbol{\beta})\}. \end{aligned} \tag{2.10}$$

Getting the logarithm of the likelihood function, we have the log-likelihood

$$l_{ML}(\theta) = -\frac{1}{2}\log(|V(\boldsymbol{\alpha})|) - \frac{1}{2}(\boldsymbol{y} - X\boldsymbol{\beta})' V^{-1}(\boldsymbol{\alpha})(\boldsymbol{y} - X\boldsymbol{\beta}). \tag{2.11}$$

By taking the derivatives of $\boldsymbol{\beta}$ and setting them to zero, yields

$$\frac{\partial l_{ML}(\theta)}{\partial \boldsymbol{\beta}} = X^{'} V^{-1}(\boldsymbol{\alpha})(\boldsymbol{y} - X\boldsymbol{\beta}) = 0,$$

which results to

$$X^{'} V^{-1}(\boldsymbol{\alpha})\boldsymbol{y} = X^{'} V^{-1}(\boldsymbol{\alpha}) X\boldsymbol{\beta}.$$

Since in this case $\boldsymbol{\alpha}$ is known, the maximum likelihood estimator **MLE** of $\boldsymbol{\beta}$ can be obtained as

$$\hat{\boldsymbol{\beta}}(\boldsymbol{\alpha}) = \Big(\sum_{i=1}^{N} X_i' W_i X_i\Big)^{-1} \sum_{i=1}^{N} X_i' W_i y_i, \tag{2.12}$$

where W_i equals V_i^{-1} and $\hat{\boldsymbol{\beta}}(\boldsymbol{\alpha})$ is the best unbiased linear estimator of $\boldsymbol{\beta}$ which is dependent on covariance parameter vector $\boldsymbol{\alpha}$.

2.6.2 Fixed Effects Estimation with unknown Covariance Parameter

If $\boldsymbol{\alpha}$ is not known, but an estimate $\hat{\boldsymbol{\alpha}}$ is available, we can set $\hat{V}_i = V_i(\hat{\boldsymbol{\alpha}}) = \hat{W}_i^{-1}$, and estimate $\boldsymbol{\beta}$ by using expression (2.12) in which W_i is replaced by $\hat{W}_i$.

$$\hat{\boldsymbol{\beta}}(\hat{\boldsymbol{\alpha}}) = \Big(\sum_{i=1}^{N} X_i' \hat{W}_i X_i\Big)^{-1} \sum_{i=1}^{N} X_i' \hat{W}_i y_i. \tag{2.13}$$

This estimator is called the empirical best linear unbiased estimator **EBLUE** [11].

Two frequently used methods for estimating $\boldsymbol{\alpha}$ are maximum likelihood estimation and restricted maximum likelihood estimation, which will be discussed and compared later.

2.7 Restricted Maximum Likelihood Estimation

In 1971 Patterson and Thompson formally described the Restricted Maximum Likelihood method [12]. REML estimators maximize the likelihood of the parameters after correcting for the fixed effects. It is a ML method that accounts for the loss of degrees of freedom due to fitting fixed effects (which is not the case for traditional ML).

2.7.1 Variance Estimation in Normal Populations

As an introductory example to introduce the $REML$ method, we can go back to parameter estimation in regression models. A ML estimation for the variance of N observations from a normal distribution with mean μ and σ^2 is

$$\hat{\sigma}^2 = \sum_i (Y_i - \mu)^2 / N$$

If we assume that μ is known, then it is an unbiased estimate of σ^2. In case that μ is not known, we replace μ by the sample mean and get

$$E(\hat{\sigma}^2) = \frac{N-1}{N}\sigma^2. \tag{2.14}$$

Due to the estimation of μ, the MLE is now biased downward. An unbiased estimate is easily obtained from expression (2.14), yielding the classical sample variance

$$S^2 = \frac{1}{N-1}\sum_{i=1}^{N}(Y_i - \bar{Y})^2$$

One method to obtain directly an unbiased estimator for σ^2, is the restricted maximum likelihood estimation ($REML$ estimation)[12, 13].

Let $\mathbf{Y} = (Y_1, \cdots, Y_N)'$ denote the vector of all measurements, and let $\mathbf{1}_N$, be the N-dimensional vector containing only ones. The distribution of $\mathbf{Y}$ is then $N(\mu\mathbf{1}_N, \sigma^2 I_N)$ where, I_N denotes the N-dimensional identity matrix. Let A be any $N \times (N-1)$ matrix with $N-1$ linearly independent columns orthogonal to the vector $\mathbf{1_N}$. We then define the vector $\mathbf{U}$ of $N-1$ so-called error contrasts by $\mathbf{U} = A'\mathbf{Y}$ which now follows a normal distribution with mean vector $\mathbf{0}$ and covariance matrix $\sigma^2 A'A$. Maximizing the corresponding likelihood with respect to the only parameter σ^2 yields

$$\hat{\sigma}^2 = \mathbf{Y}'A(A'A)^{-1}A'\mathbf{Y}/(N-1), \tag{2.15}$$

which can be shown to be equal to the classical sample variance S^2 previously derived from expression (2.14). Note that any matrix A satisfying the specified conditions leads to the same estimator for σ^2. The resulting estimator for σ^2 is often called the restricted maximum likelihood (REML) estimator since it is restricted to $(N-1)$ error contrasts [14].

2.7.2 Estimation of Residual Variance in Linear Regression

The MLE for the residual variance σ^2 in a linear regression model is

$$\hat{\sigma}^2 = (\mathbf{Y} - X(X'X)^{-1}X'\mathbf{Y})'(\mathbf{Y} - X(X'X)^{-1}X'\mathbf{Y})/N, \tag{2.16}$$

which can be easily shown to be biased downward by a factor $(N-p)/N$. Similarly, σ^2 can be estimated using a set of error contrasts $\mathbf{U} = A'\mathbf{Y}$ where A is now any $N \times (N-p)$ matrix with $N-p$ linearly independent columns orthogonal to the columns of the design matrix X. We then have that $\mathbf{U} \sim N(\mathbf{0}, \sigma^2 A'A)$, in which σ^2 is the only unknown parameter. Maximizing the corresponding likelihood with respect to σ^2 yields

$$\hat{\sigma}^2 = (\mathbf{Y} - X(X'X)^{-1}X'\mathbf{Y})'(\mathbf{Y} - X(X'X)^{-1}X'\mathbf{Y})/(N-p), \tag{2.17}$$

which is the mean squared error, unbiased for σ^2, and classically used as estimator for the residual variance in linear regression analysis.

2.7.3 REML Estimation for the Linear Mixed Model

As already been introduced the REML estimator is a maximum likelihood estimator based on a linearly transformed set of data $\mathbf{Y}^* = A\mathbf{Y}$ such that distribution of Y does not depend on $\boldsymbol{\beta}$. One way to achieve this is, to take A as a matrix which converts $\mathbf{Y}$ to a ordinary least squares (OLS) residuals,

$$A = I - X(X^{'}X)^{-1}X^{'} \tag{2.18}$$

with $\mathbf{Y}^*$ having a normal distribution with $E(\mathbf{Y}^*) = 0$ and independent from $\hat{\boldsymbol{\beta}}$. Since A is not a full rank matrix, Y has a singular distribution. In order to get a non-singular distribution one can choose $N-p$ rows of matrix A. It turns out that the resulting estimators for σ^2 and $V(\boldsymbol{\alpha})$ do not depend on which rows we use. Let A be the matrix defined in (2.18) and B a $N-p$ matrix, so that $BB' = A$ and $B'B = I$ where I indicates an Identity matrix with dimension $(N-p)\times(N-p)$ [15, 14].

Let $\mathbf{U} = B'\mathbf{Y}$ and $\boldsymbol{\alpha}$ to be fixed. The density function of $\mathbf{Y}$ is,

$$f(\boldsymbol{y}) = (2\pi)^{-\frac{N}{2}}|V|^{-\frac{1}{2}}\exp\left(-\frac{1}{2}(\boldsymbol{Y}-X\boldsymbol{\beta})^{'}V^{-1}(\boldsymbol{Y}-X\boldsymbol{\beta})\right). \tag{2.19}$$

The MLE estimator of $\boldsymbol{\beta}$ is the generalized least squares estimator,

$$\hat{\boldsymbol{\beta}} = (X^{'}V^{-1}X)^{-1}X^{'}V^{-1}\boldsymbol{y} = G\boldsymbol{y}.$$

and

$$f(\hat{\boldsymbol{\beta}}) = (2\pi)^{-\frac{p}{2}}|X^{'}V^{-1}X|^{\frac{1}{2}}\exp\left(-\frac{1}{2}(\hat{\boldsymbol{\beta}}-\boldsymbol{\beta})^{'}(X^{'}V^{-1}X)(\hat{\boldsymbol{\beta}}-\boldsymbol{\beta})\right). \tag{2.20}$$

Now we can show for $\mathbf{U} = B'\mathbf{Y}$, $E(\mathbf{U}) = \mathbf{0}$ and $\mathbf{U}$ and $\hat{\boldsymbol{\beta}}$ are independent

$$E(\boldsymbol{U}) = B^{'}E(\boldsymbol{Y}) = B^{'}X\boldsymbol{\beta} = B^{'}BB^{'}X\boldsymbol{\beta} = B^{'}AX\boldsymbol{\beta},$$

since $B^{'}B = I_{N-p}$. But we already know $BB^{'} = A$, therefore,

$$AX = (I_{N-p} - X(X^{'}X)^{-1}X^{'})X = X - X(X^{'}X)^{-1}X^{'}X = \mathbf{0}$$

hence $E(\boldsymbol{U}) = \mathbf{0}$. In order to prove that $\mathbf{U}$ and $\hat{\boldsymbol{\beta}}$ are independent, we prove that $Cov(\boldsymbol{U}, \hat{\boldsymbol{\beta}}) = \mathbf{0}$. We can write

$$\begin{aligned} Cov(\boldsymbol{U}, \hat{\boldsymbol{\beta}}) &= E(\boldsymbol{U}(\hat{\boldsymbol{\beta}}^{'} - \boldsymbol{\beta}^{'})) \\ &= E(B^{'}\boldsymbol{Y}(\boldsymbol{Y}^{'}G^{'} - \boldsymbol{\beta}^{'})) \end{aligned}$$

$$\begin{aligned} &= B^{'}E(\boldsymbol{Y}\boldsymbol{Y}^{'})G^{'} - B^{'}E(\boldsymbol{Y})\boldsymbol{\beta}^{'} \\ &= B^{'}\left[Var(\boldsymbol{Y}) + E(\boldsymbol{Y})E(\boldsymbol{Y})^{'}\right]G^{'} - B^{'}E(\boldsymbol{Y})\boldsymbol{\beta}^{'} \\ &= B^{'}\left[V + X\boldsymbol{\beta}\boldsymbol{\beta}^{'}X^{'}\right]G^{'} - B^{'}X\boldsymbol{\beta}\boldsymbol{\beta}^{'} \\ &= B^{'}VG^{'} + B^{'}X\boldsymbol{\beta}\boldsymbol{\beta}^{'}X^{'}G^{'} - B^{'}X\boldsymbol{\beta}\boldsymbol{\beta}^{'} \end{aligned}$$

and $B^{'}VG^{'}$ is,

$$\begin{aligned} B^{'}VG^{'} &= B^{'}VV^{-1}X(X^{'}V^{-1}X)^{-1} \\ &= B^{'}X(X^{'}V^{-1}X)^{-1} \\ &= B^{'}BB^{'}X(X^{'}V^{-1}X)^{-1} \\ &= B^{'}AX(X^{'}V^{-1}X)^{-1} = \mathbf{0} \end{aligned}$$

because $B^{'}X = B^{'}AX = \mathbf{0}$ as in proof for $E(\mathbf{U}) = \mathbf{0}$ and also

$$X^{'}G^{'} = X^{'}[(X^{'}V^{-1}X)^{-1}X^{'}V^{-1}]^{'} = X^{'}V^{-1}X(X^{'}V^{-1}X)^{-1} = I_{N-p}.$$

Therefore now we could show that $Cov(\boldsymbol{U}, \hat{\boldsymbol{\beta}}) = \mathbf{0}$, hence $\mathbf{U}$ and $\hat{\boldsymbol{\beta}}$ are independent.

Further, the probability density function of $\mathbf{U}$, expressed in terms of $\mathbf{Y}$, is proportional to the following ratio,

$$f(\boldsymbol{u}) = |X^{'}X|^{\frac{1}{2}}\frac{f(\boldsymbol{y})}{f(\hat{\boldsymbol{\beta}})},$$

where $\frac{f(\mathbf{y})}{f(\hat{\boldsymbol{\beta}})}$ can be obtained by substituting (2.19) and (2.20) [15]. The proportion will be,

$$\frac{f(\boldsymbol{y})}{f(\hat{\boldsymbol{\beta}})} = (2\pi)^{-\frac{N-p}{2}}|V|^{-\frac{1}{2}}|X^{'}V^{-1}X|^{-\frac{1}{2}}\exp\left(-\frac{1}{2}(\boldsymbol{Y} - X\hat{\boldsymbol{\beta}})^{'}V^{-1}(\boldsymbol{Y} - X\hat{\boldsymbol{\beta}})\right).$$

To obtain the explicit form of this ratio, we can use the following result for the ***GLM***,

$$(\boldsymbol{y} - X\boldsymbol{\beta})^{'}V^{-1}(\boldsymbol{y} - X\boldsymbol{\beta}) = (\boldsymbol{y} - X\hat{\boldsymbol{\beta}})^{'}V^{-1}(\boldsymbol{y} - X\hat{\boldsymbol{\beta}}) + (\hat{\boldsymbol{\beta}} - \boldsymbol{\beta})^{'}(X^{'}V^{-1}X)(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}).$$

Therefore the probability density function of $\mathbf{U}$ is,

$$L(\boldsymbol{\alpha}) = f(\boldsymbol{y}) = (2\pi)^{-\frac{N-p}{2}}|X^{'}X|^{\frac{1}{2}}|V|^{-\frac{1}{2}}|X^{'}V^{-1}X|^{-\frac{1}{2}}\exp\left(-\frac{1}{2}(\boldsymbol{y} - X\hat{\boldsymbol{\beta}})^{'}V^{-1}(\boldsymbol{y} - X\hat{\boldsymbol{\beta}})\right).$$

Finally, we can show that the likelihood function equals

$$L(\boldsymbol{\alpha}) = C\Big|\sum_{i=1}^{N} X_i^{\prime}V^{-1}(\boldsymbol{\alpha})X_i\Big|^{\frac{1}{2}}L_{ML}(\hat{\boldsymbol{\beta}}, \boldsymbol{\alpha}), \tag{2.21}$$

where C is a constant not depending on $\boldsymbol{\alpha}$, and where $L_{ML}(\boldsymbol{\beta}, \boldsymbol{\alpha}) = L_{ML}(\boldsymbol{\theta})$ is the maxi-

mum likelihood function given by (2.10).

Because $|\sum_{i=1}^{N} X_i' V^{-1}(\boldsymbol{\alpha}) X_i|$ in (2.21) does not depend on $\boldsymbol{\beta}$, it follows that the REML estimators for $\boldsymbol{\alpha}$ and for $\boldsymbol{\beta}$ can also be found by maximizing the REML likelihood function

$$L_{REML}(\boldsymbol{\theta}) = \Big|\sum_{i=1}^{N} X_i' V^{-1}(\boldsymbol{\alpha}) X_i\Big|^{\frac{1}{2}} L_{ML}(\boldsymbol{\theta}), \tag{2.22}$$

with respect to all parameters simultaneously ($\boldsymbol{\alpha}$ and $\boldsymbol{\beta}$). The REML estimator, $\hat{\boldsymbol{\alpha}}$, maximizes the log likelihood function,

$$l_{REML}(\boldsymbol{\alpha}) = -\frac{1}{2}\log|V(\boldsymbol{\alpha})| - \frac{1}{2}\log|X' V^{-1}(\boldsymbol{\alpha}) X| - \frac{1}{2}(\boldsymbol{y} - X\hat{\boldsymbol{\beta}})' V(\boldsymbol{\alpha})^{-1}(\boldsymbol{y} - X\hat{\boldsymbol{\beta}}). \tag{2.23}$$

The ML estimator (2.11) simply differs from the REML estimator (2.23) by the term $-\frac{1}{2}\log|X' V^{-1}(\boldsymbol{\alpha}) X|$, which is independent from $\boldsymbol{\beta}$. Therefore the REML estimator can be written as

$$l_{REML}(\boldsymbol{\beta}, \boldsymbol{\alpha}) = l_{ML}(\boldsymbol{\beta}, \boldsymbol{\alpha}) - \frac{1}{2}\log|X' V^{-1}(\boldsymbol{\alpha}) X|.$$

2.7.4 Justification of REML Estimation

Patterson and Thomson (1971) proved that when inference is based on $\mathbf{U}$ instead of $\mathbf{Y}$, no information about $\boldsymbol{\alpha}$ is lost in the absence of information on $\boldsymbol{\beta}$. More precisely, $\mathbf{U}$ is marginally sufficient for $\boldsymbol{\alpha}$ [16, 17]. Further, it has been shown that, from a Bayesian point of view, using only error contrasts to make inferences on $\boldsymbol{\alpha}$ is equivalent to ignoring any prior information on $\boldsymbol{\beta}$ and using all the data to make those inferences [17].

Several methods have been introduced to calculate ML and REML estimates by applying optimization techniques like $\boldsymbol{EM}$, $\boldsymbol{Newton - Raphson}$ and $\boldsymbol{Fisher - scoring}$ algorithms. In 1977 the $ExpectationMinimization$ (EM) have been introduced by Laird and Rubin. The EM algorithm is only used to estimate the random effects $\mathbf{b}_i$. The main advantage of the EM algorithm is that the general theory assures that each iteration increases the likelihood [18] but Laird and Ware (1982) report slow convergence of the estimators of the variance components, especially when the maximum likelihood is on or near the boundary of the parameter space. Usually it is better to use Newton-Raphson-based procedures to estimate all parameters in the model [19].

2.8 Inference for the Marginal Model

2.8.1 Inference for the Fixed Effects

As discussed in previous sections, the vector of fixed effects $\boldsymbol{\beta}$ is estimated by

$$\hat{\boldsymbol{\beta}}(\boldsymbol{\alpha}) = \Big(\sum_{i=1}^{N} X_i' V_i^{-1} X_i\Big)^{-1} \sum_{i=1}^{N} X_i' V_i^{-1} y_i = (X' V^{-1}(\boldsymbol{\alpha}) X)^{-1} X' V^{-1}(\boldsymbol{\alpha}) \boldsymbol{y}, \tag{2.24}$$

in which the unknown vector $\boldsymbol{\alpha}$ of variance components is replaced by its ML or REML estimate. $\hat{\boldsymbol{\beta}}(\boldsymbol{\alpha})$ has a multivariate normal distribution with

$$\begin{aligned} E(\hat{\boldsymbol{\beta}}(\boldsymbol{\alpha})) &= (X^{'}V^{-1}(\boldsymbol{\alpha})X)^{-1}X^{'}V^{-1}(\boldsymbol{\alpha})E(\boldsymbol{y}) \\ &= (X^{'}V^{-1}(\boldsymbol{\alpha})X)^{-1}X^{'}V^{-1}(\boldsymbol{\alpha})X\boldsymbol{\beta} = \boldsymbol{\beta}, \end{aligned}$$

, and because of

$$Var(\mathbf{y}) = V(\boldsymbol{\alpha}) \tag{2.25}$$

$$\begin{aligned} Var(\hat{\boldsymbol{\beta}}(\boldsymbol{\alpha})) &= (X^{'}V^{-1}(\boldsymbol{\alpha})X)^{-1}X^{'}V^{-1}(\boldsymbol{\alpha})Var(\boldsymbol{y})V^{'-1}(\boldsymbol{\alpha})X(X^{'}V^{-1}(\boldsymbol{\alpha})X)^{'-1} \\ &= (X^{'}V^{-1}(\boldsymbol{\alpha})X)^{-1}. \end{aligned}$$

Chapter 3

Data Management

3.1 Data collection

741 healthy subjects (49% boys) were recruited in various studies using Pulmonary Function Tests (*PFTs*) from early months after birth to school age within different projects. In total 1188 observations had been collected between the years 1998 and 2010. The *PFTs* were performed in infants using either RASP (home made) or JAEGER (commercial) devices and for school age children JAEGER. The age range for girls was 7.8 weeks to 11.97 years and in boys it was 3.12 weeks to 11.67 years. Corresponding values for height were 54 to 158 cm in girls and 59 to 151 cm in boys.

3.1.1 Infants

Healthy infants born in the maternity units at the Homerton and University College Hospitals, London were recruited up to their second birthday. These healthy controls were excluded if they had gestational age (GA)[1] less than 37 weeks or birth weight less than 2.5 kg. It has been shown that infants born small for gestational age (SGA) are at higher risk of wheezing and respiratory disease in early childhood. Furthermore, infants with low birth weight (less than 2.5 kg) may have reduced airway function in their adulthood [20],[21]. These healthy children were ineligible also if they had a hospitalization history for respiratory illnesses or had wheeze or any congenital abnormalities before recruitment. The age range of these recruited infants is 3.12 weeks to 1.98 years.

3.1.2 Pre school age

Data were collected from 3 to 5 years old children in various research projects. Families of children recruited above were also invited to return for follow up, when the child was in its pre school age. These healthy controls were also excluded if they had been hospitalized for respiratory disease, had doctor diagnosed asthma or were currently consuming any

[1] Gestational age is the common term used during pregnancy to describe how far along the pregnancy is. It is measured in weeks, from the first day of the mother's last menstrual cycle to the date of delivery. A normal pregnancy can range from 38 to 42 weeks of GA. Infants born before 37 weeks of GA are considered as preterm. Infants born after 42 weeks of GA are considered postmature.

asthma medication or other respiratory illnesses. On the day of test, parents were asked about the family history, including whether the mother currently smoked or smoked during pregnancy. Any signs of wheeze or cough in the preceding week were also ascertained. The minimum age of PSA (pre school age) children recruited was 2.73 years and maximum age 5.95 years.

3.1.3 School age

Children from 6 to 12 years were recruited in different research studies and children recruited in infants and preschool age study projects mentioned in sections 3.1.1 and 3.1.2 were invited to return for the follow up as well, when the child was in its school age. The same selection criteria were applied to these children in terms of their health at time of test and parental smoking habit at time of test. They have been asked for signs of wheeze or coughed in the preceding week as well. The age range of SA (school age) children recruited in this study was 6.04 years to 11.53 years.

3.2 Data selection

Despite recent advances and availability of some spirometric reference data from healthy black and Hispanic children, there remains a paucity of suitable equations for ethnic groups other than those of white European descent, especially among young children. As demonstrated recently, previous attempts to correct for ethnic differences have been over simplistic [22][23], therefore all children from white mothers are presented in this study. In our study population overall 489 from 741 children were from white mother descent. Applying birth weight and gestational age health criteria briefly described in previous sections on these children, the study population ended up with 416 healthy control subjects with total of 749 number of observations. Initially, scatter plots of each of PFT outcomes were visually inspected for outliers and data transcription errors. Any unlikely values were coded as missing and exclusion of an outlier or implausible value was based on agreement with Dr. Sooky Lum. There were 5 observations with extreme $FEFV$ values which have been considered as outliers and removed from the data set.

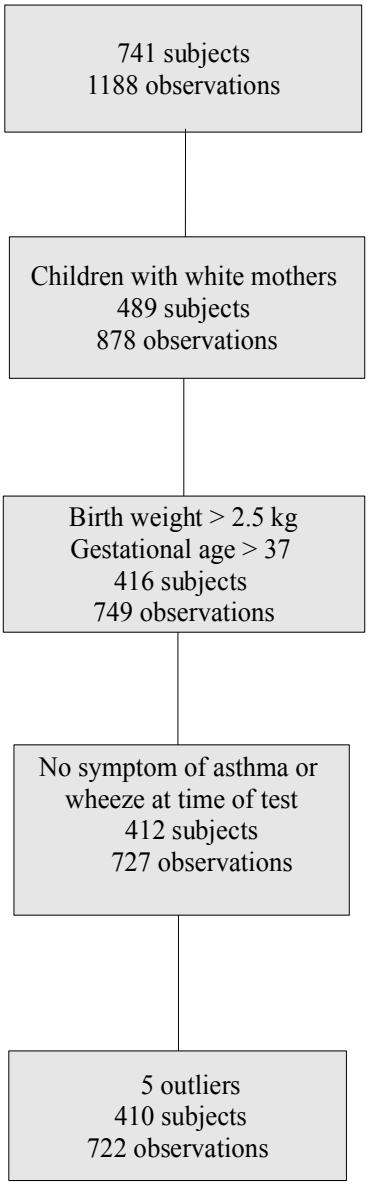


Figure 3.1: Flow chart of data cleaning procedure.

Table 3.1 presents the frequency of repeated measurements within subjects. The first column represents the number of repeated measurements, the second column is the total number of measurements obtained per number of tests (e.g. there were 268 measurements from subjects who attended two times). The last column stands for the fraction of boys in study population in terms of number of tests (e.g. 48.50 % of subjects who attended two times were boys). As expected sex was uniformly distributed during the study except for the few cases with 5 or 6 repeated measurements.

Number of tests	Number of observations	(%) Boys
1	208	52.4
2	268	48.5
3	123	41.4
4	60	53.3
5	45	11.1
6	18	100.0
Total	722	

Table 3.1: Repeated measurements of individuals in data set with all the measurement instruments present.

Summary of population characteristics at time of first test occasion is given in Table 3.2.

Healthy children	*n*
Total number	410
%Boys	49%
Test age	1.68y(1(m) - 11.97(y))
Gestational age(w)	39.78(37.00 - 43.00)
Birth weight (kg)*	3.35(0.43)
SDS test height*	0.29(1.10)
SDS test weight*	0.02(1.00)
% Maternal smoke in pregnancy	24%

Table 3.2: Details of children at time of first test occasion. In order to make the age range easier to understand, minimum age is presented in months and maximum age in years. Information of weight and height presented as *mean(sd).

3.3 Reference equations and z-score calculations

3.3.1 All age reference equations

A reliable interpretation of pulmonary function test results requires the availability and use of appropriate reference equations data to help to identify between health and disease. The "all age spirometry" study investigated different methods to develop more appropriate reference ranges which could describe the relationship between lung function, height and age more accurately during childhood. These equations can also be applied on adults and the transition between childhood and adulthood stages [24]. These equations provide smoothly changing reference curves during periods of rapid growth and transition to produce a single reference across a wide age range (5–80yrs) in people of European descent [23].

Improvements in spirometry measurements techniques have made it possible to also collect measurements in children as young as 3 years old. Therefore new available prediction equations are extended across the preschool years and are also joined to established reference equations for older children and adults. Since year 2008, a large collection of preschool data have been added to this to extend the all age models from 3 to 8 years without changing the equations in children aged more than 10 years [25]. These equations describe a multiplicative and allometric relationship, where FEV1, forced vital capacity (FVC) and forced expiratory flow at 25–75% of FVC (FEF25–75) are proportional to height raised to the power 2.5 [24] [23]. In addition to extending the outcomes already reported, reference equations were developed for FEV0.75 for preschool and school age range. The equations are also available as an Excel add-in and can be downloaded from `http://www.lungfunction.org/index.php`.

3.3.2 Infants reference equations

As briefly discussed in the previous sections a reliable interpretation of pulmonary function tests relies on appropriate reference data, which remain very limited for infants. In the data collection process for this research study, the Jaeger Master screen Baby Body (v4.67) equipment and an "in house" instrument was used to perform partial and raised volume $FEFV$ manoeuvres in healthy infants.

The need to assess whether selected prediction equations are appropriate for a given population or specific equipment is well recognized but has rarely been done in infants due to the time consuming nature of these tests and need for sedation [25].
Measurements obtained from Jaeger Master screen were showing significantly lower value than published reference data, which could lead to over diagnosis of lung disease in children with Cystic Fibrosis or other respiratory diseases. Therefore applying the published reference data for infant's $FEFV$ manoeuvres are not suitable for measurements collected by Jaeger Master screen equipment [26]. Application of an appropriate adjustment factor may minimize such errors until sufficient multi center data are available to construct reliable

equipment-specific reference ranges in this age group [26].

When using original "in-house" equipment and software (RASP), published reference data for both partial and full $FEFV$ manoeuvres appeared to be appropriate for use in ICH laboratory, as demonstrated by the mean (sd) z-scores for $FEFV$ outcomes which approximated by 0 and 1 in their local healthy controls [27] [26]. Furthermore data collected from ICH using the homemade RASP system were in close agreement with that from other centers and were included in the collated dataset for V'maxFRC prediction equations [27][26]. Given that the mean V'maxFRC z-score for data collected in ICH laboratory using the RASP system higher than that from our recent Jaeger data, it is likely that the observed differences reflect differences in hardware and software rather than population differences or changes in practice. However later, direct comparison of a limited number of infants, using an identical jacket, suggested that flows were lower when using the Jaeger system [28], and as further healthy controls were studied using this equipment they became increasingly aware of a potential bias [26]. Equipment-specific differences in lung function have been also reported previously in older subjects, [29][30], and it has been suggested that such discrepancies may be due to device-dependent characteristics such as in the integration of flow to volume, which, together with body temperature, pressure, and saturated corrections, may be inaccessible to the end-user. In contrast to adult spirometers, there are currently no accepted wave forms with which to compare outcome measures for infant $FEFV$ equipment. Therefore we can not be sure whether the previous "in-house" systems or new commercially available devices best approximate the "truth". Consequently before driving an equipment specific adjustment factor, we have chosen to use available prediction equations adjusted for sex, age and height until enough multi center data collated. Also we have two options to face the home made system (RASP) measurements discrepancies: 1. To exclude all the data measured by home made instrument from the study. 2. Until further multi-center data can be collated, we can choose to use available prediction equations.

3.3.3 Population Characteristics

Table 3.3 expresses our data as z-scores applying Jones reference equations (2009) for 3 to 80 years and Jones adjusted equations to height for infants. It shows the mean and standard error (in brackets) for selected function outcomes. As expected the mean of FEF_{75} and FEF_{25-75} is close to zero, and their standard error is approximately 1. However, we observe a high discrepancy from zero for mean of $FEV_{0.5}$ z-score. This is because of the lack of appropriate reference equations for "in house" instrument and RASP software which was discussed comprehensively in the previous section.

If we exclude the data measured by "in house" instrument which covers only the infants, we can see the mean of $FEV_{0.5}$ z-score of the population is now closer to zero and sd near to 1. Table 3.4 represents the study population characteristics at time when values measured by RASP instrument is excluded.

Controls	n	%Boys	%Girls
Observations	722		
Subjects	410	49	51
Current smoking exposure	171	41	59
Test Age †	4.17(1(m)-11.97(y))		
SDS weight *	0.059(1.05)		
SDS height *	0.39(1.10)		
$FEV_{0.5}$ z-scores *	-0.40(1.36)		
FEF_{75} z-scores *	0.18(1.23)		
FEV_1 z-scores *	0.15(0.95)		
FEF_{25-75} z-scores *	0.03(1.29)		
FVC z-scores *	-0.07(1.15)		

Table 3.3: Population characteristics at time of test, when all observations measured by Jaeger and Rasp instruments are both present in data set. In order to make the age range easier to understand, minimum age is presented in months and maximum age in years. All variables in this table are presented as *mean(sd).

Controls	n	%Boys	%Girls
Observations	366		
Subjects	227	45	55
Current smoking exposure	127	35	65
Test Age, y †	4.17(1.6(m)-11.97(y))		
SDS weight*	0.26(1.10)		
SDS height*	0.45(1.15)		
$FEV_{0.5}$ z-scores *	-0.04(0.99)		
FEF_{75} z-scores *	-0.06(0.99)		
FVC z-scores *	0.19(1.07)		
FEF_{25-75} z-scores *	-0.45(0.94)		
FEV_1 z-scores *	0.15(0.95)		

Table 3.4: Population characteristics at time of test covering data measured by Jaeger instrument. † Median(range),* Mean(sd)

3.4 Exploratory data analysis

First and before starting any statistical analysis, we conduct descriptive exploratory analyses of our data. Exploratory analyses of longitudinal data can reveal general patterns, provide insight into functional form, and identify individuals whose data do not conform to the general pattern.

3.4.1 Visualization of raw $FEFV$ outcomes

Figures 3.2 to 3.6 present the relationship between $FEFV$ outcomes ($FEV_{0.5}$, $FEV_{0.75}$, FEV_1, FVC, FEF_{25-75}) and age at time of test. This figures visualize the data when all measurement instruments (*Jaeger* and *RASP*) are presented in the dataset, i.e. 722 observations. Clearly they indicate a positive linear relation between $FEFV$ variables with age. Gender can be identified by colours. Evidently there is a gap between 2 to 3 years old and another gap between 7 to 10 years old which shows lack of collected measurements in these age ranges. Please see Appendix for more visualization on mean of $FEFV$ outcomes against test height and test age breakdown by gender and different ethnicity.

3.4.1.1 Visualization of raw $FEFV$ outcomes against test age in years.

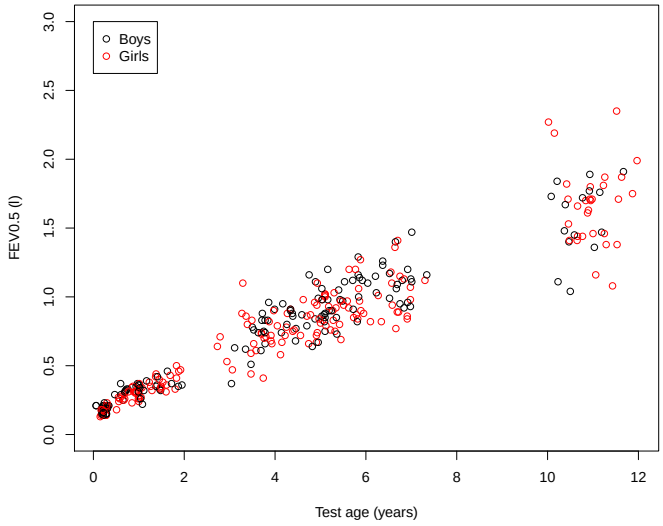


Figure 3.2: Scatter plots of raw $FEV_{0.5}$ against test age in years

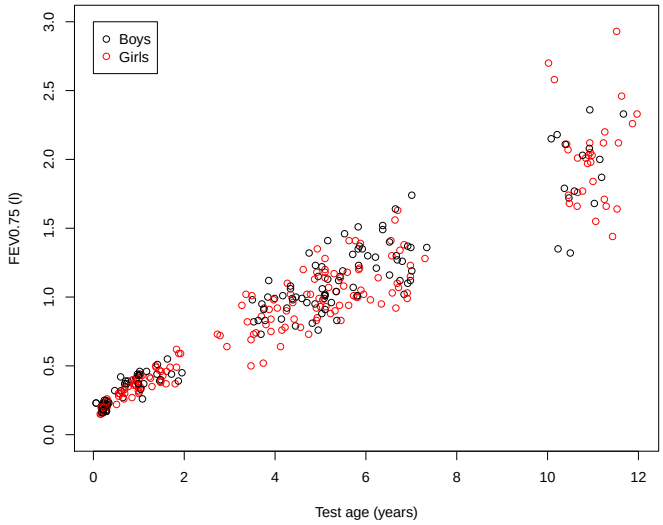


Figure 3.3: Scatter plots of raw $FEV_{0.75}$ against test age in years

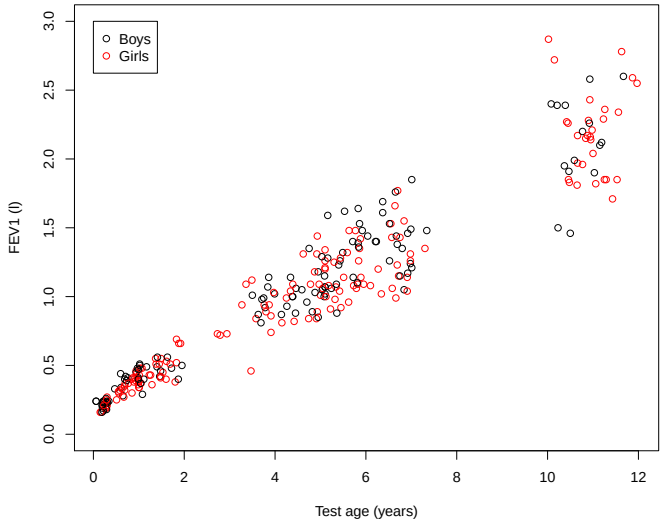


Figure 3.4: Scatter plots of raw FEV_1 against test age in years

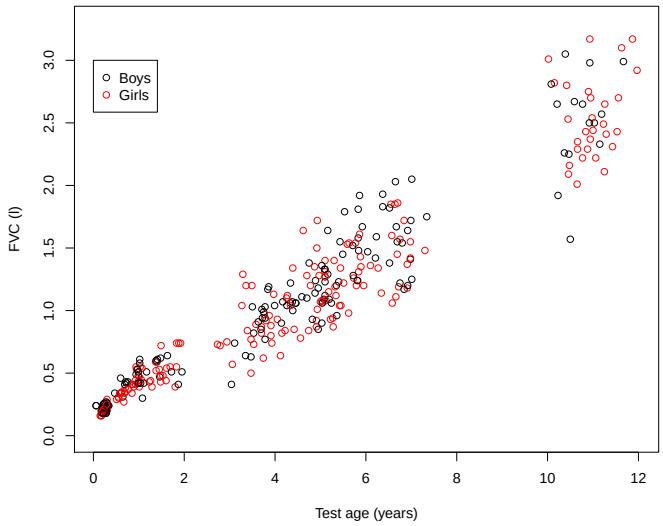


Figure 3.5: Scatter plots of raw FVC against test age in years

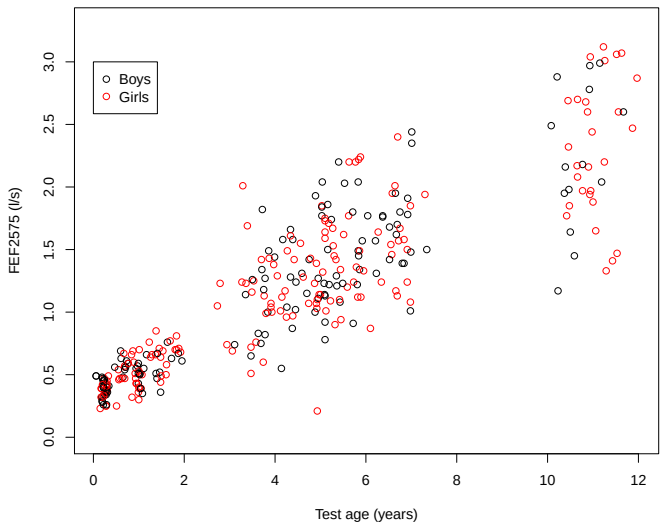
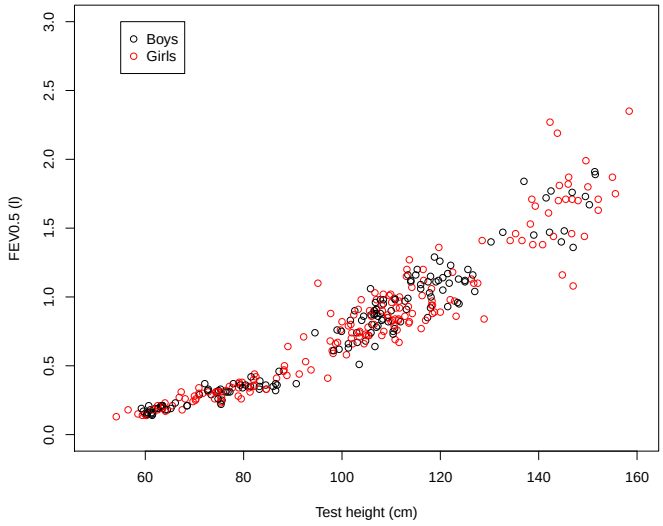


Figure 3.6: Scatter plots of raw $FEF25-75$ against test age in years

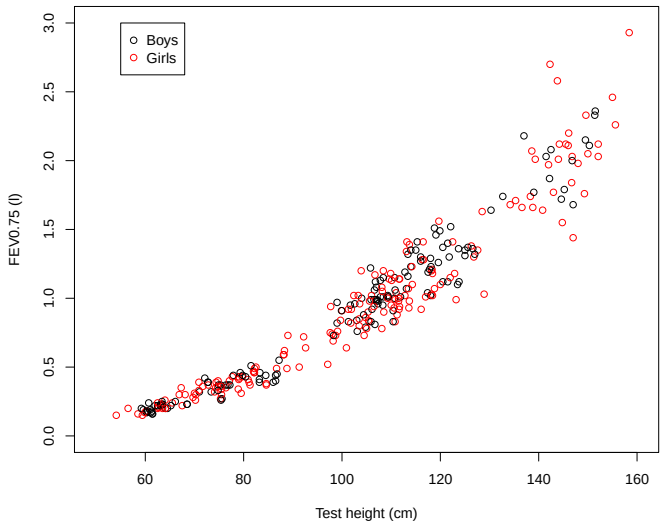
3.4.1.2 Visualization of raw $FEFV$ outcomes against test height in centimetres.

Figures 3.7 to 3.11 reflect the behaviour of $FEFV$ outcomes against test height. $FEFV$ outcomes show a curvilinear relationship with increasing height, which reflects changes in body proportions and shape during adolescence with lung growth lagging behind growth in standing height.



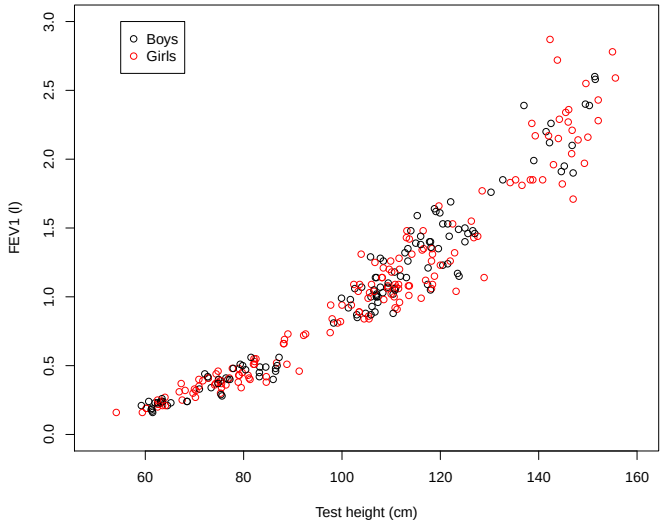
(a)

Figure 3.7: Scatter plots of raw $FEV_{0.5}$ against test height in centimetres



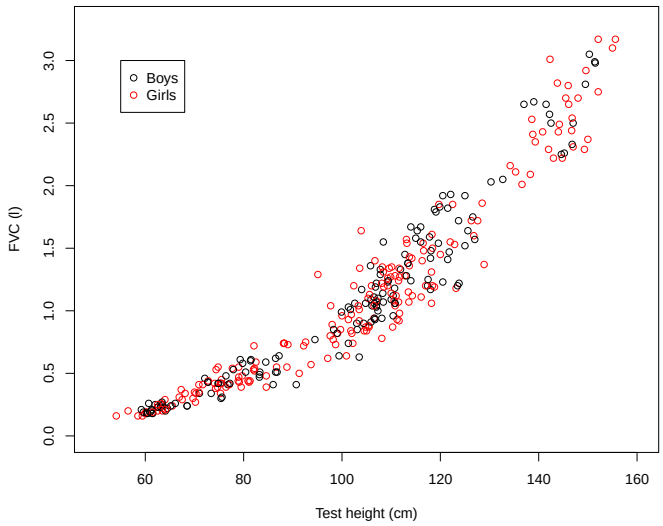
(a)

Figure 3.8: Scatter plots of raw $FEV_{0.75}$ against test height in centimetres



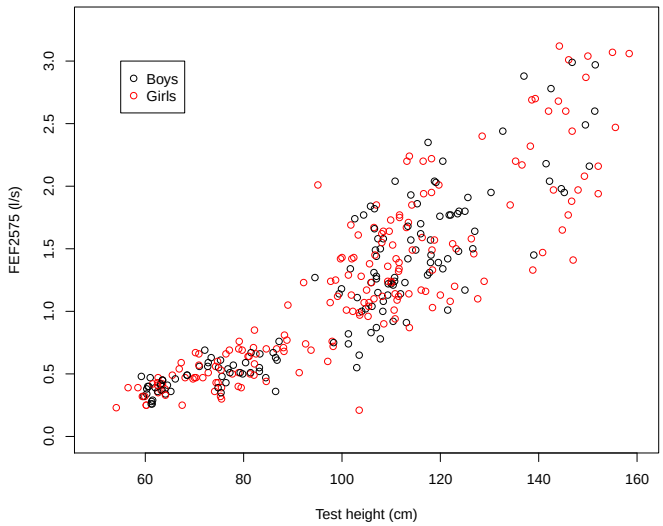
(a)

Figure 3.9: Scatter plots of raw FEV_1 against test height in centimetres



(a)

Figure 3.10: Scatter plots of raw FVC against test height in centimetres



(a)

Figure 3.11: Scatter plots of raw $FEF25-75$ against test height in centimetres

3.4.2 Visualization of $FEFV$ z-scores

Fitted z-scores were calculated for each individual in the dataset, applying Jones all age reference equations 2009 for 3 to 80 years and Jones adjusted equations to height for infants. Each of these equations are adjusted for sex, height, and age, therefore in theory the resulting z-scores should be free of any trends in sex, height or age.

Figures 3.12 to 3.20 cover plots of the individual z-scores for each spirometric parameter against height in centimetres and test age in years. For $FEV_{0.5}$ measurements the available reference equations cover only the infancy (up to 2.5 years old) and the available reference equations to calculate the $FEV_{0.75}$ z-scores are also covering the age range after infancy, i.e. preschool and school age only. For this reason figures 3.12 and 3.14 comprise only a limited age range.

We can see that for $FEV_{0.5}$, FVC and FEF_{25-75} there are a few z-scores out of normal range which their absolute value is out of the normal range ($|z-score| > 1.96sd$). Referring to our discussion in section 3.3.2, this clearly shows that the reference equations used to calculate the z-scores is not adjusted for the home made instrument (RASP) and causing the bias from mean value of 0 and standard deviation of 1.

This can wrongly guide us to misclassify these infants as subjects with abnormal results. These discrepancies can be the result of device dependency characteristics which is inevitable due to the integration of flow to volume together with the BTS correction. There are currently no accepted wave-forms with which to compare outcome measures for infant $FEFV$ equipment. We therefore cannot ascertain whether the previous "in-house" systems or new commercially available devices best approximate the "truth".

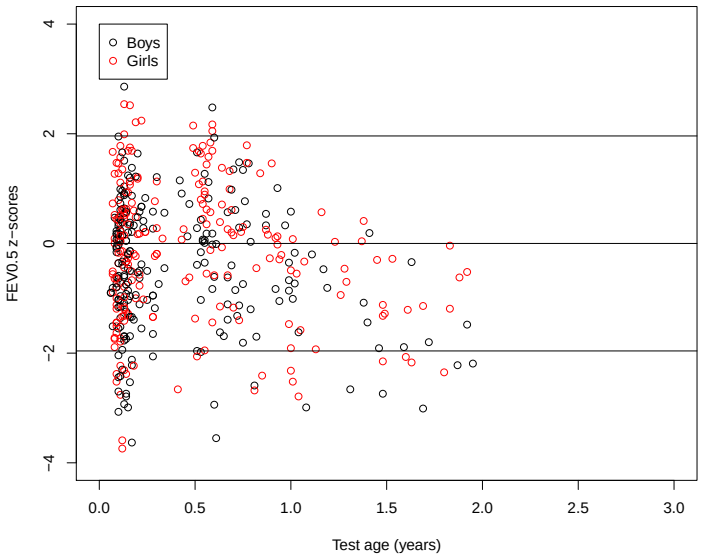


Figure 3.12: Scatter plots of $FEV_{0.5}$ z-scores in infancy against test age in years. The available, published reference equations for $FEV_{0.5}$ only covers the infancy.

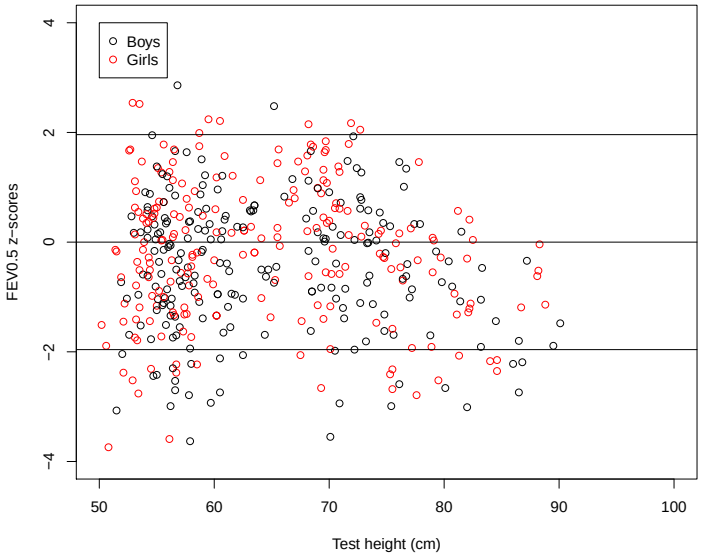


Figure 3.13: Scatter plots of $FEV_{0.5}$ z-scores in children against test height. The available, published reference equations for $FEV_{0.5}$ only covers the infancy.

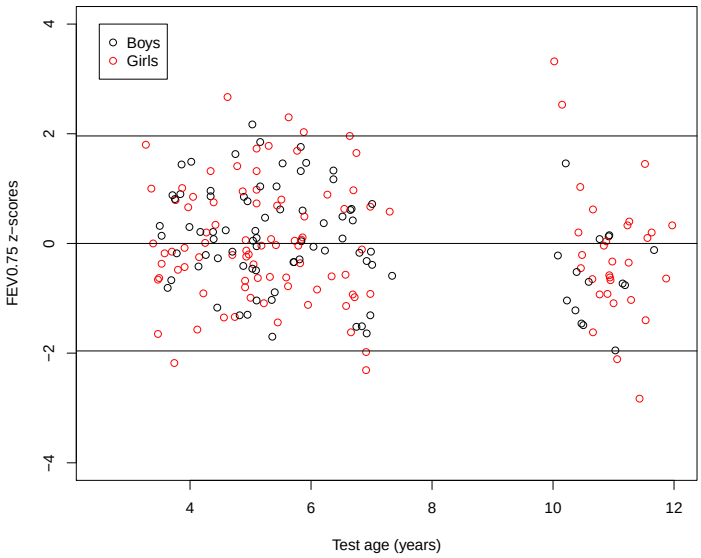


Figure 3.14: Scatter plots of $FEV0.75$ z-scores in children against test age in years. The available reference equations are covering preschool and school age children.

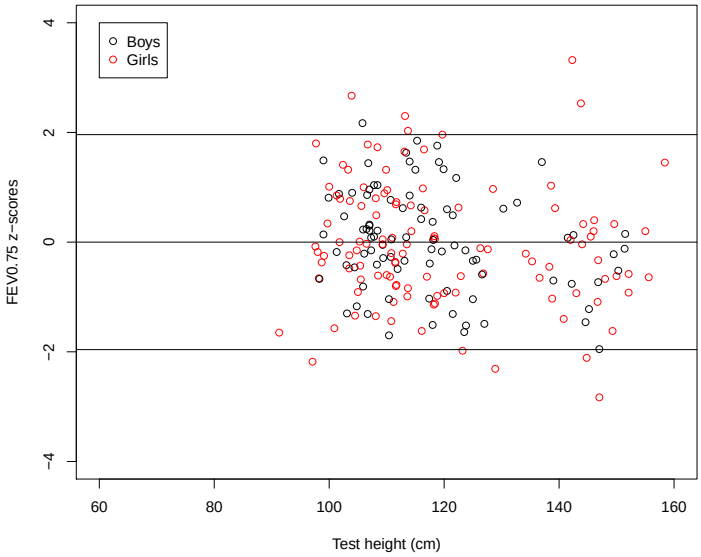


Figure 3.15: Scatter plots of $FEV0.75$ z-scores against test height. The available reference equations are covering preschool and school age children.

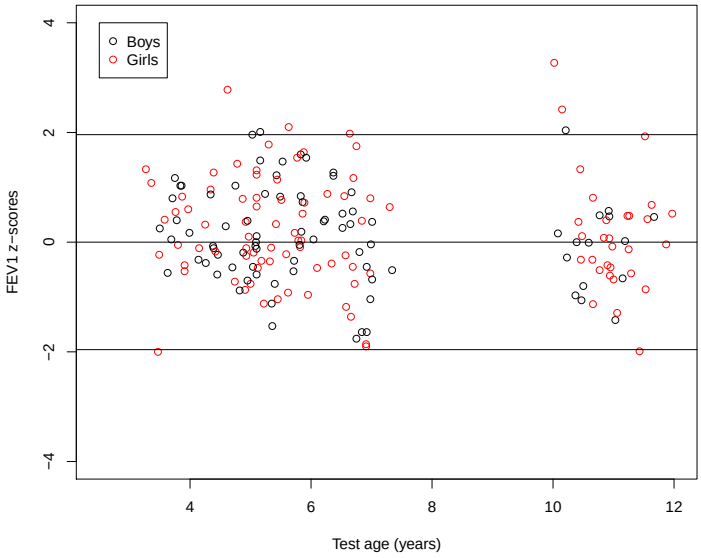


Figure 3.16: Scatter plots of $FEV1$ z-scores against test age in years. Since infants are not able to produce FEV in 1 second, the published and accessible reference equations for FEV_1 are only covering preschool and school age.

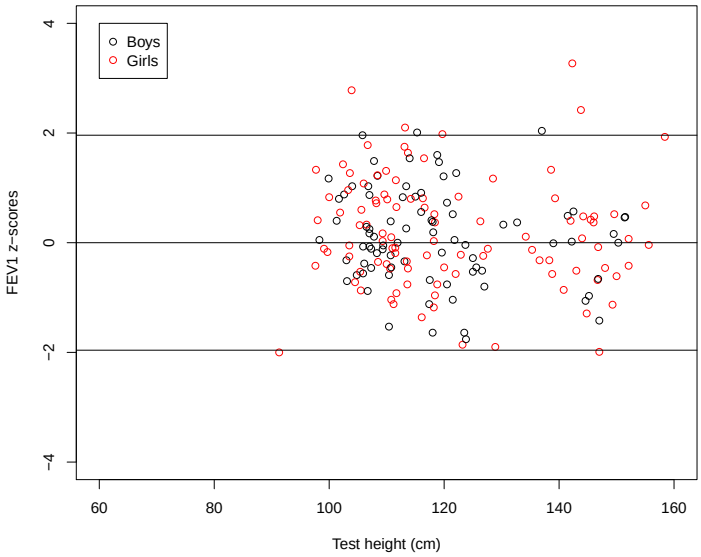


Figure 3.17: Scatter plots of $FEV1$ z-scores against test height. Since infants are not able to produce FEV in 1 second, the published and accessible reference equations for FEV_1 are only covering preschool and school age.

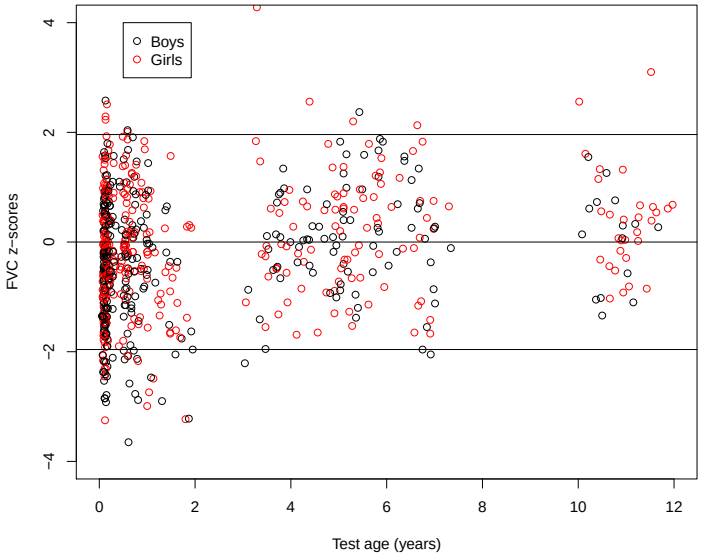
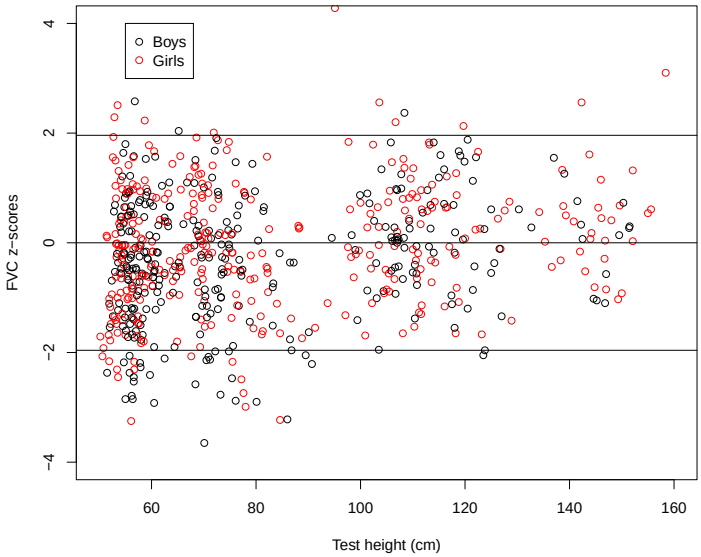


Figure 3.18: Scatter plots of FVC z-scores against test age in years. Available reference equations for FVC cover the age range from infancy up to school age.



(a)

Figure 3.19: Scatter plots of FVC z-scores against test height. Available reference equations for FVC cover the age range from infancy up to school age.

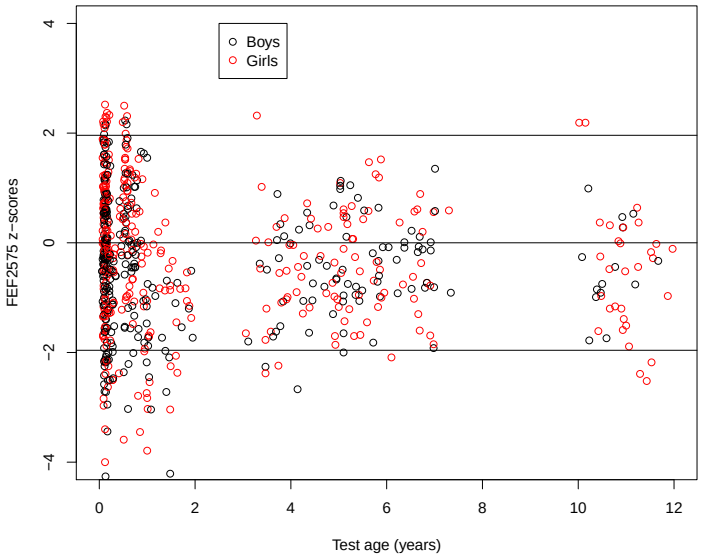


Figure 3.20: Scatter plots of FEF_{25-75} z-scores against test age in years. Available reference equations for FEF_{25-75} cover the age range from infancy up to school age.

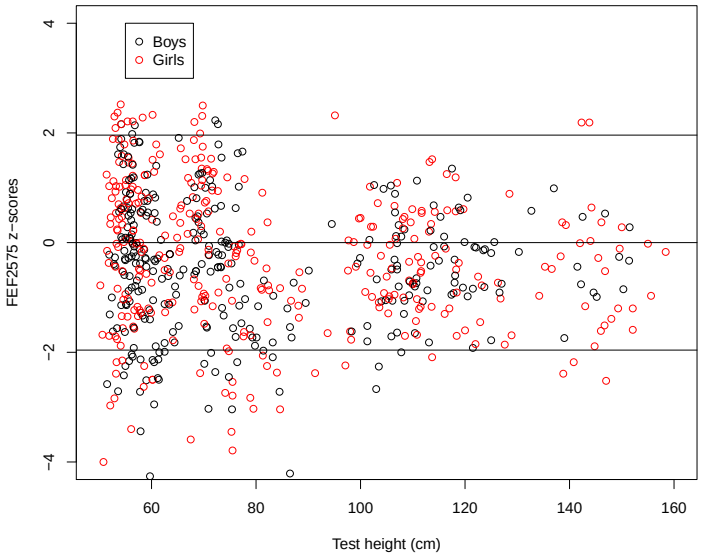


Figure 3.21: Scatter plots of FEF_{25-75} z-scores against test height. Available reference equations for FEF_{25-75} cover the age range from infancy up to school age.

3.4.3 Visualization of $FEFV$ z-scores without observations measured by RASP

Figures 3.22 to 3.31 cover the individual z-scores for each spirometric parameter against height in centimetres and test age in years, when the home made instrument measurements (RASP) are excluded. Data were available from 227 subjects with 366 observations. We saw in the previous section that the reference equations used to express the measurements as z-scores were not adjusted for in-house instrument and it was causing a bias from normal range which was leading to misclassification of some of the subjects as not normal or unhealthy children. In this section the z-scores are visualized without the data measured by home made instrument (RASP) involved in the analysis. The data measured by the Jaeger Master screen instrument are within the normal range. However, there seem to be some downward shift for ages > 1 year.

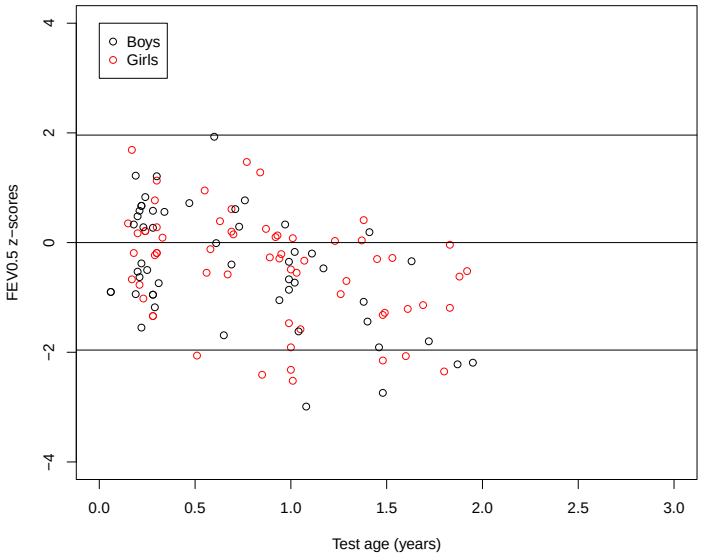


Figure 3.22: Scatter plots of $FEV_{0.5}$ z-scores in infancy against test age in years. The available reference equations for $FEV_{0.5}$ only cover the infancy.

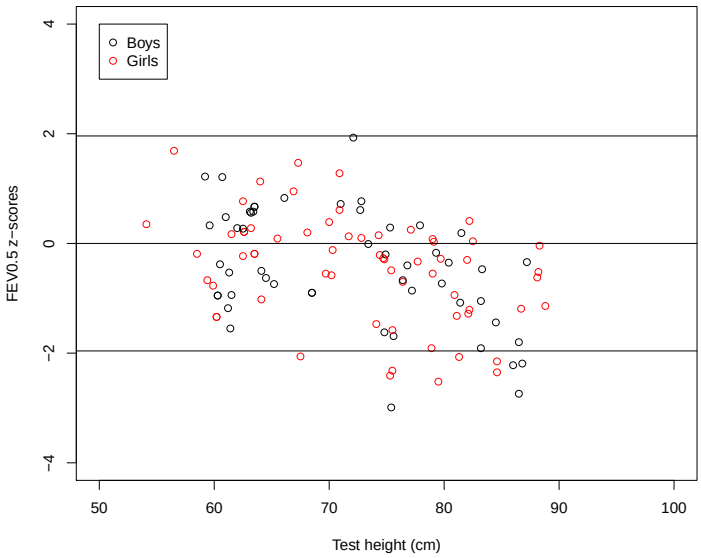


Figure 3.23: Scatter plots of $FEV_{0.5}$ z-scores in children against test height. The available reference equations for $FEV_{0.5}$ only cover the infancy.

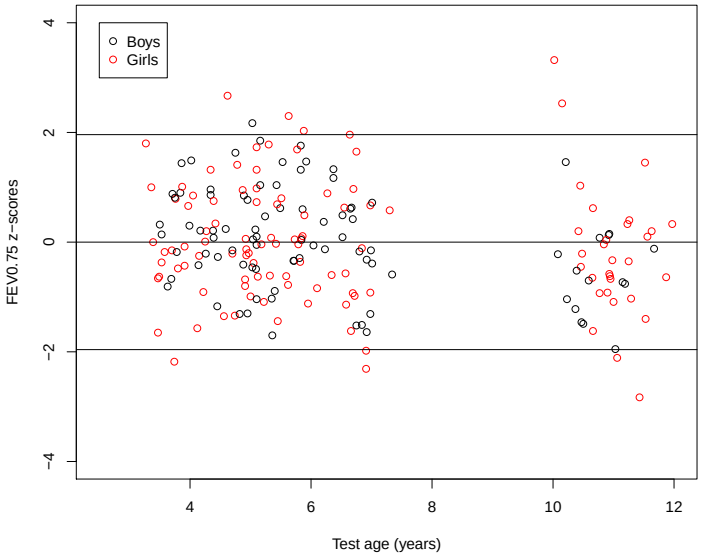


Figure 3.24: Scatter plots of $FEV_{0.75}$ z-scores in children against test age in years. The available reference equations are covering preschool and school age children.

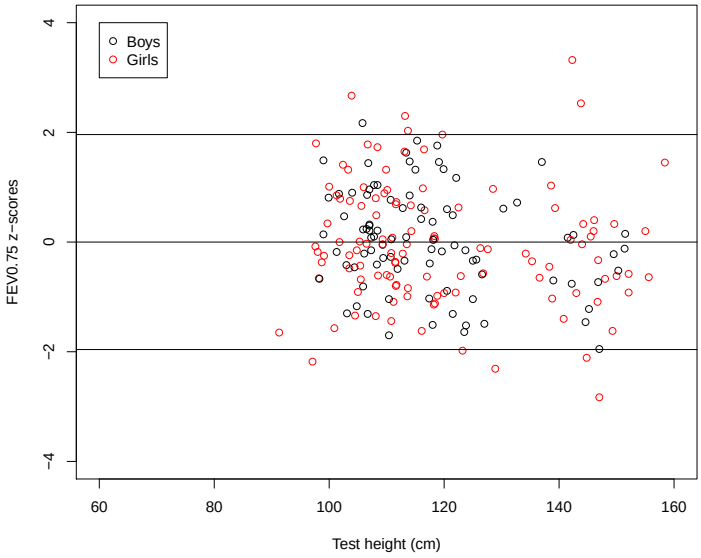


Figure 3.25: Scatter plots of $FEV_{0.75}$ z-scores against test height. The available reference equations are covering preschool and school age children.

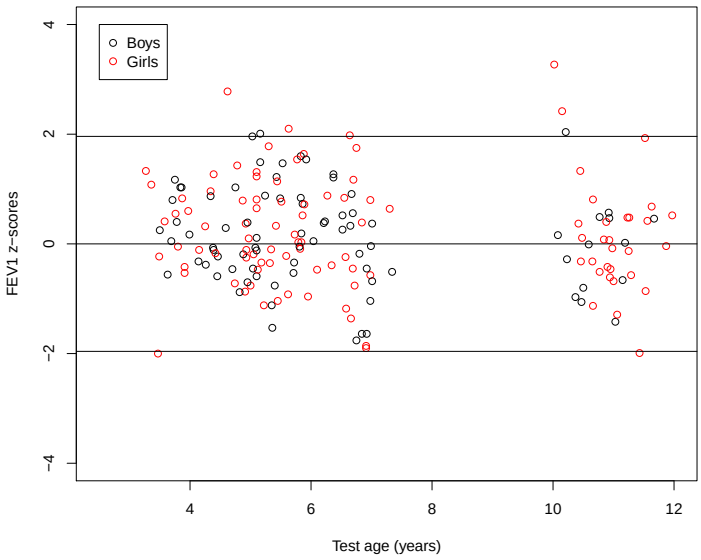


Figure 3.26: Scatter plots of FEV_1 z-scores against test age in years.

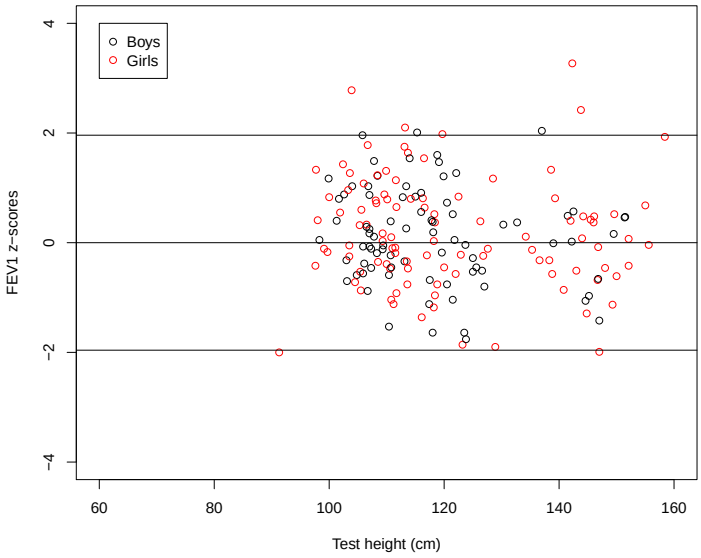


Figure 3.27: Scatter plots of FEV_1 z-scores against test height.

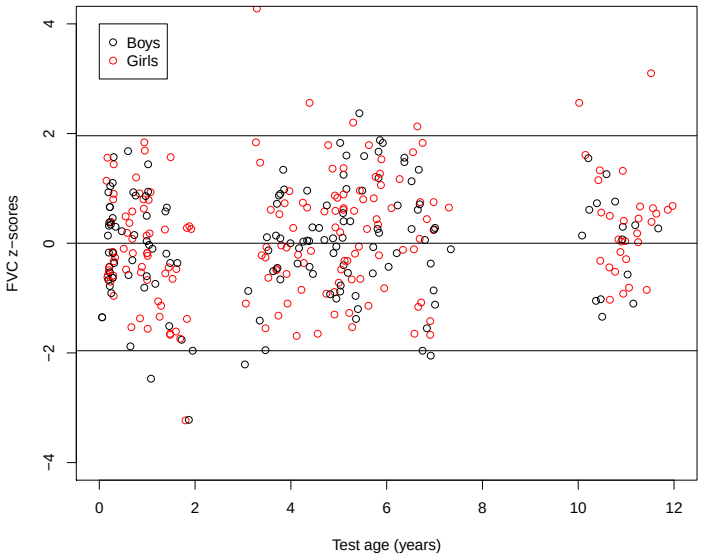


Figure 3.28: Scatter plots of FVC z-scores against test age in years.

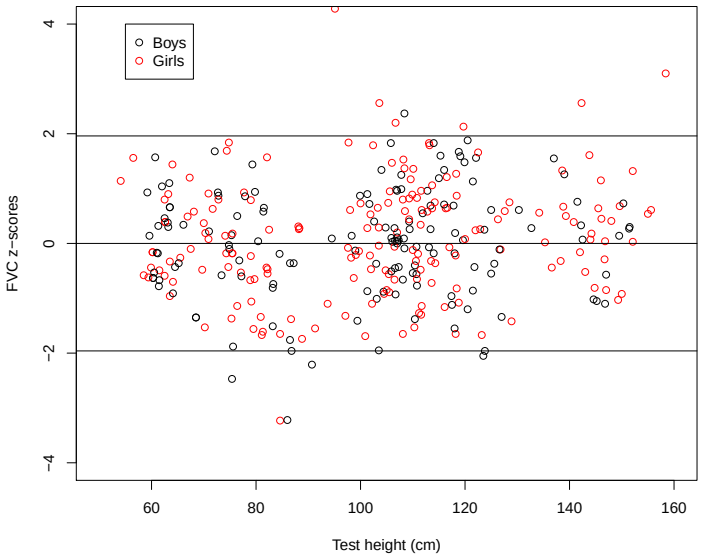


Figure 3.29: Scatter plots of FVC z-scores against test height.

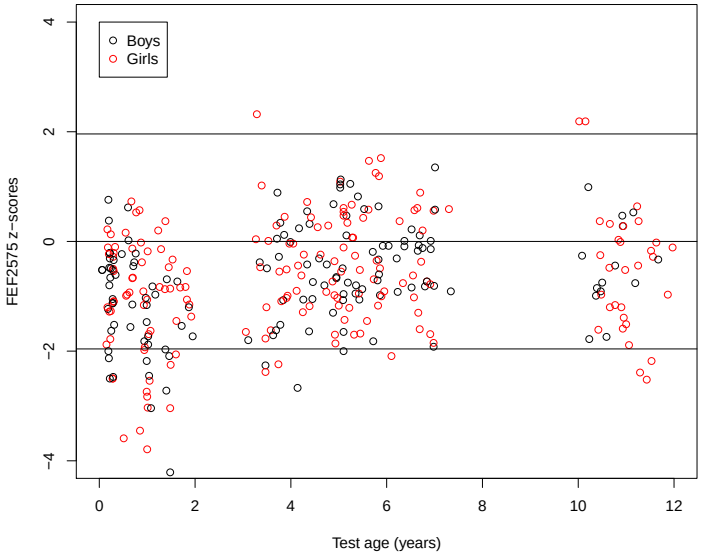


Figure 3.30: Scatter plots of FEF_{25-75} z-scores against test age in years. Available reference equations for FEF_{25-75} cover the age range from infancy up to school age

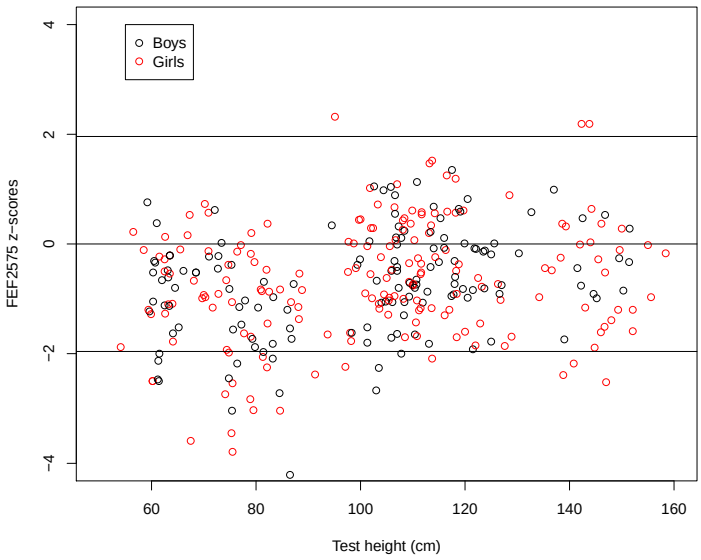
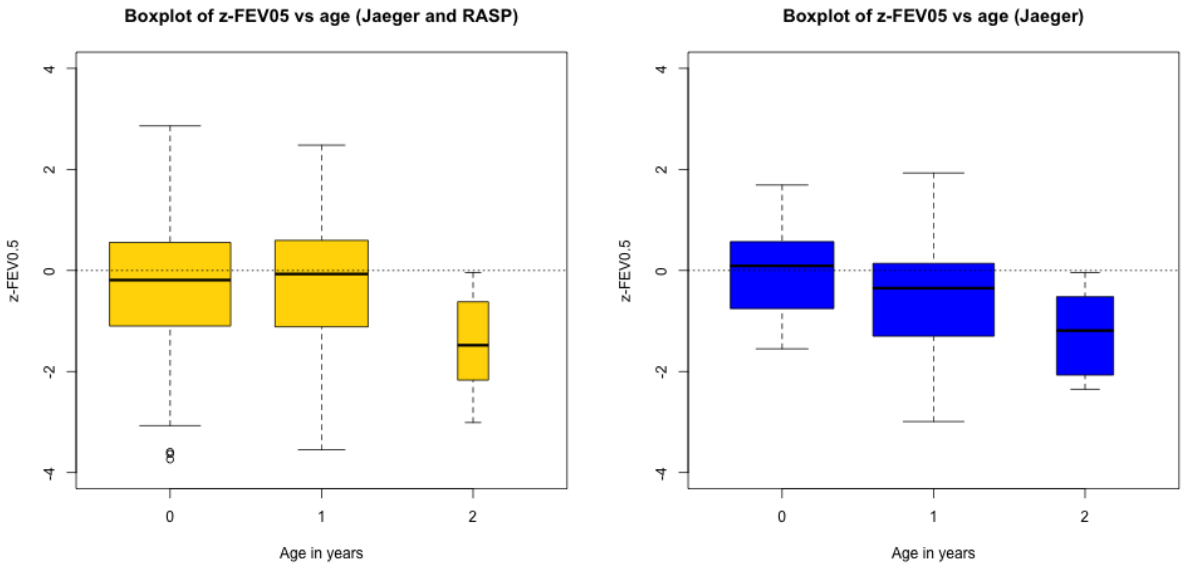


Figure 3.31: Scatter plots of FEF_{25-75} z-scores against test height.

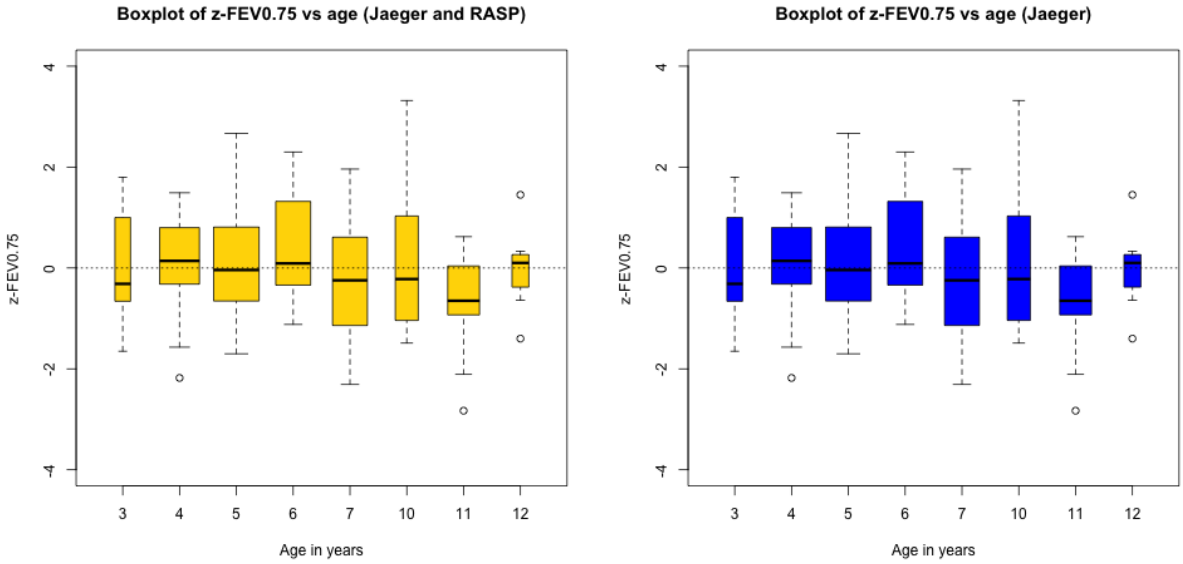
3.4.4 Group means over time

The following box-plots display differences between age groups. The spacings between the different parts of the box helps to indicate the degrees of dispersion and skewness in data, and identify outliers. Yellow box plots display the population when both measurements by RASP and Jaeger are involved and the blue ones are representing the data measured by Jaeger only. We were expecting to have the mean of z-scores to be around 0, which regarding to these figures and also to results presented in Table 3.4 is not true for some variables like $zFEV_{0.5}$ and $zFEF_{25-75}$.



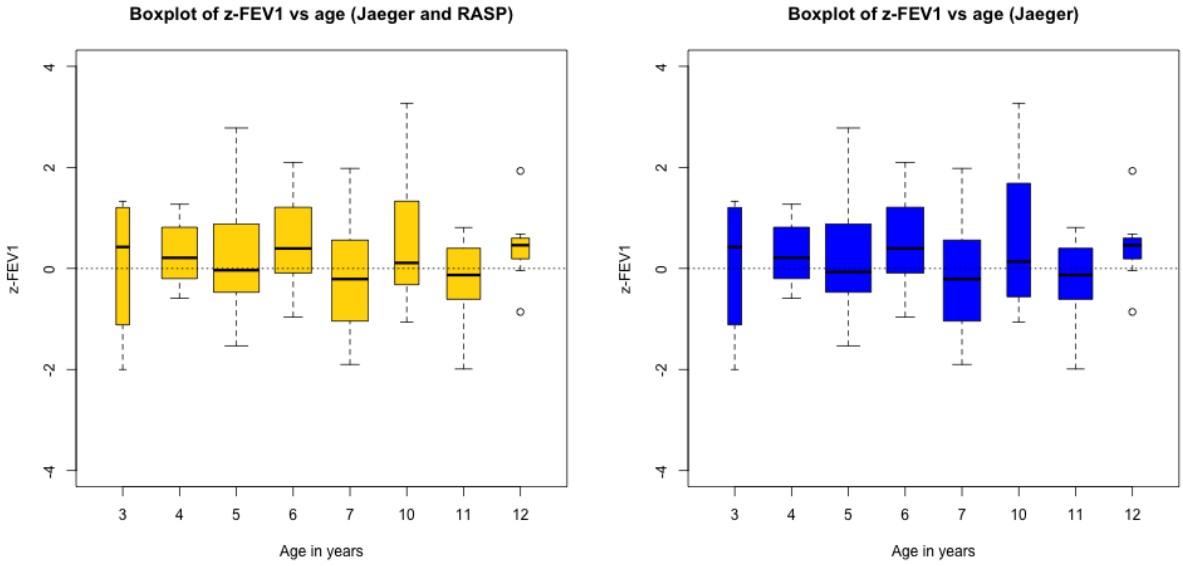
(a) (b)

Figure 3.32: $FEV_{0.5}$ z-scores cover only infancy. As we can see in both plots we have discrepancy from mean 0, specifically around age 2 years old. After exclusion of RASP from data we do not see the outliers any more. The upper quartile and lower quartile is less dispersed than the yellow box-plot and median seems to deviate from zero, however referring to table 3.4, overall mean is shifted towards 0.



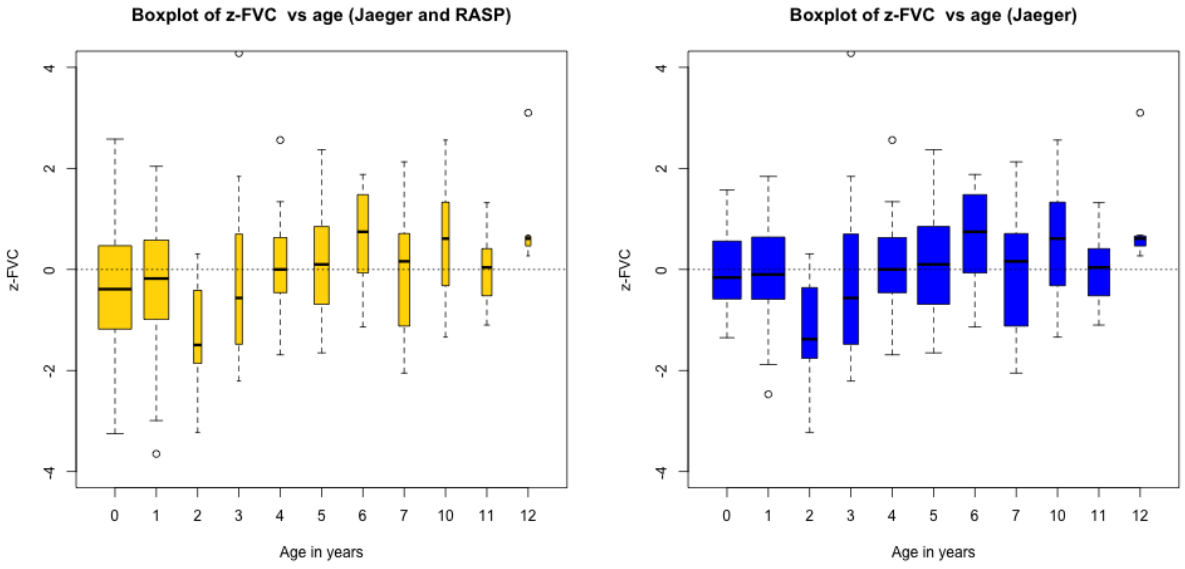
(a) (b)

Figure 3.33: Box plots of z-scores of $zFEV_{0.75}$ with respect to test age in years. $FEV_{0.75}$ z-scores cover preschool age up to school age. However we can see a gap between 7 to 10 years old, which can be seen as a large gap in Figure 3.3. Since infants are not able to produce $zFEV_{0.75}$, presence or absence of measurements from RASP did not affect the $zFEV_{0.75}$ values.



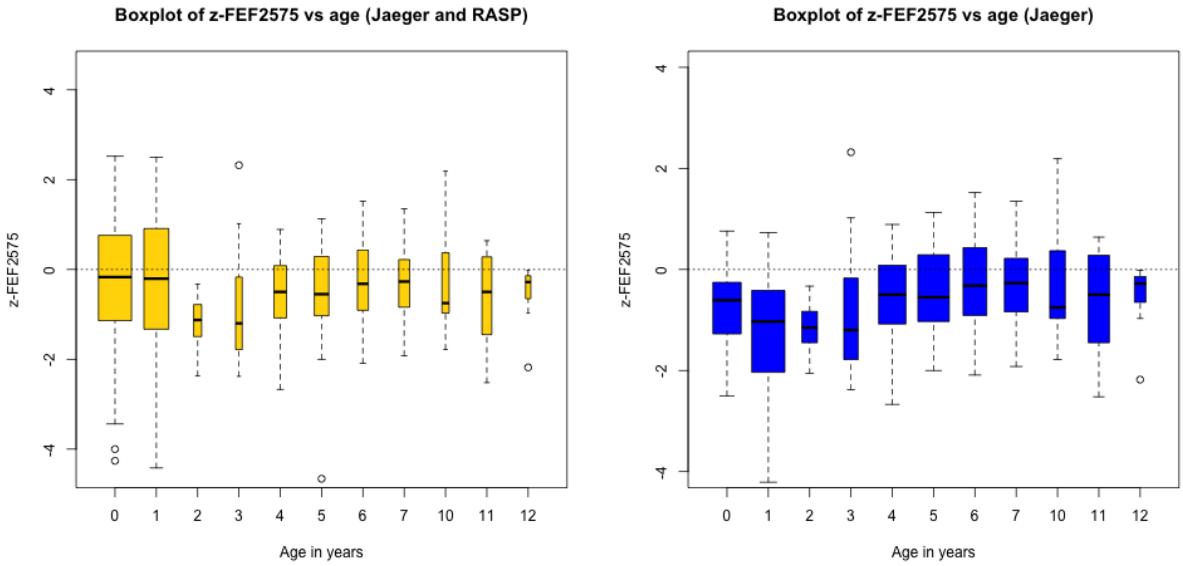
(a) (b)

Figure 3.34: Box plots of z-scores of $zFEV_1$ with respect to test age in years. FEV_1 z-scores cover preschool age up to school age. We can see a gap between 7 to 10 years old, which is also displayed as large gap in Figure 3.4. There is no change in FEV_1 z-scores before and after deletion of RASP data. The reason is that infants are not able to produce FEV_1, therefore exclusion did not have any effect on FEV_1 z-scores.



(a) (b)

Figure 3.35: Box plots of z-scores of $zFVC$ with respect to test age in years. The upper quartile and lower quartile is less dispersed after the exclusion of the RASP, in comparison to the yellow box-plots. However in both plots we see a down shift from zero in two years old infants.



(a) (b)

Figure 3.36: Box plots of z-scores of $zFEF_{25-75}$ with respect to test age in years. Exclusion of RASP measurements from data had a significant effect on FEF_{25-75} z-scores and as we can see the median have a down shift from zero in infancy population.

3.4.5 Individual profile plots

In previous sections an over all insight of collected data is presented. Since the nature of the data set is longitudinal, looking at how each person changes over time would help to see how much variation a person have had during his/her growth. The individual profile plots of the z-scores changes of our test subjects are displayed in Figure 3.37 to Figure 3.41. Since all the subjects did not recruited in test at the same time or at the same age, the time of the first occasion is set to be zero in order to define a baseline time. All other subsequent test occasions are calculated with respect to this baseline time.Most of the infants had two visits within 1 year of time, while only five of them followed up the test after 1 year.

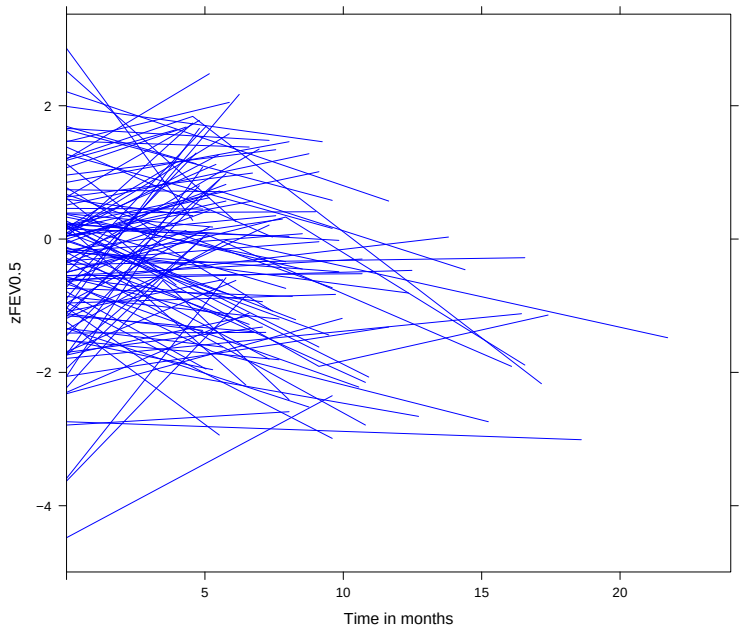


Figure 3.37: Profile plots of individual's $FEV_{0.5}$ z-scores against time (in months).

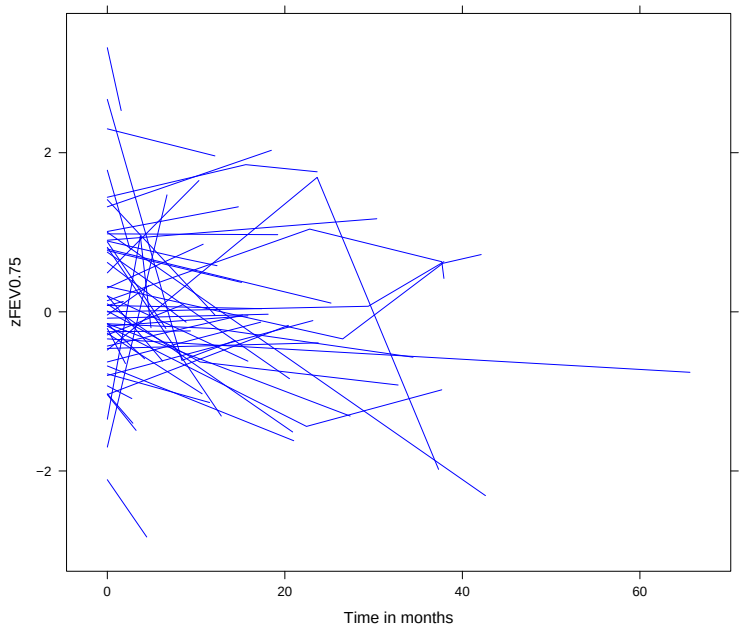


Figure 3.38: Profile plots of individual's $FEV_{0.75}$ z-scores against time (in months).

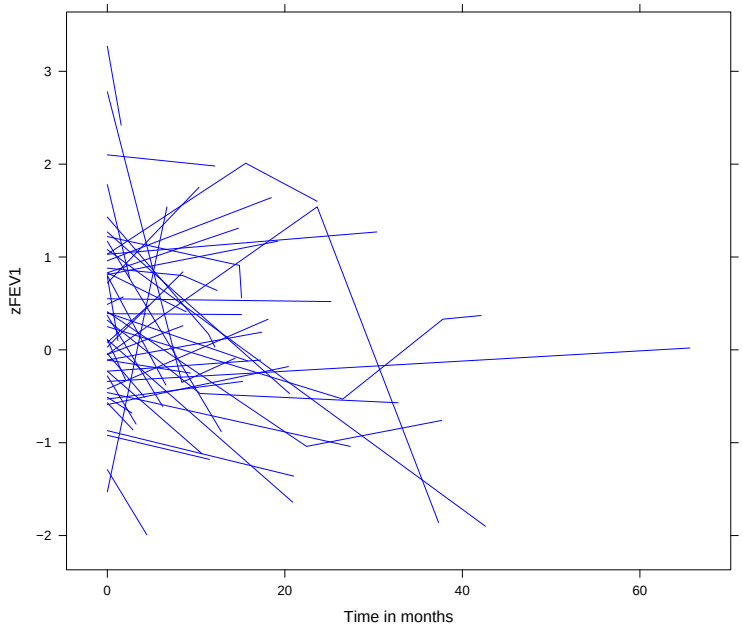


Figure 3.39: Profile plots of individual's FEV_1 z-scores against time (in months).

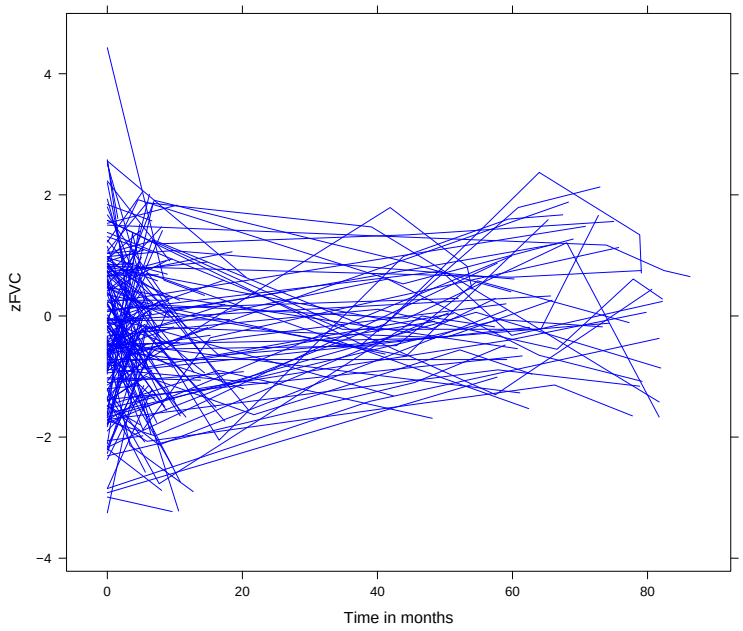


Figure 3.40: Profile plots of individual's FVC z-scores against time (in months).

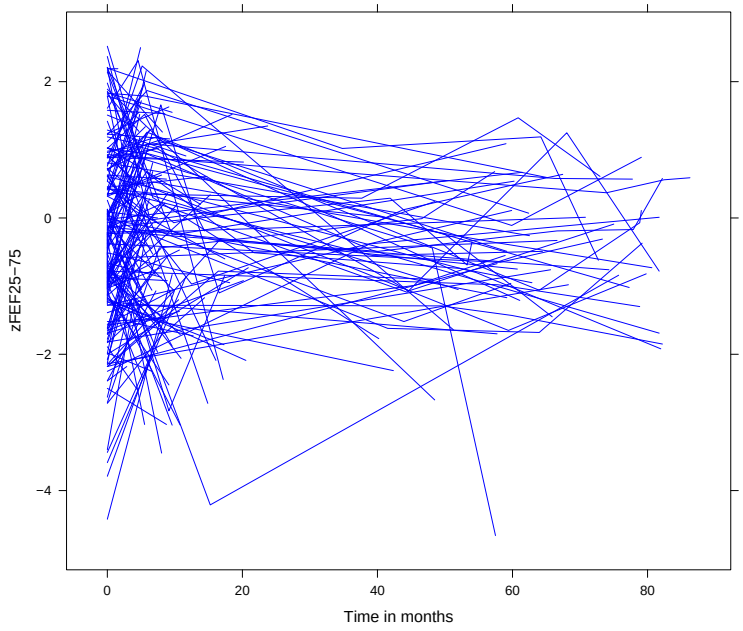


Figure 3.41: Profile plots of individual's FEF_{25-75} z-scores against time (in months).

Chapter 4

$FEFV$ Relationships

4.1 Relationships of $FEFV$ outcomes from infancy to pre-school and school age.

Infants and pre-school children normally have large airways in relation to their lung volume [31], thus they empty their lungs more rapidly than older children and adults do. Therefore the forced expiratory volume in one second (FEV_1) measurement may be difficult to achieve. Even when FEV_1 is available its clinical value remains questionable because it is often roughly equal to the forced vital capacity (FVC). In this case it may be more appropriate to report forced expiratory volume of 3/4 of a second ($FEV_{0.75}$), or forced expiratory volume of half a second ($FEV_{0.5}$) as a means for distinguishing abnormality in this age group [32]. The idea of using $FEV_{0.75}$ in young children was first introduced by Polgar et al. and Cogswell et al. in the 1970s. Subsequently it has been shown that $FEV_{0.75}$ provides similar information to FEV_1. For the same reasons, infant studies use $FEV_{0.5}$, which may prove to be a useful outcome measure in pre-school children and may facilitate longitudinal assessment from birth.

In this study we are interested to find the degree of agreement between measures in infancy and measures in pre-school age or school age. Evaluation of change in FEV outcomes between infancy and pre-school will be compared by the **paired t test**, as well as with a **Pearson correlation** comparing the former and the later values.

Another method to assess the agreement between two methods using repeated measurements are **Bland-Altman plots**, which in the following section is briefly introduced.

4.1.1 Bland-Altman plots

A visualization method for assessing the agreement between repeated measurements are so called Bland-Altman plots. For example, one might compare two scales this way, or two devices for measuring particulate matter [33]. The Bland-Altman plot is more widely known as the Tukey Mean-Difference Plot. Agreement between two methods of clinical measurement can be quantified using the differences between observations made using the two methods on the same subjects. The 95% limits of agreement, estimated by mean difference ± 1.96 standard deviation of the differences, provide an interval within which

95% of differences between measurements by the two methods are expected to lie. The results show the bias, or the average of the differences. The bias is computed as the value determined by one method minus the value determined by the other method. If one method is sometimes higher, and sometimes the other method is higher, the average of the differences will be close to zero. If it is not close to zero, this indicates that the two assay methods are producing different results [34]. The standard deviation value is used to calculate the limits of agreement, computed from equation: $(d-1.96sd, d+1.96sd)$.

To interpret Bland-Altman plots one should answer three questions from the result:

- Of which order is the average difference between methods (the bias)? However this is more a clinical question than a statistical one.
- Is there a trend? Does the difference between methods tend to get larger (or smaller) as the average increases?
- Is the variability consistent across the graph? Does the scatter around the bias line get larger as the average gets higher?

4.2 In case of unavailability of FEV_1, is $FEV_{0.75}$ or $FEV_{0.5}$ an alternative?

As already discussed in previous sections, producing the full expiratory volume in 1 second using $RVRTC$ techniques is not feasible in infancy. However although in pre-school age, children can perform reliable spirometry but because of having short expiratory time, FEV_1 can not always be determined, therefore $FEV_{0.5}$ and $FEV_{0.75}$ will be proposed as alternative measures.
The opportunity for comparing measurements made in infancy with those in later childhood have been limited, and type of investigations at different ages may not be strictly comparable.

4.2.1 Does $FEV_{0.5}$ in infancy relate to $FEV_{0.75}$ in pre-school age?

A comparison between $FEV_{0.5}$, measured during infancy, and $FEV_{0.75}$, measured during the pre-school years, is illustrated by a line plot in Figure 4.1. In total 45 children had data for both test occasions in infancy and pre-school age. Pearson correlation test indicates no significant correlation between these two measurements with a p–value of 0.08 and Pearson's correlation coefficient of $r = 0.17$.

Paired t-test is driven to compare the change in mean between infancy and pre-school age. p–value of 0.001 and the 95% confidence interval of $(-0.73, -0.17)$ show a significant

change in mean between these two growth periods. Since the confidence interval does not contain 0 it implies that there is a statistically significant change between $FEV_{0.5}$ and $FEV_{0.75}$ in infancy through pre-school age. The estimated mean difference was 0.45.

Figure 4.1 shows that 8 (32.6%) infants have z-score values outside the LLN (-1.96 z-scores) and ULN (1.96 z-scores) range. However for 2 subjects $FEV_{0.75}$ falls out of the normal range in their pre-school age.

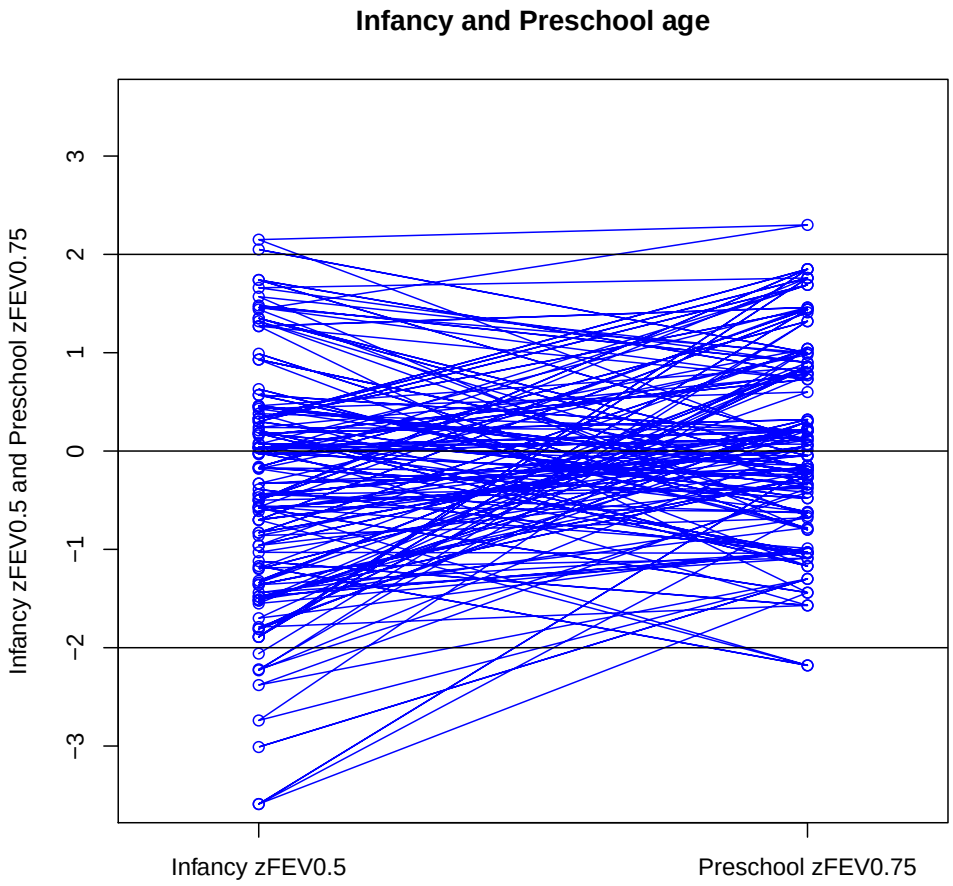


Figure 4.1: Longitudinal values for time forced expiratory volumes: $FEV_{0.5}$ (z-score) measured during infancy and $FEV_{0.75}$ (z score) measured during pre-school years.

4.2.1.1 Bland-Altman analysis

The Bland-Altman plot of the difference between the outcomes of $FEV_{0.5}$ z-scores and $FEV_{0.75}$ z-scores against their mean is presented in Figure 4.2. As we can see the difference between the outcomes tend to get slightly larger as the average increases. Each point in this plot corresponds to one of the lines in Figure 4.1.

Individual $FEV_{0.5}$ z-scores for infants (0 - 2.5 years) were calculated using the raised volume technique (RTC) and the $FEV_{0.75}$ z-scores in pre-school age by using pre-school pulmonary function testing (PFT). These were compared using Bland-Altman analysis to identify any systematic biases and differences between the two.

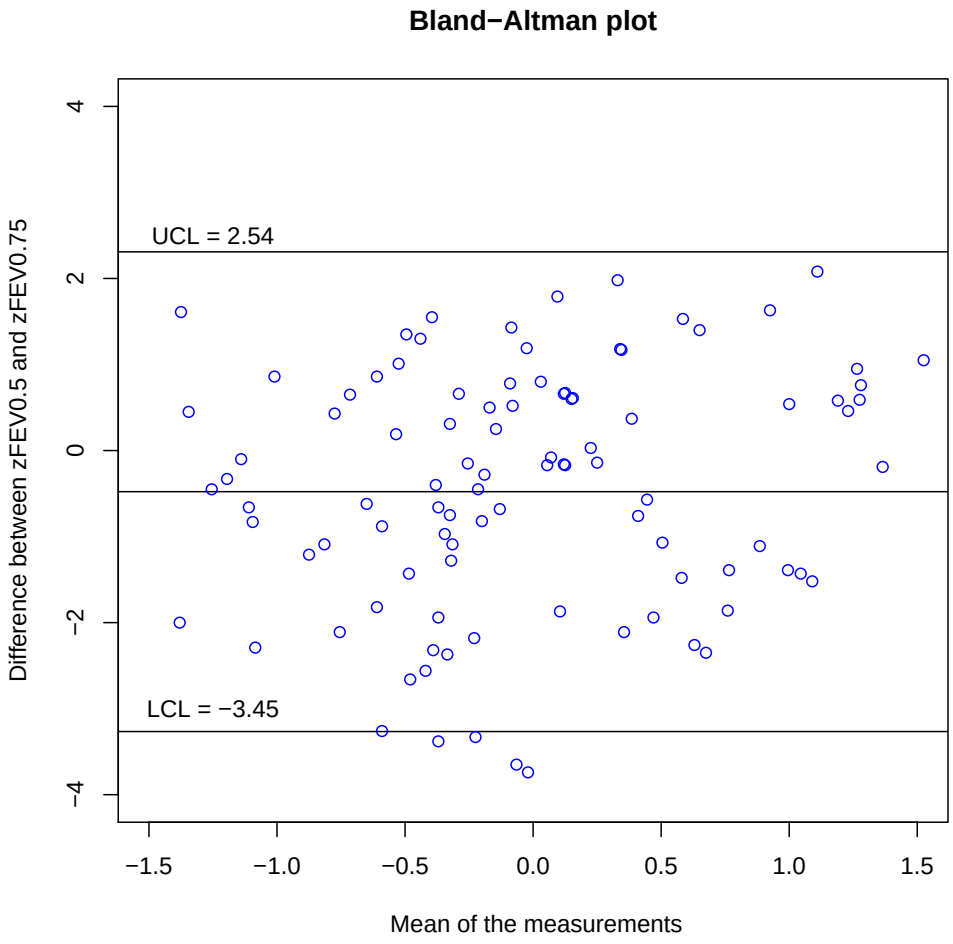


Figure 4.2: Bland-Altman plot

Comparison of the z-scores demonstrated a systematic bias where $FEV_{0.5}$ z-scores were on average 0.45 higher than the $FEV_{0.75}$ z-scores in pre-school. There is no obvious relation between the difference and the mean. Under these circumstances we can summarise

the lack of agreement by calculating the bias, estimated by the mean difference d and the standard deviation of the differences sd for pairs of $FEV_{0.5}$ and $FEV_{0.75}$ measurements. If there is a consistent bias we can adjust for it by subtracting d from the new method. In our data the mean difference is -0.45 and sd is 1.62. We would expect most of the differences to lie between $d - 1.96sd$ and $d + 1.96sd$. If the differences are normally distributed (Gaussian), 95% of the differences will lie between these limits (or, more precisely, between $d - 1.96sd$ and $d + 1.96sd$) presented in the plot by solid black lines.
If provided differences within $d \pm 1.96sd$ would not be clinically important, we could use the two measurement methods interchangeably. We shall refer to these as the "limits of agreement".

$$UCL = d + 1.96sd = 2.54 \quad LCL = d - 1.96sd = -3.45 \tag{4.1}$$

4.2.2 Does $FEV_{0.5}$ in infancy relate to FEV_1 in pre-school age?

In total 37 children had data for both test occasions in infancy and pre-school age. The Pearson correlation test shows a slightly positive correlation between $FEV_{0.5}$ in infancy and FEV_1 in pre-school age with correlation coefficient $r = 0.23$ and a p–value of 0.03. A comparison between $FEV0.5$ measured during infancy and $FEV1$ measured during the pre-school years is illustrated as a line plot in Figure 4.3.

The paired t-test is used to establish if the correlation coefficient is significantly different from zero, and, hence that there is evidence of an association between the two variables. The null hypothesis of true difference in means is equal to 0 is rejected with a p–value of 0.00003. The estimated mean difference of 0.65 shows a significant change when infancy and pre-school evaluations were compared. Figure 4.3 indicates that 11 subjects having FEV_1 z-scores inside the normal limit range, had $FEV_{0.5}$ z-scores outside of the LLN and ULN.

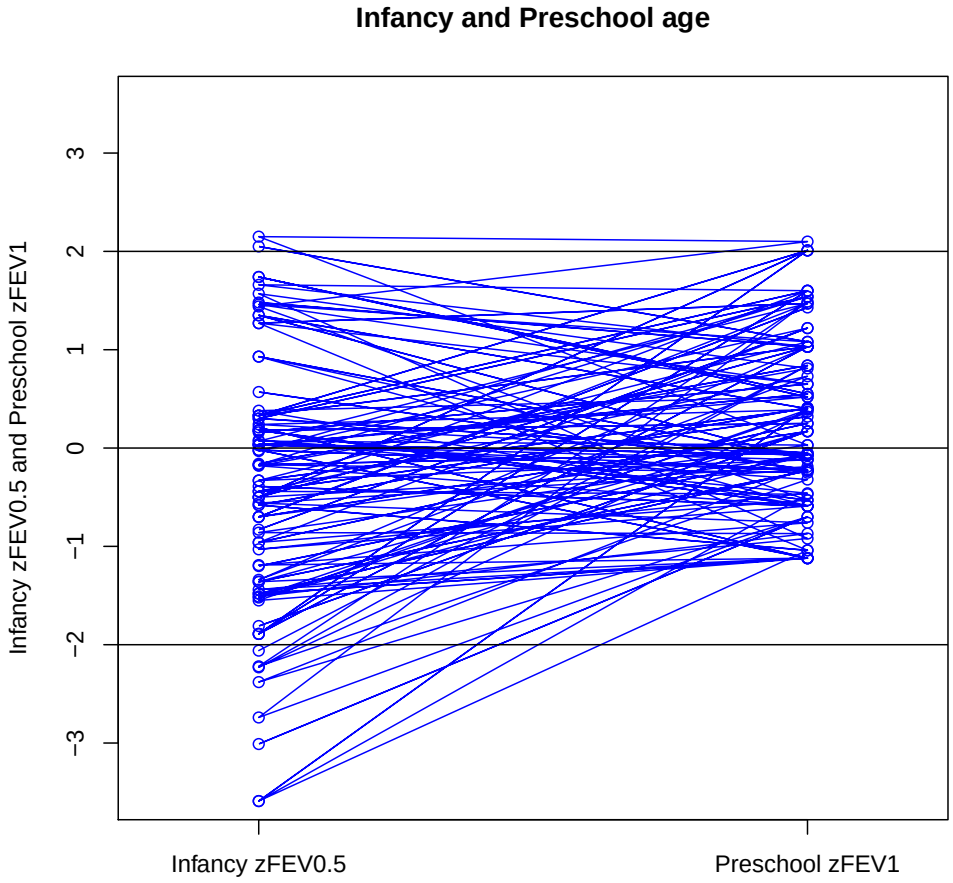


Figure 4.3: Longitudinal values for timed forced expiratory volumes: $FEV_{0.5}$ (z-score) measured during infancy and FEV_1 (z-score) measured during pre-school years.

4.2.2.1 Bland-Altman analysis

Bland-Altman analysis is used to identify any systematic differences between the two values. Comparison of the z-scores demonstrated a systematic bias where the $FEV_{0.5}$ z-scores in infancy were on average 0.65 z-scores higher than the FEV_1 in pre-school age. Figure 4.4 displays that these differences tend to get slightly larger as the average increases. We would expect most of the differences to lie between $d - 1.96sd$ and $d + 1.96sd$. The limit of agreements is plotted by solid lines.

$$UCL = d + 1.96sd = 2.02 \quad LCL = d - 1.96sd = -3.33 \tag{4.2}$$

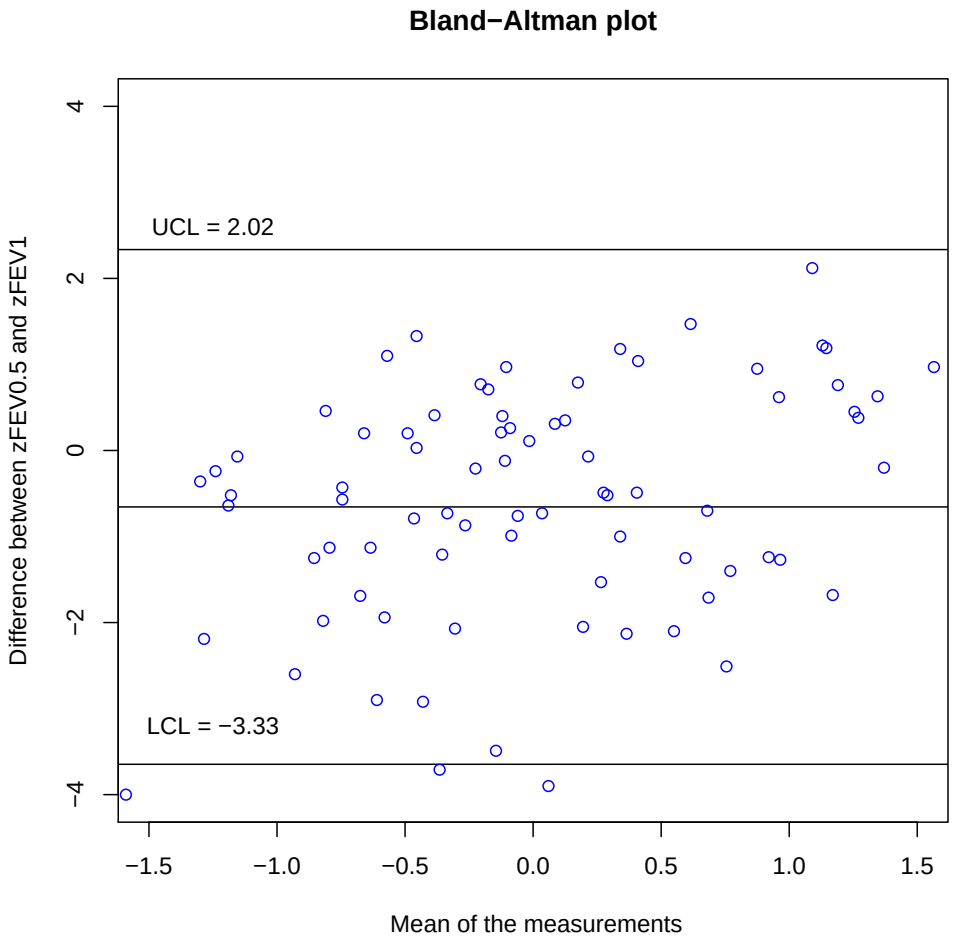


Figure 4.4: Bland-Altman plot of $FEV_{0.5}$ z-score in infancy and FEV_1 z-score in pre-school age.

4.2.3 Does FEF_{25-75} in infancy relate to FEF_{25-75} in school age?

In total 23 children had FEF_{25-75} recorded in both age groups. Pearson correlation test indicates no significant correlation between these two outcomes with $r = 0.18$ and p–value of 0.23. The result of paired t test indicates the estimated mean difference of 0.42. This shows a significant change when infancy and school age evaluations are compared.

Figure 4.5 displays the Bland-Altman plot of comparing FEF_{25-75} in infancy with FEF_{25-75} in school age. There was an obvious systematic bias in FEF_{25-75} z-scores, the mean differences increased as the mean z-score increased with the mean difference of 0.45, and 95% limits of agreement (-4.16, 0.96).

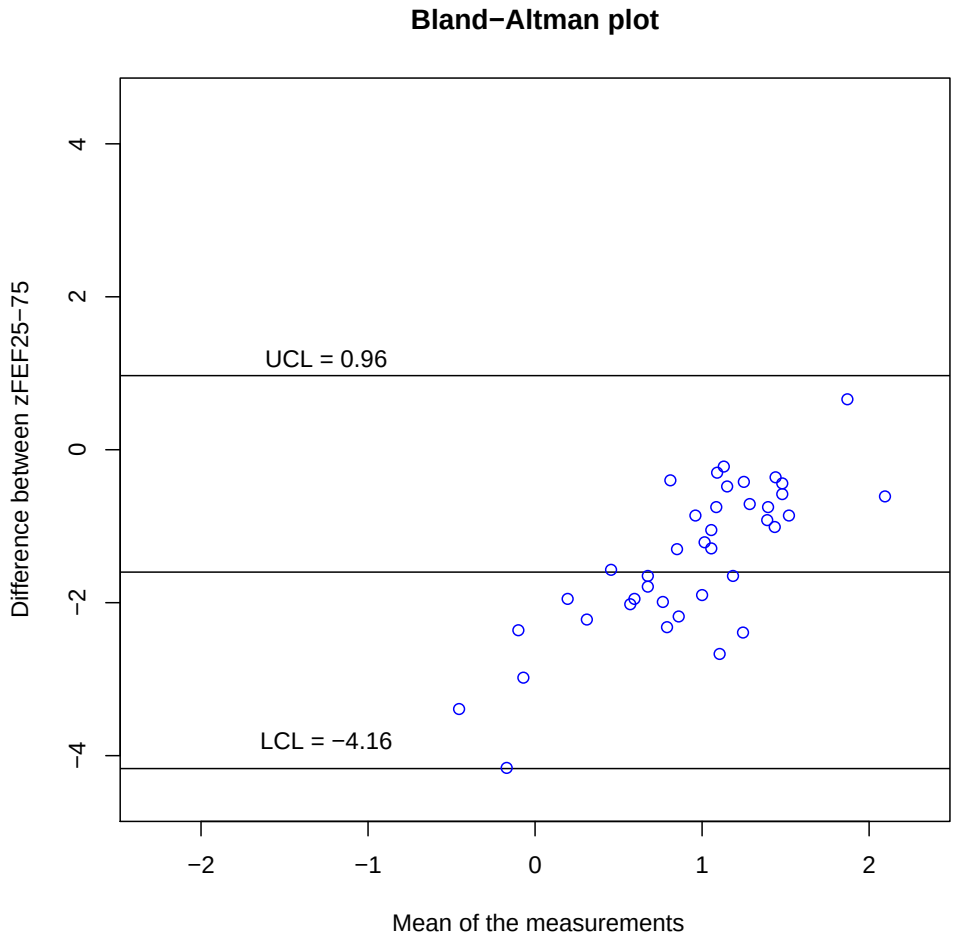


Figure 4.5: Bland-Altman plot of FEF_{25-75} z-score in infancy and FEF_{25-75} z-score in school age

Chapter 5

Between and Within Subjects Variabilities

5.1 Variation among individuals

In chapter 4 the behaviour of the **Forced Expiratory Volumes** and **Forced Expiratory Flows** is visually studied. In chapter 5 the relationships of these volumes and flows is investigated between infancy to pre-school and school age to explore between and within subjects variability. Between subject variability is a measure of the extent to which subjects, on average, perform differently from each other and within subject variability is the measure of average change for an individual in his/her growth path.

As discussed in chapter 3 a particular set of techniques which is useful to study individual's behaviour are multilevel models (Random intercept or random slope models). These models allow each individual to have a different intercept and a different slope, so it allows the explanatory variable to have different effect for each group (individuals).

$$Y_{ij} = \beta_0 + \beta_1 X_{ij} + u_{0j} + u_{1j} + e_{ij}$$

$$e_{ij} \sim N(0, \sigma_e^2)$$

$$\begin{bmatrix} u_{0j} \\ u_{1j} \end{bmatrix} \sim N(0, \Omega_u), \quad \Omega_u = \begin{bmatrix} \sigma_0^2 & \sigma_{01} \\ \sigma_{10} & \sigma_1^2 \end{bmatrix}$$

$\hat{\beta}_0$ and $\hat{\beta}_1$ are estimates of the fixed effect which are the estimations of intercept and slopes of overall population. σ_e^2 is the level one variance which is the residual variance across all occasions of measurements for individuals. In the variance matrix, σ_0^2 is the variance in intercept between individuals and σ_1^2 is the variance in slope between individuals and at the end σ_{01} indicates the covariance between intercept and slope. In the following we present the results of fitting a linear mixed effect model (random intercept model) on z-scores of $FEFV$ outcomes with the help of the *lme*4 package in R and MLwiN software developed in University of Bristol (`http://www.bristol.ac.uk/cmm/software/mlwin/`).

Graphical methods can also be used to explore the magnitude of person-to-person variability in outcomes over time. One approach is to create a panel of individual line plots for each study participant. These plots can then be inspected for both the amount of

variation from subject-to-subject in the overall "level" of the response, and the magnitude of variation in the "trend" over time in the response.

We are interested to investigate between subject variability for the measure of the extent to which subjects, on average, perform differently from each other irrespective of the experiment condition. The outputs of fitting a random intercept model to data is presented in next section. Since $FEFV$ outcomes are expressed as z-scores and been adjusted in terms of test age, height and gender, we would expect to get the estimation of β_0 and β_1 to be approximately zero but also showing a high inter-subject variability.

5.1.1 Z-score of Forced Expiratory Volume in 0.5 second ($FEV_{0.5}$)

Due to space limitations, the results of the model is presented in tables and the output of the software is skipped and plots are presenting the trajectories of model with Jaeger instrument data.

	Parameter	Estimation(SE) (all instruments, n = 273)	(Jaeger, n = 54)
Fixed Effects			
Initial status	Intercept(β_0)	-0.252(0.074)*	-0.080(0.103)
Rate of change	Slope(β_1)	-0.028(0.012)*	-0.094(0.014)*
Variance Comp.			
Level 1	Within Person(σ_e^2)	0.890(0.098)*	0.460(0.100)*
Level 2	Between Person(σ_0^2) In initial status	0.680(0.127)*	0.342(0.131)*
Goodness-of-fit			
IGLS Deviance $-2 \times$ loglikelihood		1317.105	479.727

Table 5.1: Fitting multilevel model for change in z-scores of $FEV_{0.5}$. Results are presented as estimation of the parameters model and significant results are showed by *. Standard errors are in brackets.

Interpretation of outcome:
1. Having all observations measured by Jaeger and RASP involved in the analysis, the overall intercept and slope estimations are -0.25 and -0.03 respectively. To check the significance for each of these estimations, we can easily perform $|estimate| - 2 \times sde$. If this is higher than zero (i.e. the estimate is more than two standard errors away from zero), it is very unlikely that the true value is zero, i.e. it is statistically significantly non zero. As a result we can see the estimated intercept and variance for the model having all instruments involved are significantly different from zero. This is again a confirmation for the lack of device adjusted reference equations that leads to bias the expected estimation. The estimation of σ_e^2, the level one residual variance across all occasions of measurements (within subject variability) is 0.89, this means individual's $FEV_{0.5}$ z-scores varies significantly during their infancy with standard error of 0.11. σ_0^2 the variance between individuals (between subjects variability) is significant and estimated as 0.68 with standard error 0.13.

2. Having only observations measured by Jaeger involved in the analysis, the estimate of the overall intercept and slope are -0.08 and -0.09 respectively. They are statistically not significantly different from zero. However the within individual variance is significantly non zero and estimated as 0.46 with standard error 0.10. The between individuals variability is estimated as 0.34 with standard error 0.13.

To make it easier to explore we posit a linear change trajectory on data and define a base line time. Time of the first test occasion set to be zero and then the time difference between consecutive tests is calculated in month scale. In Figure 5.1 we can see the predicted individual lines from a random slope model and the pattern of the individuals summary lines.

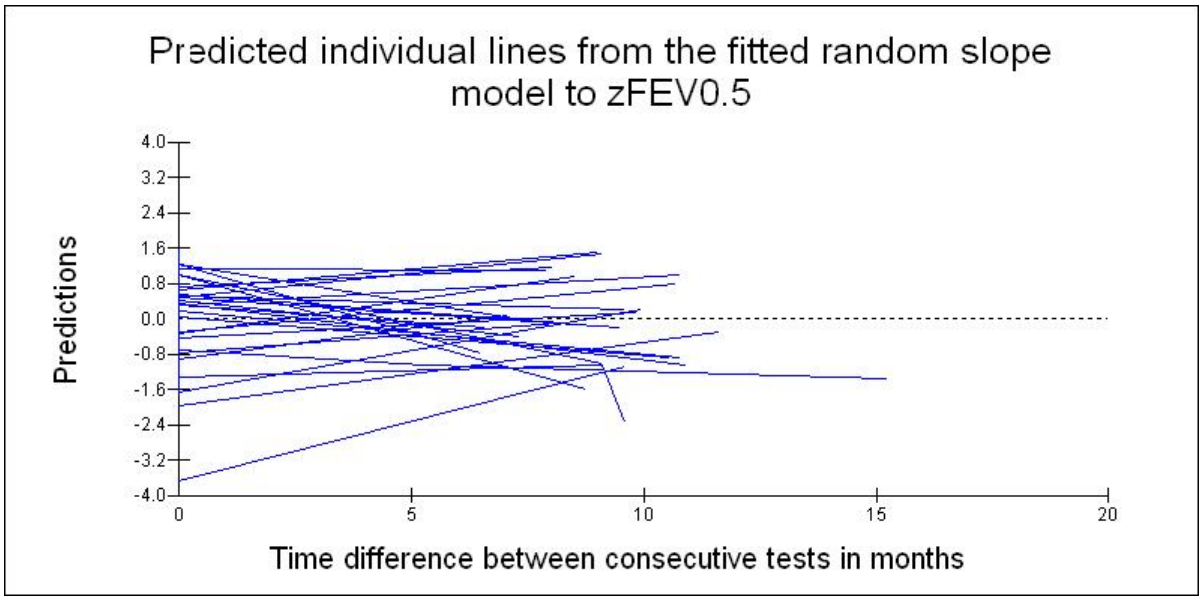


Figure 5.1: Random slope model predictions versus consecutive time difference between tests in months.

5.1.2 Z-score of Forced Expiratory Volume in 75 second ($zFEV_{0.75}$)

Table 5.2 shows results of fitting a random slope model to $FEV_{0.75}$ z-scores.

	Parameter	Estimation(SE) (all instruments, n = 113)	Estimation(SE) (Jaeger, n = 100)
Fixed Effects			
Initial status	Intercept(β_0)	-0.049(0.114)	-0.049(0.114)
Rate of change	Slope(β_1)	0.001(0.003)	0.001(0.003)
Variance Comp.			
Level 1	Within Person(σ_e^2)	0.573(0.101)*	0.740(0.118)*
Level 2	Between Person(σ_0^2) In initial status	0.382(0.180)*	0.611(0.191)*
	In rate of change(σ_1^2)	-0.008(0.006)	-0.01(0.006)
Goodness-of-fit			
IGLS Deviance $-2 \times$ loglikelihood		574.562	564.78

Table 5.2: Results of fitting a multilevel model for change in z-scores of $FEV_{0.75}$, results in this table are presented as estimation of the parameters and standard error in brackets. * represents statistical significance.

Interpretation of outcome:
1. Involving all observations measured by Jaeger and RASP in the analysis, the overall estimations of intercept and slope are -0.05 and -0.001 respectively. They are not statistically significant different from zero. The level one residual variance across all occasions of measurements σ_e^2 is estimated as 0.57 with standard error of 0.10. The estimation of σ_e^2 shows a significant change within the individual's $FEV_{0.75}$ z-scores in their pre school and school age. Besides σ_0^2, the variance in intercept between individuals shows a significant difference between individuals. It is estimated as 0.38 with standard error 0.18 and furthermore the estimation of the between individual variance in rate of change (slope) is not significantly different from zero and is estimated as -0.008 with standard error 0.006.

2. Involving only observations measured by Jaeger in the analysis, the estimation of overall intercept and slope are -0.05 and 0.001 respectively. This shows a non significant difference from zero similar to the case when all instruments were presented. The within individual variance is estimated as 0.74 with standard error 0.12 and analogue to the previous condition $FEV_{0.75}$ z-scores of individuals change significantly in their preschool

and school age. The estimated value of between individuals variability in intercept shows a significant difference and is estimated as 0.61 with standard error 0.19. However the estimation of the between individual variance in rate of change (slope) does not show a significance difference between individuals; it is -0.011 with standard error 0.006.
Figure 5.2 shows the predicted individual lines from a random slope model.

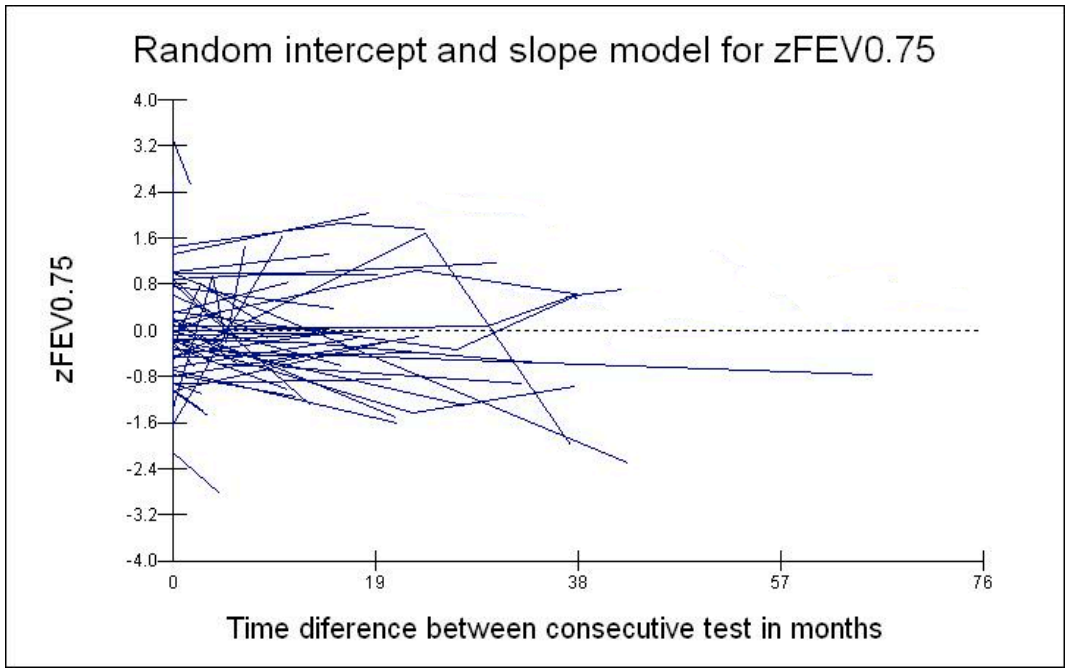


Figure 5.2: Random slope model predictions versus consecutive time difference between tests in months.

5.1.3 Z-score of Forced Expiratory Volume in 1 second ($zFEV_1$)

Table 5.3 shows the results of summary table of fitting a random slope model to $zFEV_1$.

	Parameter	Estimation(SE) (all instruments, n = 103)	(Jaeger, n = 90)
Fixed Effects			
Initial status	Intercept(β_0)	0.161(0.099)	0.16(0.100)
Rate of change	Slope(β_1)	-0.002(0.003)	-0.001(0.003)
Variance Comp.			
Level 1	Within Person(σ_e^2)	0.7113(0.11)*	0.511(0.093)*
Level 2	Between Person(σ_0^2) In initial status	0.190(0.134)	0.470(0.159)*
	In rate of change(σ_1^2)	-0.011(0.005)*	-0.016(0.005)*
Goodness-of-fit			
IGLS Deviance $-2 \times$ loglikelihood		490.738	479.72

Table 5.3: Results of fitting a multilevel model for change in z-scores of FEV_1, results in this table are presented as estimation and standard error in brackets. * represents statistical significance.

Interpretation of outcome:
1. Having all observations measured by Jaeger and RASP in the analysis, the estimations of the overall intercept and slope are 0.16 and -0.002 respectively. This means the estimated intercept and slope are not significantly different from zero. The estimated value for σ_e^2, the level one residual variance across all occasions of measurements is 0.71 with standard error 0.11, which presents a significant change in FEV_1 z-scores from preschool to school age within individuals. The estimation of σ_0^2 the variance in intercept between individuals is not significantly different from zero and is estimated as 0.19 with standard error 0.13. Estimation of the between individual variance in rate of change (slope) is -0.011 and standard error 0.005.

2. Having only observations measured by Jaeger in the analysis, the estimations of overall intercept and slope is 0.16 and -0.001 respectively that implies there is not a significant difference to zero for the overall intercept and slope estimation. In this case the within individual variance estimation (0.51) with standard error (0.09) shows a significant change

in individual's FEV_1 z-scores from preschool to school age. The between individuals variability in intercept is estimated as 0.47 with standard error 0.15 and the estimation of the between individual variance in rate of change (slope) is -0.016 with standard error 0.005. Plot 5.3 shows predicted individual lines from a random slope model.

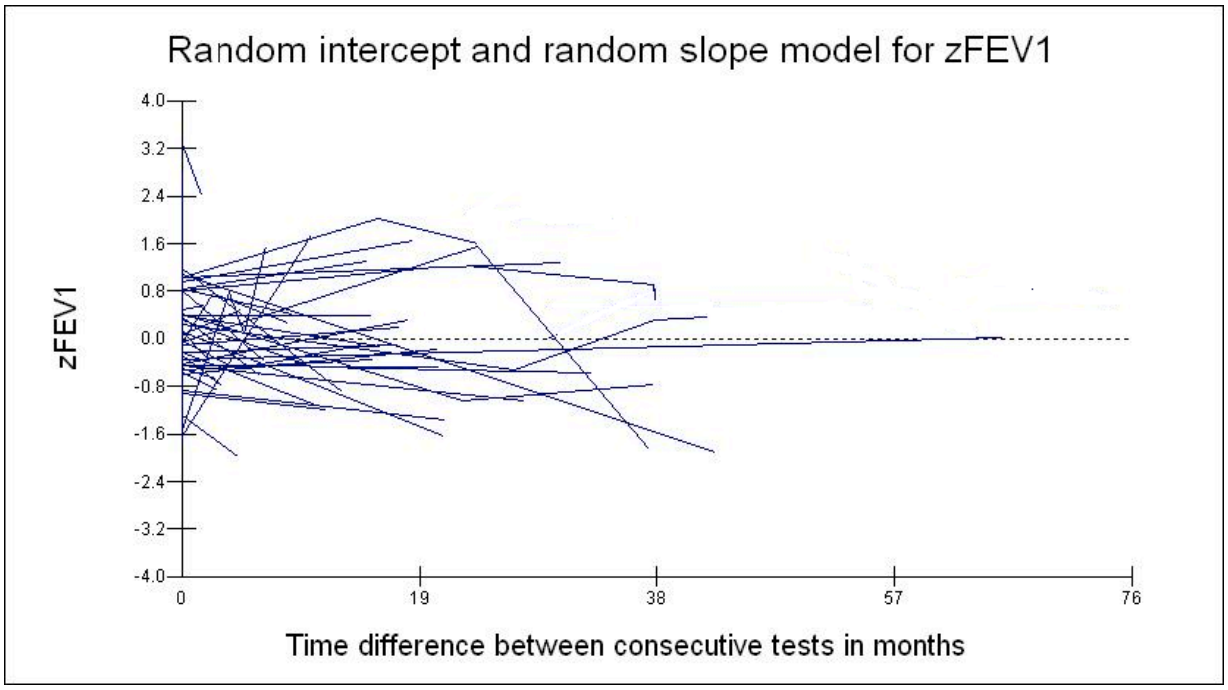


Figure 5.3: Random slope model predictions versus consecutive time difference between tests in months.

5.1.4 Z-score of Forced Expiratory Flow 25-75% ($zFEF_{25-75}$)

Table 5.4 shows results of the summary table of fitting a random slope model to $zFEF_{25-75}$.

	Parameter	Estimation(SE) (all instruments, n = 347)	(Jaeger, n = 149)
Fixed Effects			
Initial status	Intercept(β_0)	-0.280(0.066)*	-0.78(0.077)*
Rate of change	Slope(β_1)	0.004(0.004)	0.004(0.004)
Variance Comp.			
Level 1	Within Person(σ_e^2)	0.881(0.078)*	0.644(0.083)*
Level 2	Between Person(σ_0^2) In initial status	0.840(0.083)*	0.418(0.100)*
	In rate of change(σ_1^2)	-0.009(0.003)*	-
Goodness-of-fit			
IGLS Deviance			
$-2 \times$ loglikelihood		2084.365	915.49

Table 5.4: Results of fitting a multilevel model for change in z-scores of FEF_{25-75}, results in this table are presented as estimation and standard error in brackets. * represents statistical significance.

Interpretation of outcome:
1. When all observations measured by Jaeger and RASP are involved in the analysis, the overall intercept and slope estimations are -0.28 and 0.004 respectively. The intercept is significantly different from zero. However the estimated slope is not statistically significant. Similar to the rest of the analysis individuals show a significant variability in their FEF_{25-75} z-scores in their growth path with the estimated value of 0.88 with standard error 0.07. σ_0^2 the variance in intercept between individuals is estimated as 0.84 with standard error 0.13. Estimation of the between individual variance in rate of change (slope) is -0.009 with standard error 0.003. Both estimations for variance in intercept and slope between individuals show a significant difference between subjects.

2. Analogue to the first analysis when involving only observations measured by Jaeger, the estimated value for overall intercept and slope shows that intercept is statistically significant, however the estimated value for slope is not significant. The within individual

variance estimation is 0.64 with standard error 0.08 that means overall individuals have a significant change in their FEF_{25-75} z-scores from their infancy to school age. The between individuals variability in intercept is estimated as 0.41 with standard error 0.10 and the estimation of the between individual variance in rate of change (slope) is -0.009 with standard error 0.003.

Figure 5.4 presents the plot of predicted individual lines of the fitted model.

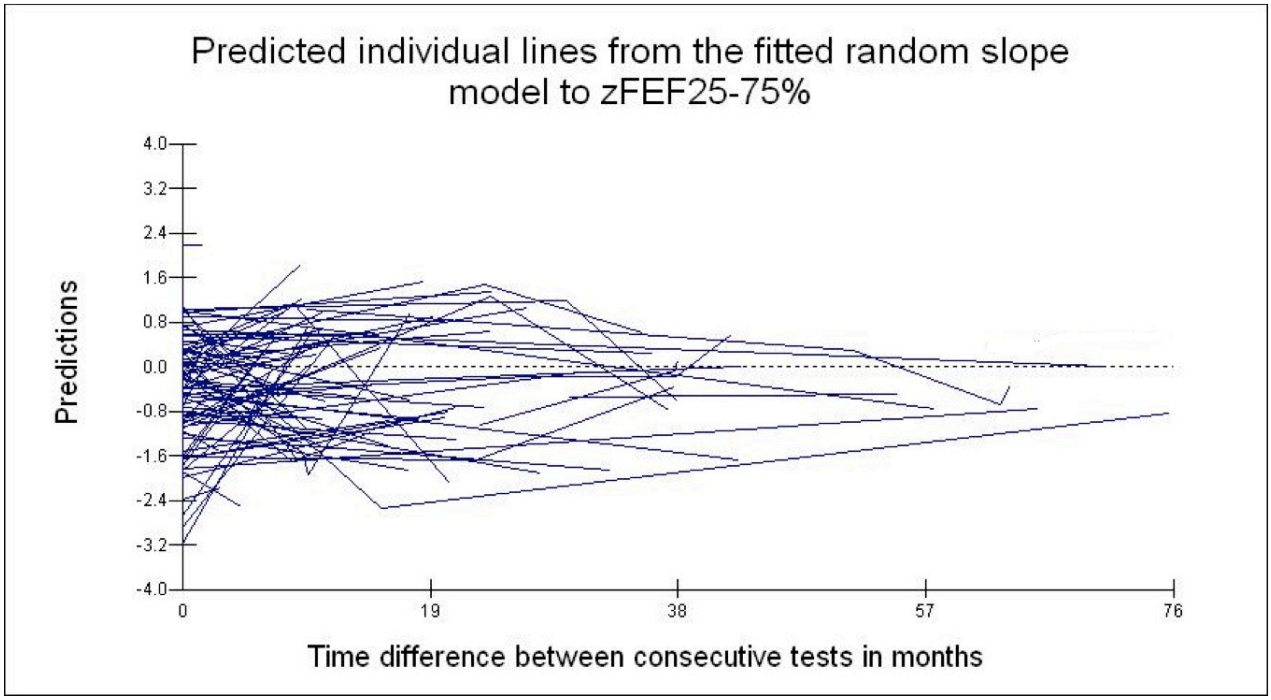


Figure 5.4: Random slope model predictions versus consecutive time difference between tests in months.

5.1.5 Z-score of Forced Vital Capacity ($zFVC$)

Table 5.5 shows results of summary table of fitting a random slope model to $zFVC$.

	Parameter	Estimation(SE) (all instruments, n = 348)	Estimation(SE) (Jaeger, n = 153)
Fixed Effects			
Initial status	Intercept(β_0)	-0.241(0.057)*	0.058(0.074)
Rate of change	Slope(β_1)	0.006(0.002)*	-0.001(0.002)
Variance Comp.			
Level 1	Within Person(σ^2_ϵ)	0.930(0.079)*	0.82(0.106)*
Level 2	Between Person(σ^2_0) In initial status	0.461(0.099)*	0.400(0.139)*
	In rate of change(σ^2_1)	-0.003(0.002)	-0.013(0.004)*
Goodness-of-fit			
IGLS deviance $-2 \times$ loglikelihood		2019.350	960.43

Table 5.5: Results of fitting multilevel model for change in z-scores of FVC, results in this table are presented as estimation and standard error in brackets. * represents statistical significance.

Interpretation of outcome:
1. Involving all observations measured by Jaeger and RASP in the analysis, the overall intercept estimation and slope are -0.24 and 0.006 respectively. σ^2_ϵ, the level one residual variance across all occasions of measurements is 0.93 with standard error of 0.07. σ^2_0 the variance in intercept between individuals is estimated as 0.46 with standard error 0.09. Estimation of the between individual variance in rate of change (slope) is -0.003 and standard error 0.002.

2. Having only observations measured by Jaeger in the analysis we get the estimation of overall intercept and slope as 0.06 and -0.001 respectively. The within individual variance estimation is 0.82 with standard error 0.11. The between individuals variability in intercept is estimated as 0.40 with standard error 0.14 and the estimation of the between individual variance in rate of change (slope) is -0.01 and standard error 0.004.

Figure 5.5 shows the plot of predicted individual lines of the fitted model.

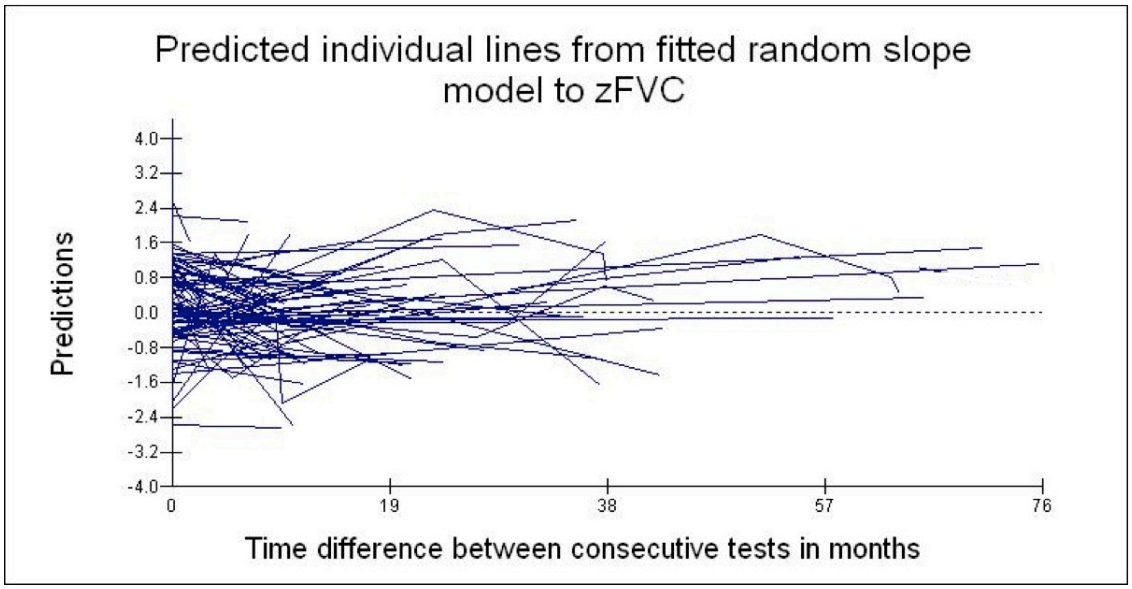


Figure 5.5: Random slope model predictions versus consecutive time difference between tests in months.

Chapter 6

Conclusions and Future Works

In this final chapter, we will conclude by describing the progress made towards the goals we described in abstract. We suggest some future research directions that could provide the next steps along the path to more precise and applicable techniques.

6.1 Conclusions

In this study, we investigated the change in Forced Expiratory Volumes from infancy to school age and also the within and between subjects variability. In order to accomplish this goal, we used only the measurements of white European descent babies as the Jones reference equations (2009) used in this study are not adjusted for other ethnics. However we adjusted the data to age, gender and height. In the next steps we cleaned the data to get all the healthy babies.

Plots of calculated z-scores of $FEV_{0.5}$ in infants, showed that the z-scores of $FEV's$ measured by the home-made instrument RASP are lower than z-scores of $FEV's$ measured by Jaeger. In this phase of study we decided to continue the research with two cases: 1. All measurements of both instruments were involved 2. Only measurements of Jaeger instrument involved.

We applied the paired t-test and Bland-Altman plots to this data in order to analyse the agreement between $FEV_{0.5}$ in infancy with $FEV_{0.75}$ and FEV_1 in preschool and also FEF_{25-75} in infancy to school age. The results of the paired t-test showed a significant change in mean between these two growth periods and the Bland-Altman method calculated the mean difference and its standard deviation. These analyses on this data set showed a significant variation from infancy to preschool or school age which make tracking the children from their infancy measures not applicable.

Additionally we were interested to find the intra subject variability for various respiratory parameters in healthy children. Applying multilevel regression analysis using `R` package "`lme4`" and also using the software "`MLwin`" we estimated the between and within individuals variation from their first measured FEV to their last measurements. The results of this analysis showed that individuals have a statistically significant variation within themselves and also there is also a significant change and difference between the individuals as well.

6.2 Future works

Following the investigations described in this thesis, a number of projects could be followed:

1. Collection of additional data from each individual: one severe limitation we were facing along this study was lack of enough data for each child to help us completely understand how the individuals change along their growth.

2. Further refinement of reference equations: since there are organizations such as Portex unit that are applying their own home-made instrument to measure the lungs forced expiratory flows specially in infancy, a research to find better reference equations adjusted for different instrument is a key step for all other spirometry studies that involve tracking the infants from their early ages to their preschool and school age.

3. Overcoming ethnic differences: a 3D approach to body physique for predicting lung function in children. This is a new on going research project at Portex unit. The study aims to determine which measurements of body physique, such as sitting height, chest depth, chest width or fat free mass, are the most predictive of lung function in children. This will help create reference ranges for paediatric spirometry that are based on body physique rather than ethnicity, allowing earlier diagnosis and treatment of lung disease in all children.

Appendix

Plot .1 to .8 are presenting the change in mean of $FEFV$ outcomes for different ethnicity. Only white children are studied in this thesis, due to lack of generalized reference equations applied to different ethnicities, an overall view of change in mean for different ethnicity would be beneficial for future research. Numbers on the lines show the size of population in each age or height group.

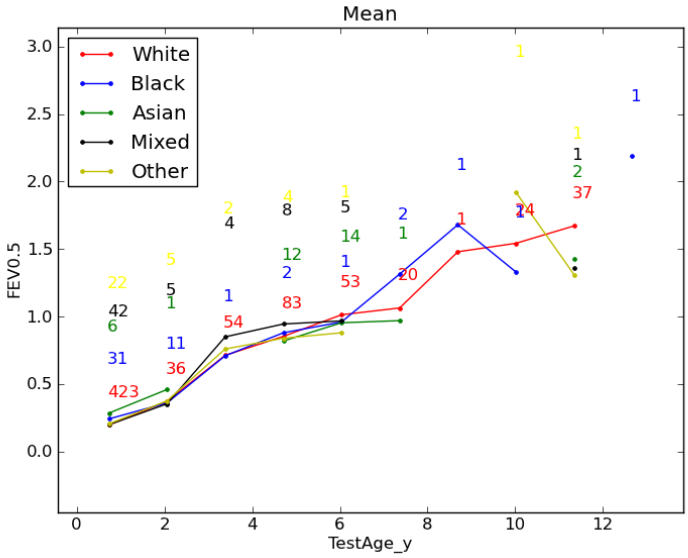


Figure .1: Mean plot of raw $FEV_{0.5}$ against test age in years for different ethnicity.

Plot .9 to .16 are presenting the change in mean of $FEFV$ outcomes for boys and girls against age and height. Numbers on the lines show the size of population in each age or height group.

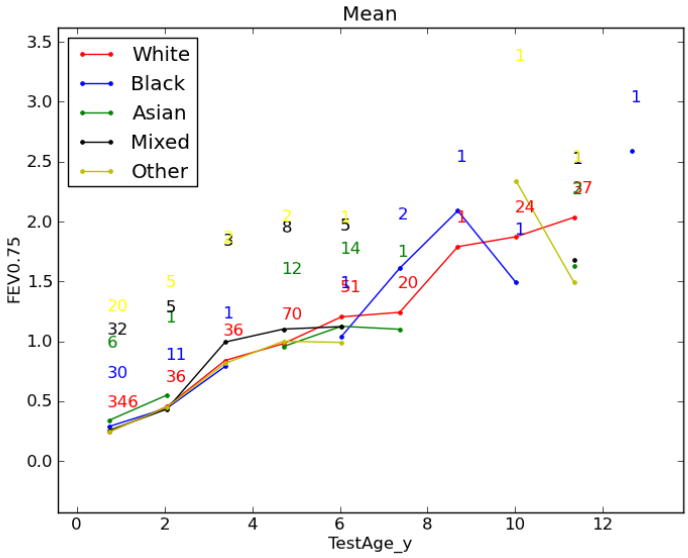


Figure .2: Mean plot of raw $FEV_{0.75}$ against test age in years for different ethnicity.

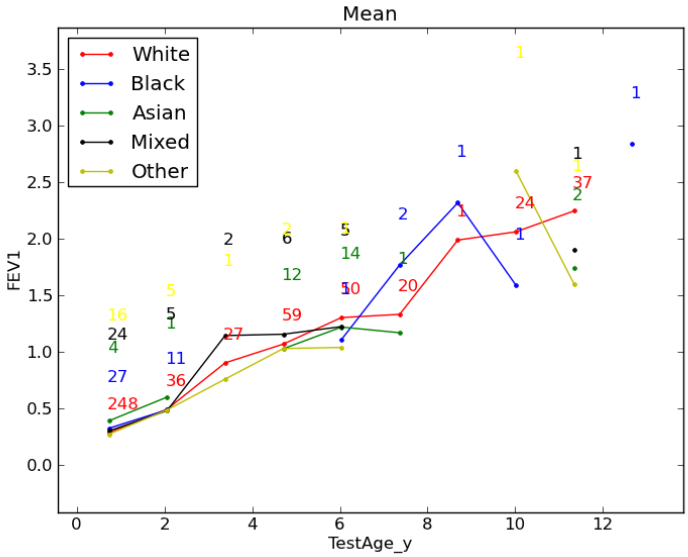


Figure .3: Mean plot of raw FEV_1 against test age in years for different ethnicity.

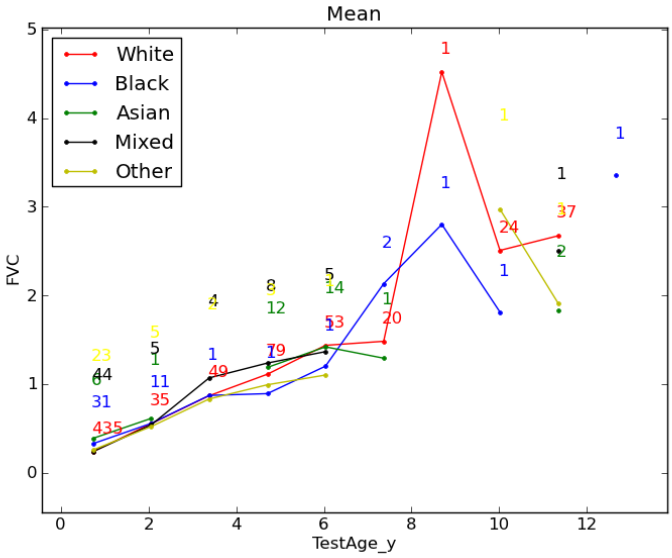


Figure .4: Mean plot of raw FVC against test age in years for different ethnicity.

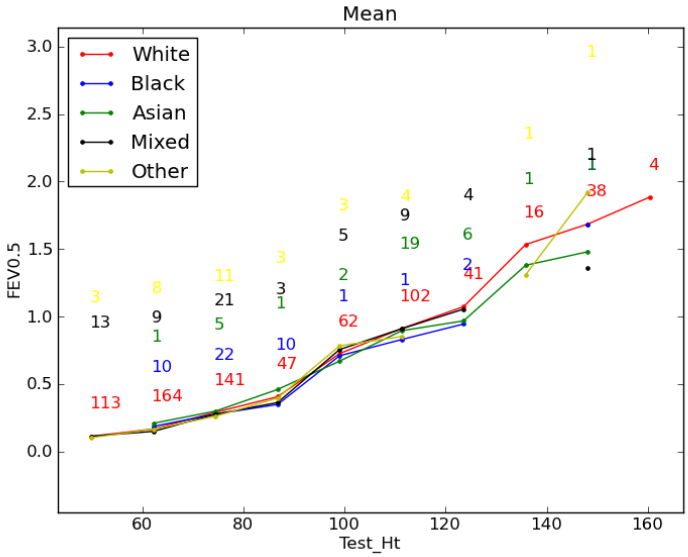


Figure .5: Mean plot of raw $FEV_{0.5}$ against test height in centimetres for different ethnicity.

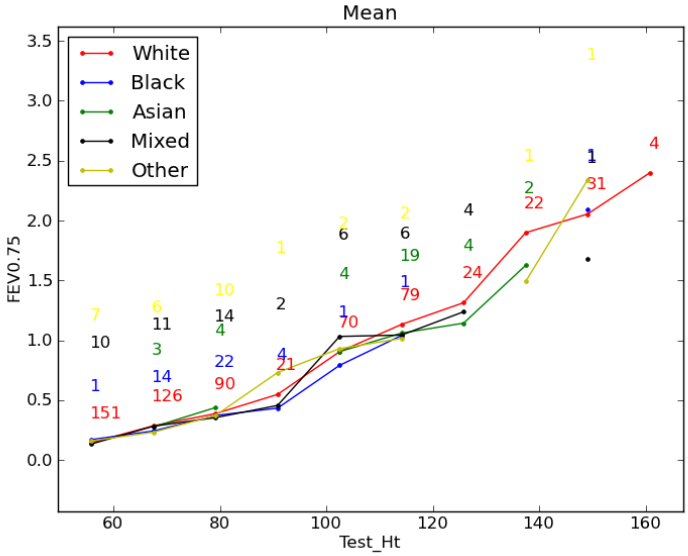


Figure .6: Mean plot of raw $FEV_{0.75}$ against test height in centimetres for different ethnicity.

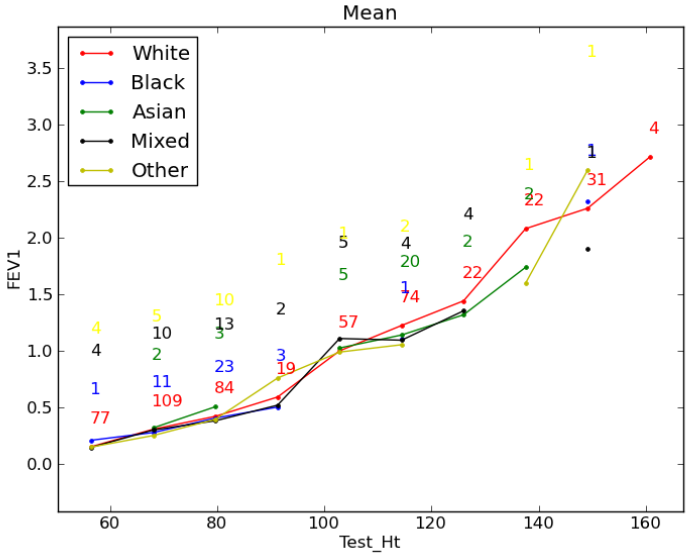


Figure .7: Mean plot of raw FEV_1 against test height in centimetres for different ethnicity.

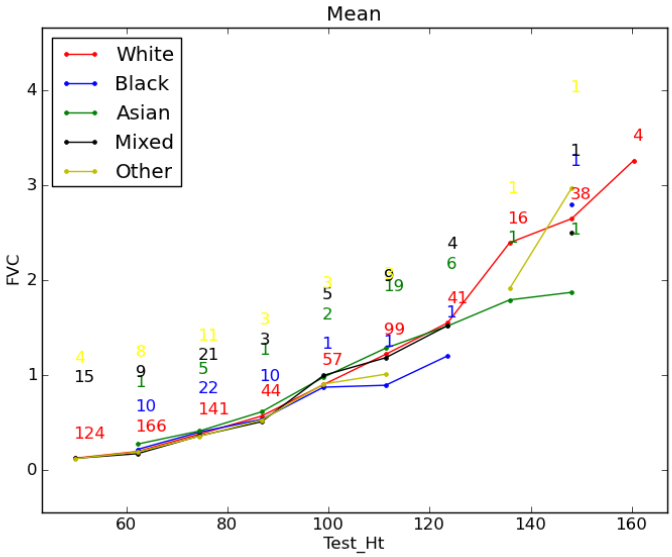


Figure .8: Mean plot of raw FVC against test height in centimetres for different ethnicity.

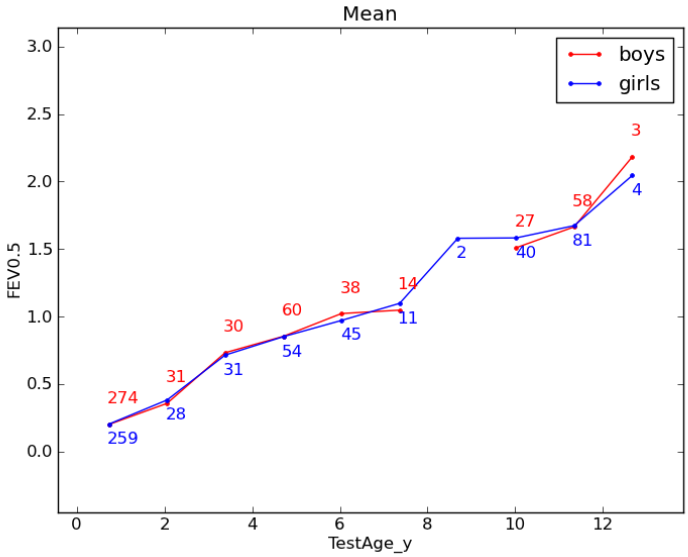


Figure .9: Mean plot of raw $FEV_{0.5}$ against test age in years for different gender.

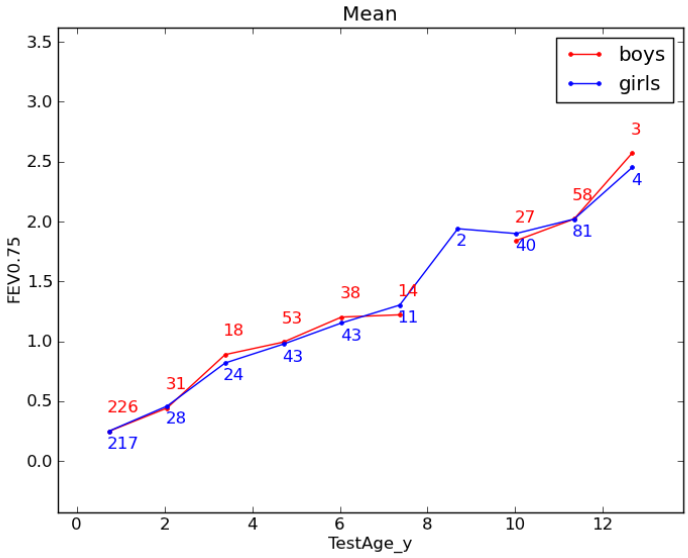


Figure .10: Mean plot of raw $FEV_{0.75}$ against test age in years for different gender.

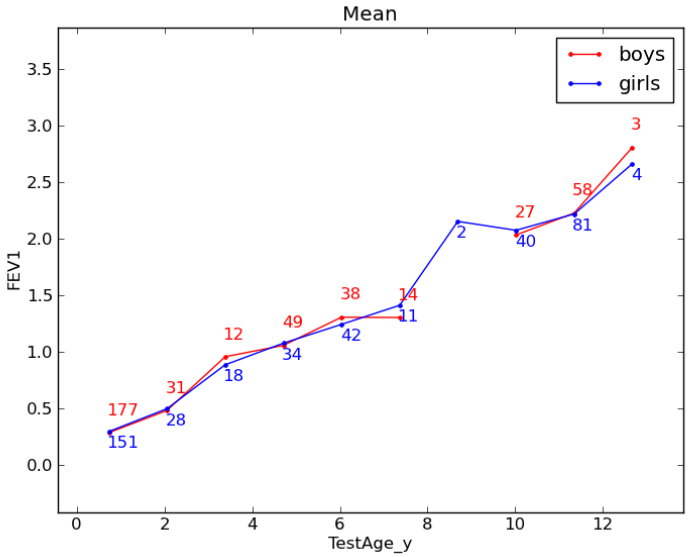


Figure .11: Mean plot of raw FEV_1 against test age in years for different gender.

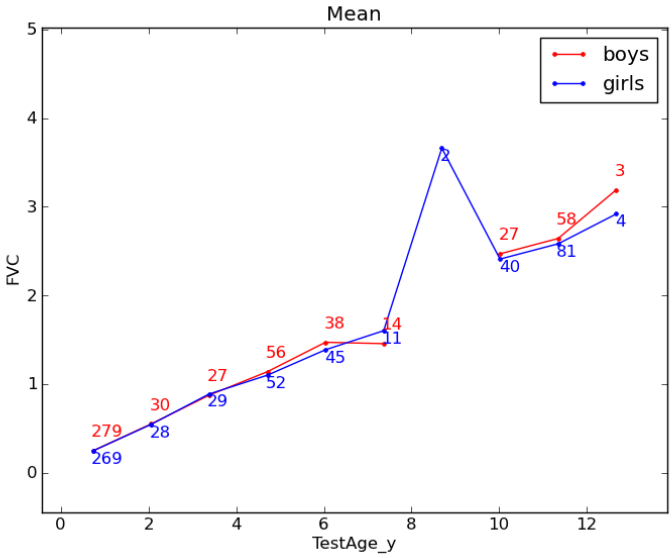


Figure .12: Mean plot of raw FVC against test age in years for different gender.

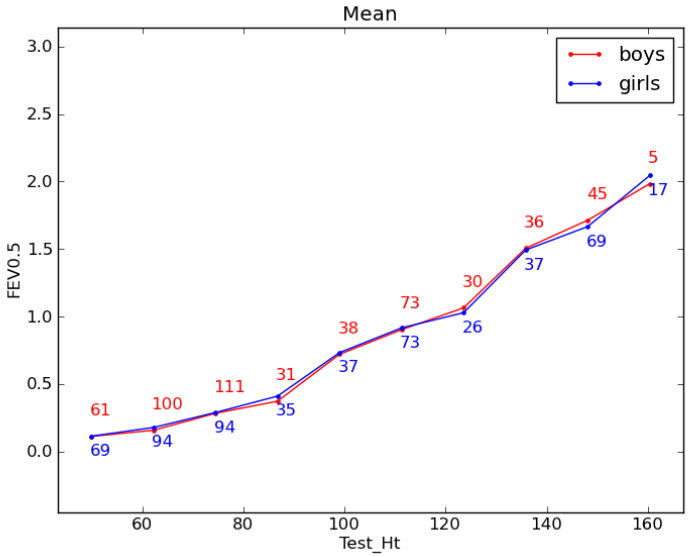


Figure .13: Mean plot of raw $FEV_{0.5}$ against test height in centimetres for different gender.

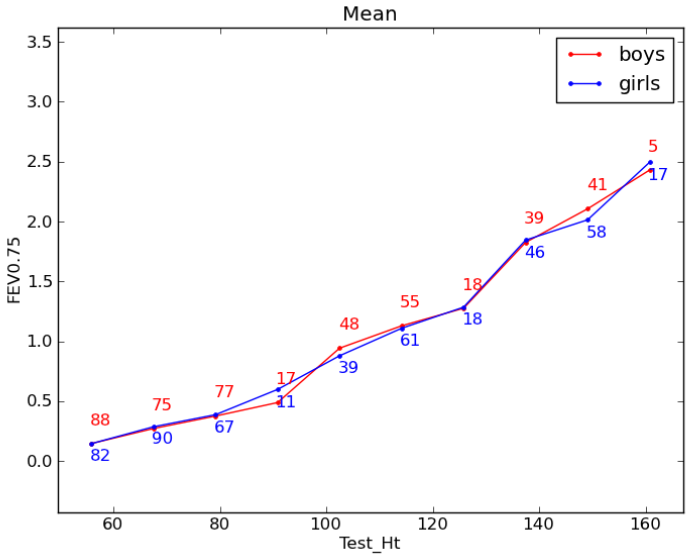


Figure .14: Mean plot of raw $FEV_{0.75}$ against test height in centimetres for different gender.

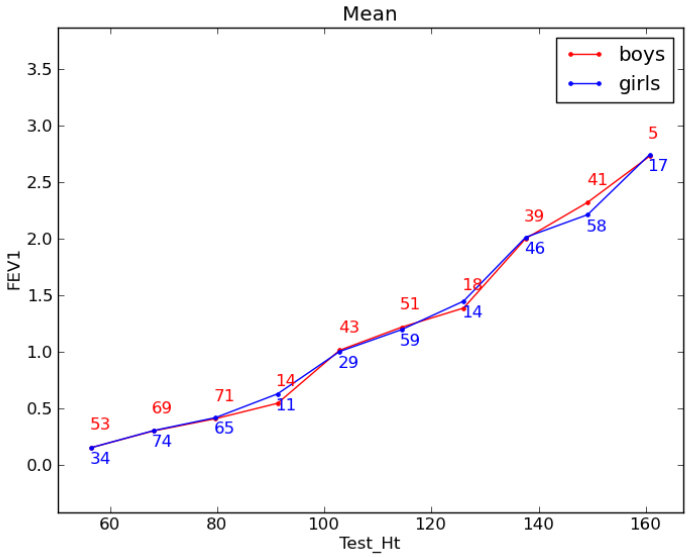


Figure .15: Mean plot of raw FEV_1 against test height in centimetres for different gender.

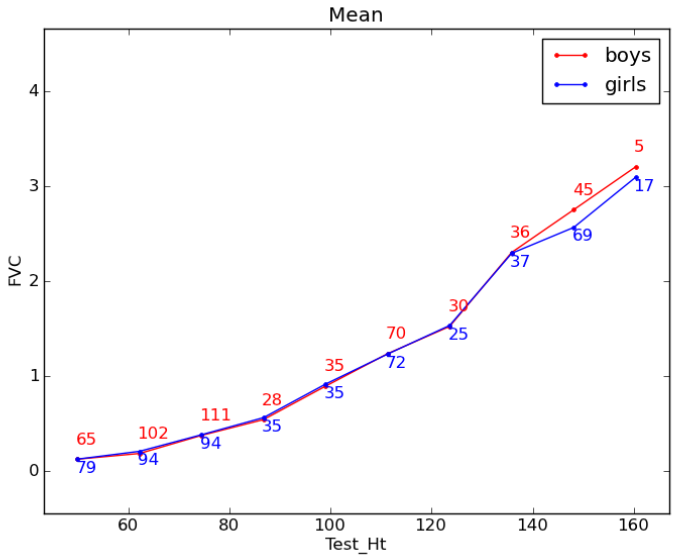


Figure .16: Mean plot of raw FVC against test height in centimetres for different gender.

Bibliography

[1] Kendig EL, Wilmott RW.
Kendig and Chernick's disorders of the respiratory tract in children.
8th ed. Philadelphia, PA: Saunders/Elsevier; 2012.

[2] Beydon N, Davis SD, Lombardi E, Allen JL, Arets HGM, Aurora P, et al.
An official American Thoracic Society/European Respiratory Society statement: pulmonary function testing in preschool children.
Am J Respir Crit Care Med. 2007 Jun;175(12):1304–45.

[3] Lum S, Stocks J.
Forced expiratory manoeuvres.
In: Paediatric Lung Function. vol. 47. European Respiratory Society Journals Ltd; 2010-03. p. 46–65.

[4] Cole TJ, Green PJ.
Smoothing reference centile curves: the LMS method and penalized likelihood.
Stat Med. 1992 Jul;11(10):1305–19.

[5] Verbeke G, Molenberghs G.
Linear mixed models for longitudinal data.
Springer Series in Statistics. New York: Springer; 2009.

[6] Michael Patetta BLOSJTDTCT Paul Marovich, Wolfinger R.
Longitudinal Data Analysis with Discrete and Continuous Responses.
SAS Institute Inc.; 2002.

[7] Fitzmaurice GM, Laird NM, Ware JH.
Applied Longitudinal Analysis.
Wiley Interscience; 2004.

[8] Laird NM, Ware JH.
Random-effects models for longitudinal data.
Biometrics. 1982 Dec;Vol. 38(No. 4):963–974.

[9] Searle SR, Casella G, McCulloch CE.
Variance components.
New York: Wiley; 1992.

[10] Longford NT.
Random coefficient models. vol. 11.
Oxford: Clarendon Press; 1993.

[11] West B, Welch KB, Galecki AT.
Linear mixed models: a practical guide using statistical software.
Chapman and Hall/CRC; 2007.

[12] Patterson TR H D.
Recovery of inter-block information when block sizes are unequal.
Biometrics. 1971;(58):545–554.

[13] Harville DA.
Bayesian inference for variance components using only error contrasts.
Biometrika. 1974;61(2):383–385.

[14] Verbeke G, Molenberghs G.
Linear mixed models for longitudinal data.
New York: Springer; 2000.

[15] Diggle P, Liang KY, Zeger SL.
Analysis of longitudinal data. vol. 13.
Oxford: Clarendon Press; 1994.

[16] Sprott DA.
Marginal and conditional sufficiency.
Biometrika. 1975;62(3):599–605.

[17] Harville DA.
Maximum Likelihood Approaches to Variance Component Estimation and to Related Problems.
JASA. 1977;72(358):320–338.

[18] Dempster AP, Laird NM, Rubin DB.
Maximum likelihood from incomplete data via the EM algorithm.
JRStatSoc B. 1977;39(1):1–38.

[19] Wolfinger R, Tobias R, Sall J.
Computing gaussian likelihoods and their derivatives for general linear mixed models.
SIAM JSciComp. 1994;15(6):1294–1310.

[20] Lum S, Kirkby J, Welsh L, Marlow N, Hennessy E, Stocks J.
Nature and severity of lung function abnormalities in extremely pre-term children at 11 years of age.
Eur Respir J. 2011 May;37(5):1199–207.

[21] Lum S, Hoo AF, Dezateux C, Goetz I, Wade A, DeRooy L, et al.
The association between birthweight, sex, and airway function in infants of nonsmoking mothers.
Am J Respir Crit Care Med. 2001 Dec;164(11):2078–84.

[22] Hankinson JL, Odencrantz JR, Fedan KB.
Spirometric reference values from a sample of the general U.S. population.
Am J Respir Crit Care Med. 1999 Jan;159(1):179–87.

[23] Stanojevic S, Wade A, Stocks J.
Reference values for lung function: past, present and future.
Eur Respir J. 2010 Jul;36(1):12–9.

[24] Stanojevic S, Wade A, Stocks J, Hankinson J, Coates AL, Pan H, et al.
Reference ranges for spirometry across all ages: a new approach.
Am J Respir Crit Care Med. 2008 Feb;177(3):253–60.

[25] Stanojevic S, Wade A, Cole TJ, Lum S, Custovic A, Silverman M, et al.
Spirometry centile charts for young Caucasian children: the Asthma UK Collaborative Initiative.
Am J Respir Crit Care Med. 2009 Sep;180(6):547–52.

[26] Lum S, Hoo AF, Hulskamp G, Wade A, Stocks J.
Potential misinterpretation of infant lung function unless prospective healthy controls are studied.
Pediatr Pulmonol. 2010 Sep;45(9):906–13.

[27] Hoo A, Dezateux C, Hanrahan J, Cole T, Tepper R, Stocks J.
Sex-specific prediction equations for V'maxFRC in infancy: a multi-center collaborative study.
Am J Respir Crit Care Med. 2002;165:1084–1092.

[28] Hulskamp G, Lum S, Hoo A, et al.
V'maxFRC—validation of a new commercially available equipment.
Eur Respir J. 2002 Sep;20:224s.

[29] Künzli N, Kuna-Dibbert B, Keidel D, Keller R, Brändli O, Schindler C, et al.
Longitudinal validity of spirometers-a challenge in longitudinal studies.
Swiss Med Wkly. 2005 Aug;135(33-34):503–8.

[30] Orfei L, Strachan DP, Rudnicka AR, Wadsworth MEJ.
Early influences on adult lung function in two national British cohorts.
Arch Dis Child. 2008 Jul;93(7):570–4.

[31] Aurora P, Stocks J, Oliver C, Saunders C, Castle R, Chaziparasidis G, et al.
Quality control for spirometry in preschool children with and without lung disease.
Am J Respir Crit Care Med. 2004 May;169(10):1152–9.

[32] Piccioni P, Borraccino A, Forneris MP, Migliore E, Carena C, Bignamini E, et al.
Reference values of Forced Expiratory Volumes and pulmonary flows in 3-6 year children: a cross-sectional study.
Respir Res. 2007;8:14.

[33] Bland JM, Altman DG.
Statistical methods for assessing agreement between two methods of clinical measurement.
Lancet. 1986 Feb;1(8476):307–10.

[34] Motulsky H.

Prism 4 Statistics Guide −Statistical analyses for laboratory and clinical researchers. vol. 4.0.
GraphPad Software; February 2005.