Comparison of growth curve models for assessing height in a South African birth cohort

Jacqui Niehaus

Supervised by Francesca Little

MSc Biostatistics

2018

Department of Statistical Sciences

University of Cape Town

Contents

List of Figures

List of Tables

1 Acknowledgements

I thank the Drakenstein Child Health Study for the use of data used in this dissertation. In particular, the Drakenstein clinical and research team for their hard work and commitment, the data team, the lab teams, and the administrative staff at the Western Cape Government Health Department for their support of the study. I especially thank the families who participated in the study. I also thank the Bill & Melinda Gates Foundation (OPP1017641), South African Medical Research Council, National Research Foundation South Africa, National Institute of Health, and H3Africa (1U01AI110466-01A1) for their support.

I would like to extend my sincere appreciation to my supervisor, Francesca Little, for her continued guidance and support for this dissertation. The insightful reviews and comments have been invaluable in my learning experience during this process.

2 Abstract

Childhood malnutrition is a major concern in low- to middle- income populations. This dissertation uses longitudinal data on height measurements of babies between 0 and 4 years of age to construct growth curves, which serve as a tool for assessing the health and nutritional progress of children. We wish to characterise the way height changes over time and identify predictors of that change. Various mixed effect models were fit and compared to neural networks in terms of model fit, interpretability of parameters as well as predictive power. The best fitting mixed-effect model was the Berkey-Reed 2nd order model. The neural network compared well with this model, indicating that neural networks may serve as a useful alternative to modelling longitudinal growth data.

Subsequently, logistic regression was used to explain the relationship between various pre- and post-natal risk factors for stunting, a shortfall in height relative to age. The results were compared to a random forest model. Methods for variable importance in classification problems using tree-based methods were explored. The random forest model appeared to perform similarly to the logistic regression model in terms of predictive power and variable interpretation.

This dissertation contributes in investigating the possibility of using machine learning techniques to identify probable correlates of childhood malnutrition.

3 Introduction

3.1 Objective

The objective of this dissertation is to:

- Apply mixed effect modelling to longitudinal measurements of children's height in order to:
  - Compare methods for estimating growth curves that best fit the measurements in this cohort
  - Identify predictors of the growth trajectory

- Explore the relationship between stunting and various pre- and post-natal risk factors
- Compare traditional statistical techniques with machine learning algorithms in the child growth setting, namely:
  - Investigate the potential of using an artificial neural network as an alternative modelling strategy to mixed effects models in assessing and predicting longitudinal growth data.
  - Investigate the potential of using random forests as an alternate classification model to logistic regression in assessing and predicting stunting.

3.2 Motivation

According to R. E. Black et al. (2013), the prevalence of childhood malnutrition has increased in low- and middle-income countries. Stunting is of particular prevalence in under-nourished populations. Stunting is defined by the World Health Organization (2010) as a shortfall in height relative to the child's age. Akombi et al. (2017) showed that in Africa, stunting rates increased by 24% between 1990 and 2012. Stunting is an indication that a child is failing to develop. It often occurs due to a child being subjected to continual malnutrition at early stages of their lives. This malnutrition can begin in the womb, due to the mother not receiving adequate nutrition. She is therefore unable to sustain her child's development while pregnant. Stunting can then continue after birth as a result of poor feeding practices, infections and environmental factors.

Akombi et al. (2017) show that stunting mostly occurs during the first 24 months following birth, continuing to 5 years of age. According to Akombi et al. (2017), risks associated with chronic malnutrition include reduced brain development resulting in learning difficulties later in life, weakened immune systems resulting in more illness and higher chances of serious diseases later in life, such as diabetes and cancer.

Growth curve assessment can be used as a means for assessing the health and nutritional progress of children. The ability to predict abnormal height trajectories in children early on will allow preventative measures to be taken. This kind of surveillance also provides a way to indirectly measure the quality of life of an entire population since covariates of stunting are related to standards of living.

3.3 Background

Longitudinal data are useful in providing information about individual changes. By collecting data over many time points, changes over time in individual children can be separated from differences between children. However, W. Johnson, Balakrishna, and Griffiths (2014) shows that special models are needed for such data due to the non-independence of responses, heteroscedasticity and serial correlation. Considering the effect of nutritional inputs on the health of a child, we would expect each child to improve their health by the addition of certain nutritional inputs. However the growth rate would differ from child to child. Furthermore there may be differences in the point at which a gradual increase in health would occur for any additional nutritional inputs. These differences in growth rates between any two children is the subject-specific effect.

Mixed effect models are considered the most appropriate statistical methodology for analysing longitudinal data in that they take account of within-subject association through the addition of subject-specific random effects. The fixed-effect component represents the population growth curve and the random-effect component allows for individual variation around the population curve. Various popular growth models are fitted and compared in this dissertation. These include non-structural and structural models. Non-structural models, mainly polynomials, do not assume a specific form of the growth curve. They often fit the data very well, however their parameters cannot be interpreted. Structural models imply a functional form of the growth curve and allow for some biologically interpretable parameters. Models that have been developed to model child growth include the Count model, Berkey-Reeds model, Jenss-Bayley model and Karlberg model.

Artificial Neural Networks (ANNs) have been shown to fit better models than traditional statistical techniques in many cases. The advantages of ANNs are that they easily recognise patterns in the data, they can be fitted to any kind of dataset, they do not require any model assumptions and the model is adaptively formed based on features present in the data.

Recent studies suggest that ANNs may also be useful in modelling longitudinal data, specifically where the actual relationship between the variables is not known or cannot be specified. Neural networks allow for a flexible nonlinear model to be developed with little knowledge about the actual relationship that exists between the variables. The

parameters of the model (weights) are estimated by the neural network learning from the data. This differs from the mixed effect model where the form of the model has to be specified. The following approaches have been investigated to model longitudinal data using ANNs in previous studies:

1) Using a time variable as input along with other covariates. This method is used by Ganesan et al. (2014), where LMEs and ANNs are compared for modelling the growth data of sheep. The time variable is rescaled such that each month is associated with a numeric value between 0 and 1. This produces a month-specific intercept for time. Each observation is fed into the network independently, i.e. there is no allowance for the longitudinal nature of the data.

2) Tandon, Adak, and Kaye (2006) use a 'mixed-effect neural network' that uses an unspecified nonlinear function of the input variables to predict the course of disease in Alzheimer's patients. In this method, the neural network is used only to estimate the nonlinear function that relates the covariates to the outcome. The network just has the input layer and output layer (no hidden layers) and then a random subject effect is added. The ANN is not allowed to learn automatically from the data.

3) Crane-Droesch (2017) created semiparametric panel data models using Neural Networks. This method allows the neural network to fit the nonparametric component of the model only.

4 Data

4.1 Data description

The Drakenstein Child Health Study (DCHS) is a population birth study in the Drakenstein area in Paarl, a peri-urban area, 60km outside Cape Town, South Africa (H. J. Zar et al. 2015). Mothers were enrolled antenatally and the mother-baby pairs were followed until 5 years of age. Detailed socio-demographic, nutritional and psychosocial data were collected at multiple time points antenatally and postnatally.

The outcome variable of interest in this dissertation is height/length. Measurements were taken using a Seca length-measuring mat (Seca, Hamburg, Germany) on a firm surface by two staff members. Children less than 2 years of age were measured by lying down on the board (recumbent length), and children over 2 years of age were measured while standing up. All measurement were taken twice per child to improve accuracy. The time points used are birth (0 months), 6 weeks (1.5 months), 10 weeks (2.5 months), 14 weeks (3.5 months), 6 months, 9 months, 12 months, 24 months, 30 months, 36 months, 42 months and 48 months. Too few children in the study had reached 54 and 60 months at the time of this analysis for these time points to be considered.

Height measurements were converted to height-for-age z-scores using the WHO reference ranges. Growth reference ranges are used during anthropometric assessment of children to evaluate the growth, nutritional status and well-being of children (World Health Organization 2010). A growth standard indicates the optimal growth and the growth reference is the distribution used for comparison. The percentiles of this distribution are used to classify sex-age specific anthropometric measures. The z-score is the number of standard deviations away from the mean of the WHO reference population, when the distribution is normal. Stunting is defined as having a height-for-age z-score of less than or equal to -2. Children with a z-score of less than -3 are classified as having severe stunting.

The following measures were taken to clean the data before analysis:

- According to the WHO anthroplus manual, z-scores falling outside of the interval (-6, 6) are biologically implausable. Such scores were therefore seen as data errors and removed from the analyses.

- Non-increasing height measurements between successive measurements were seen as data errors and removed.
- Only children with at least 3 height measurements at different occasions were included, to ensure a meaningful growth trajectory for improved model covergence.

4.2 Data summary

There were 1157 children in the study in total. After data cleaning, 1024 are left for the analysis. Table 4.1 shows the number of measurements per age. We can see that the data is unbalanced, i.e. some children did not attend every visit. Not all children in the study have reached 48 months yet. An advantage of mixed-effect models is that they do not require data measurements to to be at the same time point for each individual nor that all individuals need to have complete data.

Table 4.1: Number of observations per age

age	0	1.5	2.5	3.5	6	9	12	18	24	30	36	42	48
	968	588	631	807	843	764	705	753	673	646	529	482	378

The following variables will be investigated in this dissertation to determine their association with the growth model:

- Numeric covariates
  - Maternal age (years)
  - Maternal height (cm)
  - Gestational age (weeks). This is the duration of the pregenancy in weeks at which the baby is born.
- Categorical covariates
  - Sex (male versus female)
  - Clinic. TC Newman services a predominantly mixed race community and Mbekweni services a predominantly Black-African community.
  - Maternal HIV status (Yes/No)
  - Tobacco during pregnancy (Yes/No)
  - Alcohol during pregnancy (Yes/No)
  - Socioeconomic status (SES). This was measured using a composite score

of four factors: assets and market access (a score based on the number of specific assets owned as well as access to retail and financial facilities), household income (a score based on the total amount of combined household income per month), employment status (dichotomise variable of employed or unemployed), and educational achievement (a score based on highest scholastic level). This was adapted from the South African Stress and Health Study. The combined SES score was calculated by summing the scores of each of the four variables. Participants were then split into four levels based on their SES score (low, low-to-moderate, moderate-to-high, or high SES) (Budree et al. 2017).

- Maternal marital status (Single, Not married but in a married like relationship, married, divorced, widowed)
- Initiated breastfeeding (Yes/No). This is an indicator variable to identify whether breastfeeding was initiated or not

Table 4.2 shows a summary of the numeric covariates. We can see that some covariate values are missing from the data. Mothers are between 17 and 45 years of age, with a mean age of 26.77. Maternal height ranges between 134cm and 189cm. Gestational age ranges between 26 weeks and 48 weeks. A normal gestational age is considered to be between 38 and 42 weeks (World Health Organization 2010).

Table 4.2: Numeric covariate summary

	N	Missing	Mean	SD	Min	Q1	Median	Q3	Max
Maternal age (years)	965	59	26.77	5.75	17.71	22.08	25.97	30.84	44.84
Maternal height (cm)	952	72	159.3	6.81	134	155	159	164	189
Gestational age (weeks)	1020	4	38.52	2.64	26	38	39	40	48

Table 4.3 shows the number and percentage in each class of the categorical covariates. We can see that there are slightly more male babies in the study. A larger proportion of babies were born at Mbekweni. The mother's data shows that 20.8% had HIV, 23.3% had tobacco during pregnancy and 14.3% had alcohol during pregnancy. There is quite an even distribution of socio-economic quartiles present in the study. Most mothers in the study were single (never married). A largr proportion (92.6%) of mothers initiated breastfeeding. Almost a quarter (24.7%) of babies were born pre-maturely, while 3.1% were born post-maturely.

Table 4.3: Factor covariate summary

	Level	N	%
Sex	Female	491	47.9
	Male	533	52.1
Clinic	TC Newman	464	45.3
	Mbekweni	560	54.7
Maternal HIV status	Yes	213	20.8
	No	752	73.4
	Missing	59	5.8
Tobacco during pregnancy	Yes	239	23.3
	No	618	60.4
	Missing	167	16.3
Alcohol during pregnancy	Yes	146	14.3
	No	711	69.4
	Missing	167	16.3
Socio economic quartile	Low	249	24.3
	Low to moderate	268	26.2
	Moderate to high	263	25.7
	High	243	23.7
	Missing	1	0.1
Maternal marital status	Single (never married)	605	59.1
	Not married, but in a marriage like relationship	187	18.3
	Married	224	21.9
	Divorced	3	0.3

	Level	N	%
	Widowed	3	0.3
	Missing	2	0.2
Initiated breastfeeding	Yes	948	92.6
	No	76	7.4
Gestation status	Premature (<38 weeks)	253	24.7
	Normal (38-42 weeks)	735	71.8
	Postmature (>42 weeks)	32	3.1
	Missing	4	0.4

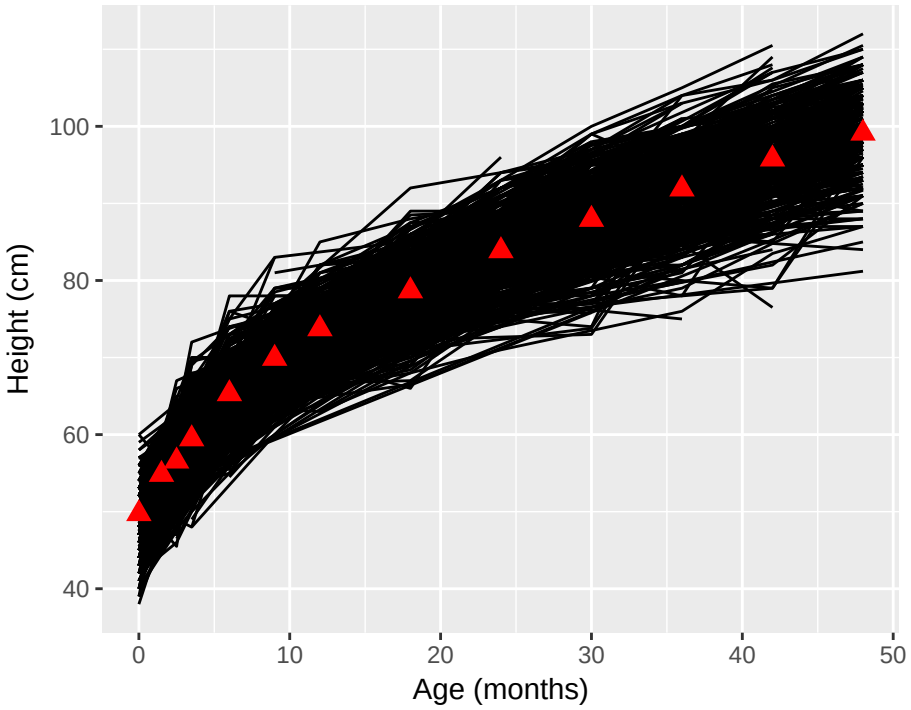


Figure 4.1: Individual height profiles

4.3 Data visualisation

Figure 4.1 shows the individual height profiles of the children. The triangles indicate the mean height at each age. Children seem to vary by birth height as well as growth rate. The growth seems to increase more rapidly during the early months, with the rate of growth slowing down as age increases. Therefore a model that captures the acceleration and deceleration in growth will be required. The variance across time is also seen to be non-constant and increasing as the time increases.

Figure 4.2 shows a boxplot of the height distribution over time. The plot confirms that the variation in height increases with increasing age, as the boxes become bigger and the whiskers longer with age. There are more outliers at the shorter end of the height measurements than the taller end. This indicates a greater proportion of abnormally short babies. However, excluding outliers, the height distribution does not appear to be skewed in any direction.

Plotting the height-for-age z-scores over time provides insight into any abnormal growth trajectories. Figure 4.3 shows the individual z-score profiles. Values falling near zero

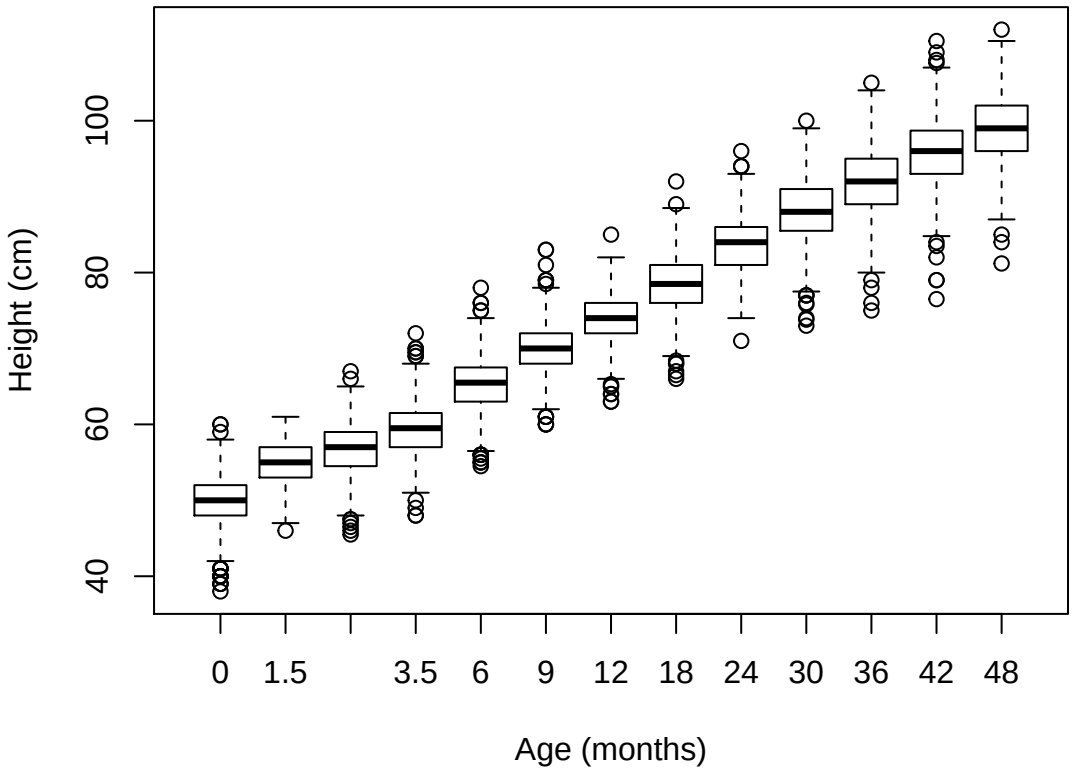


Figure 4.2: Height distribution

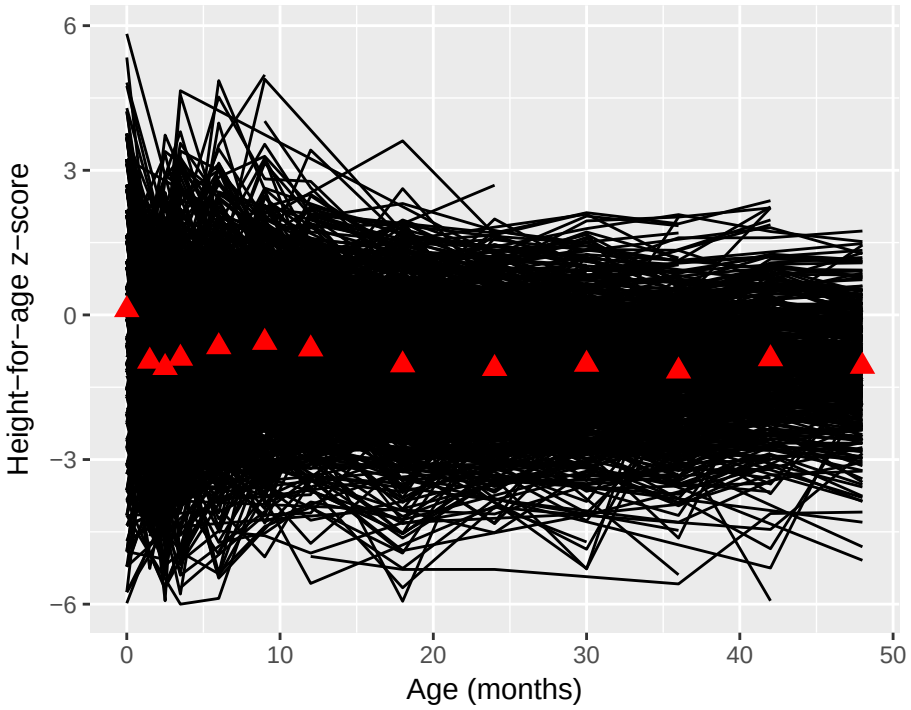


Figure 4.3: Individual Height-for-age z-score profiles

indicate normal height-for-age according to the WHO growth reference standards. The mean profile, indicated by the triangles, indicates that the average height of the children is consistently below the WHO reference population mean. There are quite a few children whose z-score falls consistently below -2, indicating that they are stunted. Some children seem to recover from a stunting period, returning to a normal height-for-age.

Figure 4.4 shows the height-for-age z-score distribution over time. There is greater variation in the z-scores at birth. The median height and inter-quartile range are less than zero post birth, indicating that the children are generally short in comparison to the WHO standards. There are many outliers at the low end of the z-score range, indicating that there are several cases of stunting in the population. The inter-quartile range reduces with age, indicating that differences in heights amongst the population decreases with age.

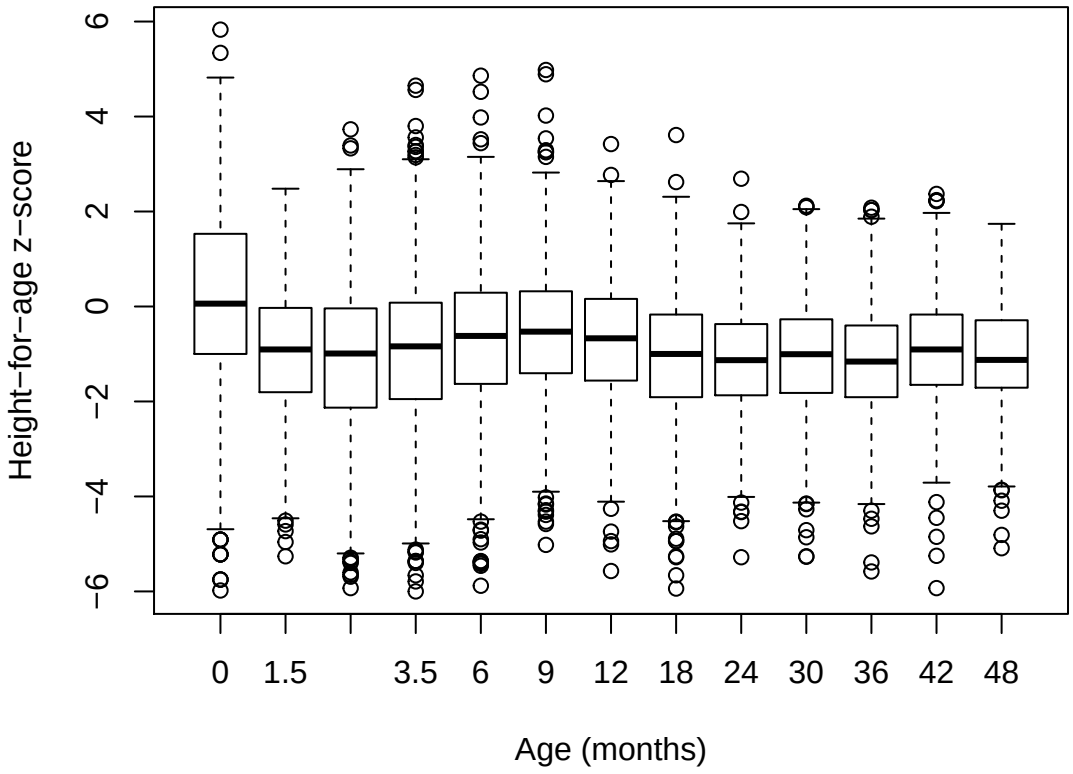


Figure 4.4: Height-for-age z-score distribution

5 Mixed-effect Models

5.1 Model fitting strategy

The following models (in increasing order of complexity) were progessively fit and assessed. Unconditional models refer to models in which no covariates are used, i.e. height is a function of age only. Conditional models refer to models in which covariates were introduced into the model.

- Unconditional growth models
  - Linear change over time
  - Non-linear change over time
- Conditional growth models
  - Effect of constraints on the intercept
  - Effect of constraints on the slope
  - Effect of constraints on higher order terms or other shape parameters
- Allow for different within-child error covariance structures

The types of models fit were as follows:

1) Linear model:

$$y_{ij} = \beta_0 + \beta_1 t_{ij}$$

2) Quadratic model:

$$y_{ij} = \beta_0 + \beta_1 t_{ij} + \beta_2 t_{ij}^2$$

3) Cubic model:

$$y_{ij} = \beta_0 + \beta_1 t_{ij} + \beta_2 t_{ij}^2 + \beta_3 t_{ij}^3$$

4) Fractional Polynomial model:

$$y_{ij} = \beta_0 + \beta_1 log(t_{ij} + 2) + \beta_2 (t_{ij} + 2)^{0.5}$$

5) Count model:

$$y_{ij} = \beta_0 + \beta_1 t_{ij} + \beta_2 ln(t_{ij} + 1)$$

6.1) Berkey-Reeds 1st order (Reed1) model:

$$y_{ij} = \beta_0 + \beta_1 t_{ij} + \beta_2 ln(t_{ij}) + \beta_3 \frac{1}{t_{ij}}$$

6.2) Berkey-Reeds 2nd order (Reed2) model:

$$y_{ij} = \beta_0 + \beta_1 t_{ij} + \beta_2 ln(t_{ij}) + \beta_3 \frac{1}{t_{ij}} + \beta_4 \frac{1}{t_{ij}^2}$$

7) Logistic growth model:

$$y_{ij} = \frac{\phi_{1i}}{1 + exp[-(t_{ij} - \phi_{2i})/\phi_{3i}]}$$

8) Gompertz growth model:

$$y_{ij} = \alpha exp(\beta(1 - exp(-\mu t_{ij})))$$

9) Exponential growth model:

$$y_i = \beta_0 exp(\beta_1 t_i)$$

10) Jenss-Bayley model:

$$y_{ij} = \beta_0 + \beta_1 t_{ij} - exp(\beta_2 + \beta_3 t_{ij})$$

11) Karlberg growth model:

$$y_i = \beta_0 + \beta_1[1 - exp(-\beta_2 t_{ij})]$$

where

y_{ij} represents height of child i at measurement occasion j and t_{ij} represents age of the child i at measurement occasion j.

$t_{ij} = \{0, 1.5, 2.5, 3.5, 6, 9, 12, 18, 24, 30, 36, 42, 48\}$ represents the age in months at which height measurements were taken.

These models were chosen since we require a function that captures the rapid growth following birth and then subsequent deceleration in growth.

The polynomial models (1-4) do not assume a specific form of the growth curve. Increasing the order of the polynomial usually fits the data very well however the parameters cannot be interpreted. The structural models (5-11) imply a functional form and the parameters allow some biological interpretation.

5.2 Linear Mixed effects models

The first six models are all linear in the parameters and were therefore fit to height measurements using linear mixed effects (LME) models . The general structure of an LME is given by:

$$\boldsymbol{y_i} = \boldsymbol{X_i}\boldsymbol{\beta} + \boldsymbol{Z_i}\boldsymbol{b_i} + \boldsymbol{\epsilon_i}$$

where

$\boldsymbol{y_i}$ is the $\boldsymbol{n_i} \times \mathbf{1}$ vector of observed heights

n_i is the number of measurements taken for the ith child

$i = 1, ...N$, where $N = \sum_i n_i$

$\boldsymbol{X_i}$ is the $\boldsymbol{n_i} \times \boldsymbol{p}$ matrix of the fixed effects

$\boldsymbol{\beta}$ is the $\boldsymbol{p} \times \mathbf{1}$ vector of the coefficients

$\boldsymbol{Z_i}$ is the $\boldsymbol{n_i} \times \boldsymbol{q}$ matrix of the random effects

$\boldsymbol{b_i} \sim \boldsymbol{N(0, \Psi)}$ is the $\boldsymbol{q} \times \mathbf{1}$ subject effect

$\boldsymbol{\epsilon_i} \sim \boldsymbol{N(0, \sigma_\epsilon^2 I_n)}$ is the $\boldsymbol{n_i} \times \mathbf{1}$ vector of random errors

The random components were systematically added to the fixed effects in each model. We start with a random intercept since we can see from the trajectory plots that there is a subject-to-subject variability for the birth height. We then incorporate a random effect on the slope and higher order growth terms, testing for significance at each stage.

5.2.1 Linear growth model

This is the baseline model that allows for individual variation in intercept and growth rates. It has been used by Grimm, Ram, and Hamagami (2010) to model height between 3 and 19 years of age. The parameters of the linear growth model can be easily interpreted. The intercept refers to birth height. The linear slope represents the average rate of change over the observation period. A limitation of this model is that if height is projected to future ages outside the observation period, individual trajectories will continue towards positive infinity. It also does not capture the change in growth rate as age increases. Grimm, Ram, and Hamagami (2010) showed that the linear growth model may only be valid for modelling growth over small age ranges.

Let y_{ij} be the height for child i at time j, then

$y_{ij} = \beta_{0i} + \beta_{1i} t_{ij} + \epsilon_{ij}$

where $\beta_{0i} = \beta_0 + b_{0i}$ and $\beta_{1i} = \beta_1 + b_{1i}$

The fixed-effects regressor matrix is given as:

$$\boldsymbol{X_i} = \begin{bmatrix} 1 & t_0 \\ 1 & t_1 \\ 1 & t_2 \\ \vdots & \vdots \\ 1 & t_{13} \end{bmatrix}$$

$\boldsymbol{\beta} = \begin{pmatrix} \beta_0 \\ \beta_1 \end{pmatrix}$ with β_0 is related to the baseline height at birth and β_1 is related to the slope (linear growth velocity), $\boldsymbol{b_i} = \begin{pmatrix} b_{0i} \\ b_{1i} \end{pmatrix}$, $\boldsymbol{b_i} \sim \boldsymbol{N(0, \Psi)}$ where Ψ is the variance-covariance matrix of $\boldsymbol{b}$ and $\boldsymbol{\epsilon_i} \sim N(0, \sigma^2 \boldsymbol{I})$ where σ^2 is the overal residual variance.

This model determines a line of best fit between height and age. We would expect this to provide a poor fit for the data since we know that child growth does not occur at a constant rate.

5.2.1.1 Random intercept only:

If we start with just a random intercept and constant slope, $\boldsymbol{b_i} \sim N(0, \sigma_b^2)$ is a one-dimensional random-effects vector describing the shift in intercept for each subject. Because there is a common growth rate, these shifts are constant preserved for all values of age. The matrix $\boldsymbol{\Psi} = \sigma_b^2$ will be a 1x1 matrix in this case. It represents the variance of the measurements in the population at a fixed value of age. The random-effects regressor matrix is given as:

$$Z_i = \begin{bmatrix} 1 \\ 1 \\ 1 \\ \vdots \\ 1 \end{bmatrix}$$

5.2.1.2 Random intercept and slope:

Adding a random slope parameter, the random-effects regressor matrix is given as:

$$Z_i = \begin{bmatrix} 1 & t_0 \\ 1 & t_1 \\ 1 & t_2 \\ \vdots & \vdots \\ 1 & t_{13} \end{bmatrix} \; and \; \Psi = \begin{bmatrix} \sigma_{b0}^2 & \sigma_{b01}^2 \\ \sigma_{b01}^2 & \sigma_{b1}^2 \end{bmatrix}$$

Table 5.1 shows the model fit parameters for the two linear models. Note that Maximum Likelihood Estimation was used for fitting purposes to allow for model comparison. The Akaike Information criterion (AIC) and Bayesian Information Criterion (BIC) were used to compare models. The likelihood ratio test statistic was used to determine if additional random effects were required in the model.

Table 5.1: Comparison of linear models

	Random effects	df	AIC	BIC	Test	p-value
Linear1	Intercept	4	53373	53401		NA
Linear2	Intercept and age	6	53053	53095	1 vs 2	<0.001

Both the AIC statistic and the likelihood ratio test statistic comparing the nested models indicates that the random effect for slope significantly improves the model fit.

Table 5.2 shows the fixed effects of Linear2, the preferred linear model. Table A.1 in the Appendix shows the random effect components of the model. From this we derive the covariance matrix as:

$$\Psi = \begin{bmatrix} \sigma^2_{b0} & \sigma^2_{b01} \\ \sigma^2_{b01} & \sigma^2_{b1} \end{bmatrix} = \begin{bmatrix} 1.4946^2 & (0.578)(1.4946)(0.0887) \\ (0.578)(1.4946)(0.0887) & 0.0887^2 \end{bmatrix} = \begin{bmatrix} 2.23383 & 0.076626 \\ 0.076626 & 0.007868 \end{bmatrix}$$

Table 5.2: Linear2 Fixed Effects

	Estimate	Std.Error	DF	t-value	p-value
(Intercept)	56.2	0.0858	7737	655	<0.001
age	1.04	0.00471	7737	221	<0.001

Figure 5.1 shows the standarised residuals of Linear2. These are the subject-level residuals obtained by subtracting the fitted values at the subject level from the actual values (and dividing by the estimated within group standard error). The fitted values at the subject level are obtained by adding together the population fitted values (based on the fixed effect estimates) and the estimated contributions of the random effects. This results in a conditional residual.

We can see that the residuals follow a trend and are not scattered around 0, indicating poor fit. Figure A.1 shows a histogram of the residuals and indicates that within-group errors are fairly normally distributed. Figure A.2 shows that the assumption of normality seems reasonable for both random effects. Figure 5.2 shows the estimated profiles superimposed on the actual values. We can see that the actual profiles are not linear. The model should be capable of capturing the non-linear developmental patterns in individual growth. The height curve gradually departs negatively from a straight line as age increases. To create such a curve, a quadratic age term was added in the next model.

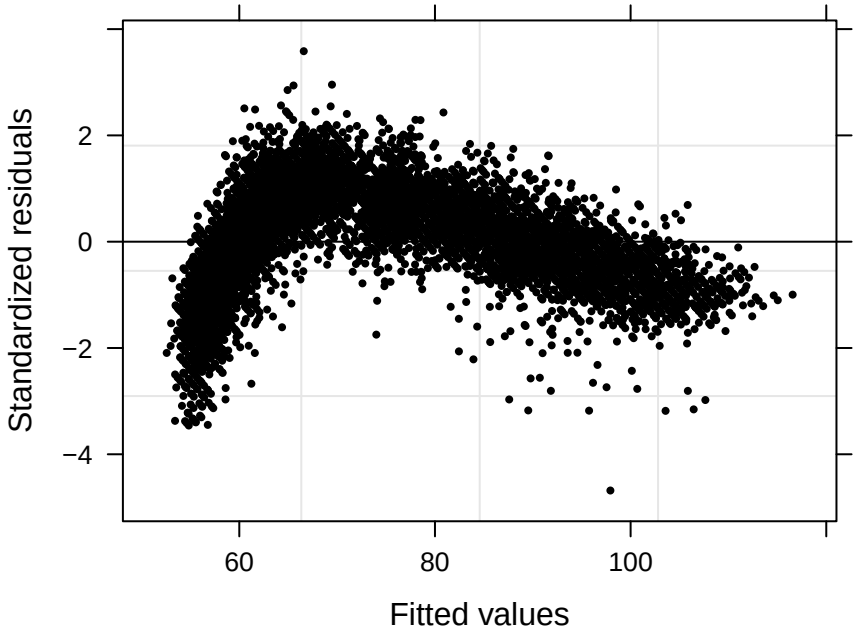


Figure 5.1: Linear2 Residuals

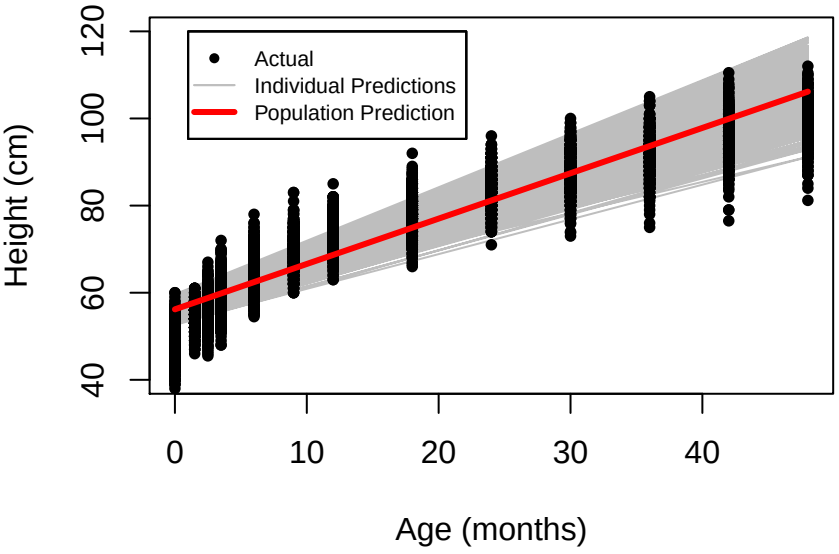


Figure 5.2: Linear2 Estimated Profiles

5.2.2 Quadratic growth model

Since a non-linear trend was identified in the trajectory plot, we try to improve on the linear growth model by adding a quadratic term. The quadratic model has been used by M. M. Black and Krishnakumar (1999) to model height between 0 and 6 years of age, and by Grimm, Ram, and Hamagami (2010) to model height between 3 and 19 years of age. The intercept still represents the birth height. However interpretation of the slope parameters is tricky since the linear and quadratic slopes both affect the rate of change. As with the linear model, expected trajectories will be unrealistic outside of the observation period. Grimm, Ram, and Hamagami (2010) showed that quadratic models are inadequate in modelling child growth over long age intervals .

For y_{ij} = height of child i at time j, let

$y_{ij} = \beta_{0i} + \beta_{1i}t_{ij} + \beta_{2i}t_{ij}^2$

Here the fixed-effects regressor matrix is given as:

$$X_i = \begin{bmatrix} 1 & t_0 & t_0^2 \\ 1 & t_1 & t_1^2 \\ 1 & t_2 & t_2^2 \\ \vdots & \vdots & \vdots \\ 1 & t_{13} & t_{13}^2 \end{bmatrix}$$

The intercept refers to height at birth, the t and t^2 terms allow for a polynomial curve for growth.

We progressively fit models that add random effects to the intercept, intercept $+t$, and intercept $+t+t^2$ terms respectively. Table 5.3 shows the required model fit parameters.

Table 5.3: Comparison of quadratic models

	Random effects	df	AIC	BIC	Test	p-value
Quadratic1	Intercept	5	47253	47289		NA
Quadratic2	Intercept and age	7	46761	46811	1 vs 2	<0.001
Quadratic3	Intercept, age and age^2	10	46722	46793	2 vs 3	<0.001

The AIC and likelihood ratio test statistics indicate that random effects on both the linear and quadratic changes are needed.

Therefore,

$$\beta_{0i} = \beta_0 + b_{0i} \quad \beta_{1i} = \beta_1 + b_{1i} \quad \beta_{2i} = \beta_2 + b_{2i}$$

Since each model parameter takes on a different value for each child, each parameter has variance and there exists a covariance between parameters. The underlying variance-covariance structure is:

$$\Psi = \begin{bmatrix} \sigma^2_{b0} & & \\ \sigma^2_{b01} & \sigma^2_{b1} & \\ \sigma^2_{b02} & \sigma^2_{b12} & \sigma^2_{b2} \end{bmatrix}$$

where σ^2_{b0}, σ^2_{b1} and σ^2_{b2} are the variances of the three random effects σ^2_{b01}, σ^2_{b12} and σ^2_{b02} are the covariances between the random effects. The unstructured variance-covariance structure is used since it allows the variance-covariance estimates to be distinct. For height curves, the variances of the random effects are unlikely to be the same because they explain different aspects of the individual curves (W. Johnson, Balakrishna, and Griffiths 2014)

Table 5.4, A.3 and A.4 shows the fixed effect estimates, random effect estimates and covariance estimates respectively. The positive coefficient for age refers to the increase in height with increasing age, while the negative coefficient for age^2 reflects a slowing down of this increase.

Table 5.4: Quadratic3 Fixed Effects

	Estimate	Std.Error	DF	t-value	p-value
(Intercept)	52.5	0.0895	7736	586	<0.001
age	1.84	0.00989	7736	186	<0.001
I(age^2)	-0.0196	0.000224	7736	-87.7	<0.001

The residual error has decreased from the linear model. However the standardized

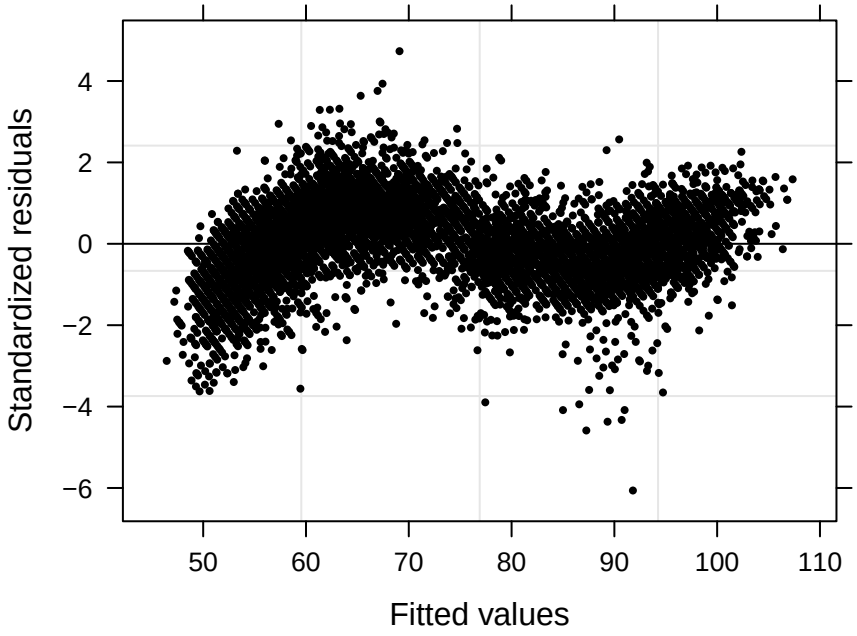


Figure 5.3: Quadratic3 Residuals

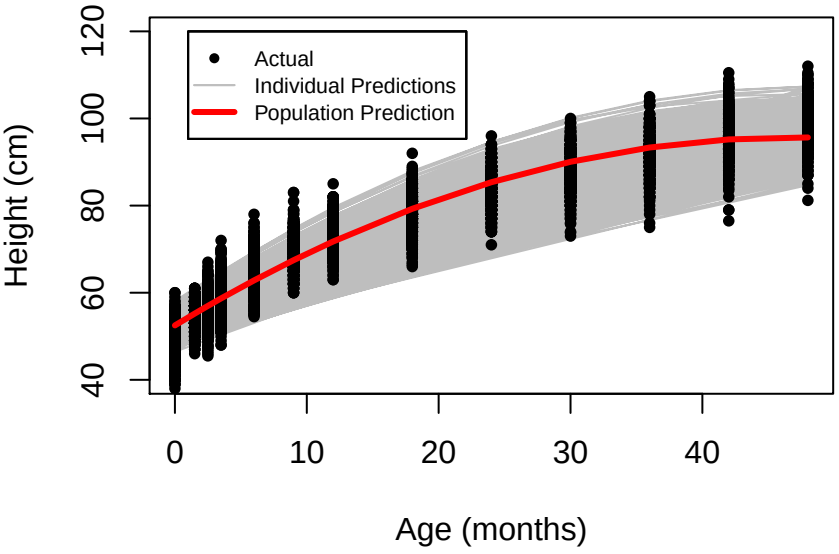


Figure 5.4: Quadratic3 Estimated Profiles

residuals still follow a pattern and therefore the fit can be further improved. Figure A.3 shows a histogram of the residuals and indicates that within-group errors are fairly normally distributed. Figure A.4 shows that the assumption of normality seems reasonable for all random effects.

Figure 5.4 shows the estimated profiles superimposed on the actual values. We can see that the actual profiles may follow a quadratic shape, however the fitted values do not seem to adequately capture the growth.

5.2.3 Cubic change

We continue the fitting of polynomial models and add a 3rd order term. The intent here is to better capture the nonlinear developmental pattern. However, as with the quadratic model, there is no biological interpretation of the slope parameters. Steele (2008) showed that a cubic polynomial model adequately modelled the height of boys between 11 and 14 years of age.

Table 5.5 shows the model fit parameters for the four cubic models.

Table 5.5: Comparison of cubic models

	Random effects	df	AIC	BIC	Test	p-value
Cubic1	Intercept	6	44362	44405		NA
Cubic2	Intercept and age	8	43290	43346	1 vs 2	<0.001
Cubic3	Intercept, age and age^2	11	43182	43260	2 vs 3	<0.001
Cubic4	Intercept, age, age^2 and age^3	15	43192	43299	3 vs 4	0.730168

The likelihood ratio test statistic indicates that the random effect on the cubic change is not needed. Therefore Cubic3 is the preferred model.

Table 5.6, A.5 and A.6 shows the fixed effect estimates, random effect estimates and covariance estimates respectively. Comparing the fixed-effect coefficients with the quadratic model, we see a greater positive increase in height with age, followed by a slowing down through the negative estimated coefficient for age^2, but with a fur-

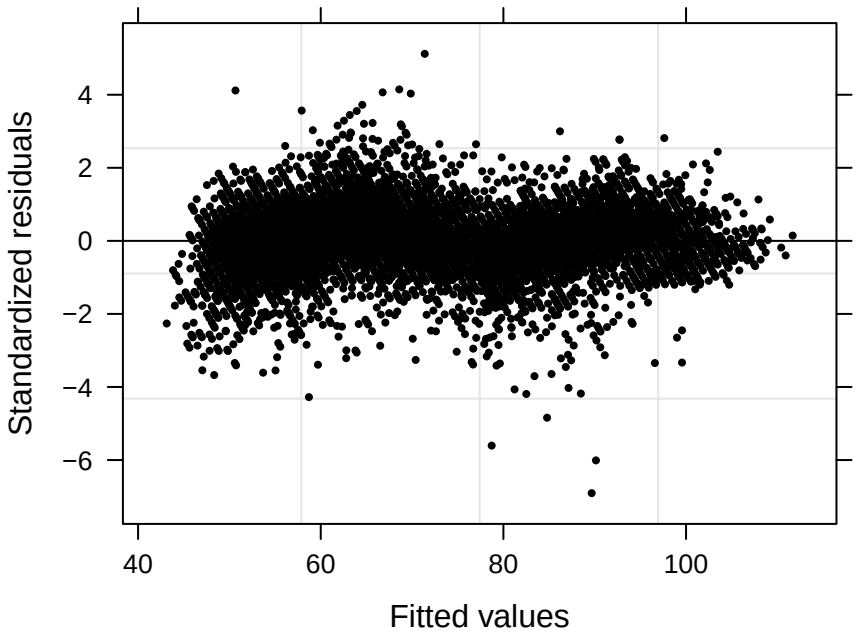


Figure 5.5: Cubic3 Residuals

ther correction through the cubic term which allows for a more rapid increase in height in later ages.

Table 5.6: Cubic3 Fixed Effects

	Estimate	Std.Error	DF	t-value	p-value
(Intercept)	50.5	0.0954	7735	529	<0.001
age	2.66	0.0148	7735	180	<0.001
I(age^2)	-0.0698	0.000757	7735	-92.2	<0.001
I(age^3)	0.000755	1.11e-05	7735	68.3	<0.001

Figure 5.5 shows that the standardized residuals are now scattered around zero, however there is still a slight trend. Figure 5.6 shows that the predicted growth curve allows for a faster initial increase in height that subsequently slows down, and then increases again. The quadratic model has been overlayed to show how the cubic term improves the predictions. Figures A.9 and A.10 show that the normality assumptions

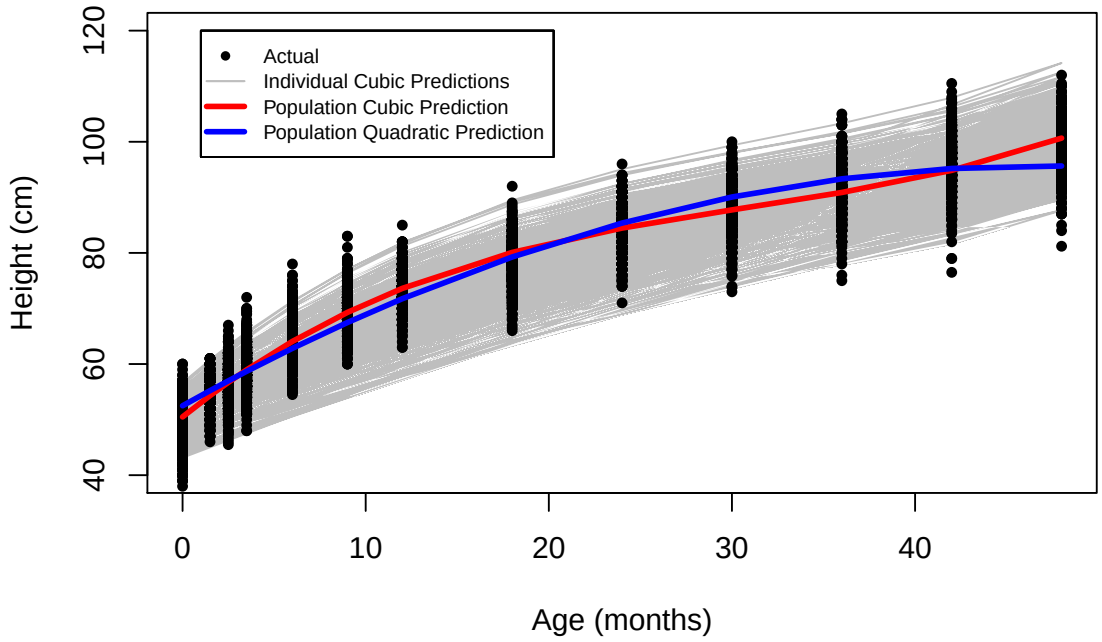


Figure 5.6: Cubic3 Estimated Profiles

are satisfied.

5.2.4 Fractional Polynomial Model

Studies such as Tilling et al. (2014) have shown fractional polynomials to be useful in modelling child growth. They allow for an initially fast rate of growth which slows over time. They give a realistic, smooth function. However they have the disadvantage of the parameters not having a biological interpretation, as with other non-structural models. Fractional polynomial models are still linear in their parameters, however they are more flexible than conventional polynomials because logarithms, non-integer powers and repeats of powers are allowed. The form of the fractional polynomial used was the best-fitting option from W. Johnson, Balakrishna, and Griffiths (2014) which modelled child growth under a similar setting.

The model for the height y_{ij} of child i at age t_{ij} is

$$y_{ij} = \beta_{0i} + \beta_{1i}t_{ij} + \beta_{2i}log(t_{ij}+2) + \beta_{3i}(t_{ij}+2)^{0.5} + \epsilon_{ij}$$

The random effects were progressively added to each term. Table 5.7 shows model fit parameters. The model with a random effect on the $(t_{ij}+2)^{0.5}$ did not coverge.

Table 5.7: Comparison of Fractional Polynomial models

	Random effects	df	AIC	BIC	Test	p-value
Fractional1	Intercept	6	43937	43980		NA
Fractional2	Intercept and age	8	42807	42864	1 vs 2	<0.001
Fractional3	Intercept, age and log(age+2)	11	42616	42694	2 vs 3	<0.001

The AIC and likelihood ratio test statistic shows that all random effects are needed and therefore Fractional3 is the preferred model.

Table 5.8, A.7 and A.8 shows the fixed effect estimates, random effect estimates and covariance estimates respectively. We see that the residual error of the fractional polynomial model is smaller than the previous polynomials, indicating a better fit than the linear, quadratic and cubic models.

Table 5.8: Fractional3 Fixed Effects

	Estimate	Std.Error	DF	t-value	p-value
(Intercept)	36.7	0.421	7735	87.3	<0.001
age	-0.0594	0.0344	7735	-1.72	0.084568
I(log(age + 2))	4.69	0.406	7735	11.5	<0.001
I((age + 2)^0.5)	6.57	0.496	7735	13.2	<0.001

Figure 5.7 shows the standardised residuals of the fractional polynomial model. We can see that they are randomly scattered around 0, indicating a good fit. Figures A.7

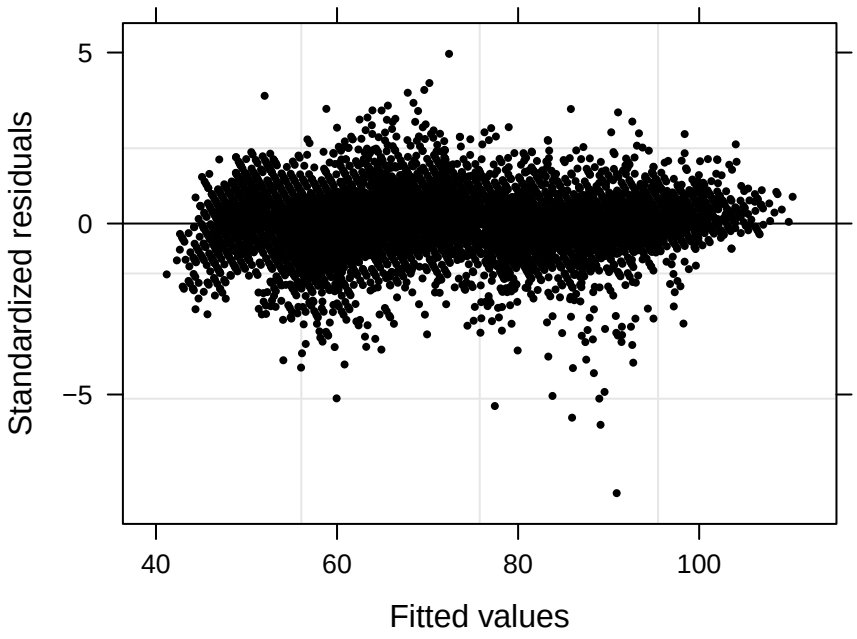


Figure 5.7: Fractional3 Residuals

and A.8 show that the normality assumptions are satisfied.

Figure 5.8 shows the estimated profiles of the fractional polynomial model, overlayed with the actual measurements. We can see that the model captures the trend well.

5.2.5 Count model

The Count model, developed by Earl Count in 1943, was put forward for modelling anthropometric variables including weight, height and head circumference. The intent was to develop a structural model, i.e. one that has biologically interpretable parameters, that is linear in the parameters. The rationale for this was that linear models allow for easier fitting and analysis of parameters (C. S. Berkey 1982). It has been widely used for the study of early childhood growth.

The model for the height y_{ij} of child i at age t_{ij} is

$$y_{ij} = \beta_{0i} + \beta_{1i}t_{ij} + \beta_{2i}ln(t_{ij}) + \epsilon_{ij}$$

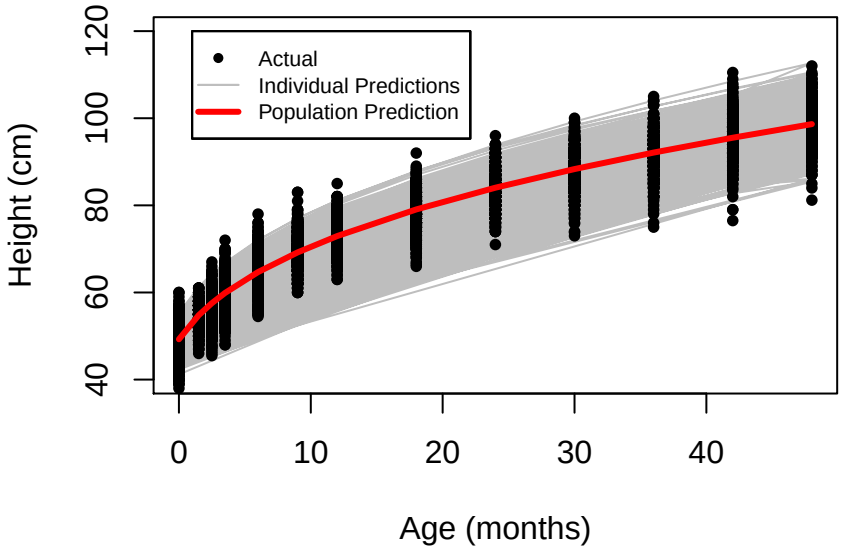


Figure 5.8: Fractional3 Estimated Profiles

where

$$\boldsymbol{\beta_i} = \begin{bmatrix} \beta_{0i} \\ \beta_{1i} \\ \beta_{2i} \end{bmatrix} = \begin{bmatrix} \beta_0 \\ \beta_1 \\ \beta_2 \end{bmatrix} + \begin{bmatrix} b_{0i} \\ b_{1i} \\ b_{2i} \end{bmatrix} = \boldsymbol{\beta} + \boldsymbol{b_i}$$

$$\boldsymbol{b_i} \sim \boldsymbol{N(0, \Psi)}, \quad \boldsymbol{\epsilon} \sim N(0, \sigma^2)$$

The model parameters can be interpreted as follows:

β_{0i} is related to the baseline height at birth. β_{1i} is related to the linear component of the growth velocity. β_{2i} is related to the deceleration in growth velocity (Chirwa et al. 2014).

These parameters are allowed to vary with child. The fixed effects, $\boldsymbol{\beta}$, represent the population average of the individual parameters, $\boldsymbol{\beta_i}$. The random effects, $\boldsymbol{b_i}$, represent the deviations of the $\boldsymbol{\beta_i}$ from their population average. The Count model is not defined at age=0. Since we require a model that fits the neonatal period, it is important to

include birth height in the modelling. Therefore the model was modified as follows:

$$y_{ij} = \beta_{0i} + \beta_{1i}t_{ij} + \beta_{2i}ln(t_{ij} + 1) + \epsilon_{ij}$$

This has been the approach taken in other studies, such as K. B. Simondon et al. (1992).

The random effects were progressively added to each term. Table 5.9 shows the model fit parameters.

Table 5.9: Comparison of Count models

	Random effects	df	AIC	BIC	Test	p-value
Count1	Intercept	5	44696	44732		NA
Count2	Intercept and age	7	43736	43786	1 vs 2	<0.001
Count3	Intercept, age and log(age+1)	10	43628	43699	2 vs 3	<0.001

The AIC and likelihood ratio test statistics show that the random effects on all terms are needed and therefore Count3 is the preferred model.

Table 5.10 shows the fixed effect estimates. We can see that the baseline height at birth is $\beta_0 = 48.4cm$, the linear growth rate is $\beta_1 = 0.517cm$ per month and deceleration in growth velocity is $\beta_2 = 6.98cm$ per month. To understand this deceleration in growth velocity, figure 5.9 shows the growth curve for different values of β_2. We can see that a higher β_2 value results in a slower deceleration in growth velocity.

Tables A.9 and A.10 show the random effect estimates and covariance estimates respectively. The residual error of the Count model is smaller than the linear, quadratic and cubic models, but larger than the fractional polynomial model. However the count model has the advantage of interpretable parameters, which the fractional polynomial model does not.

Table 5.10: Count3 Fixed Effects

	Estimate	Std.Error	DF	t-value	p-value
(Intercept)	48.4	0.106	7736	455	<0.001

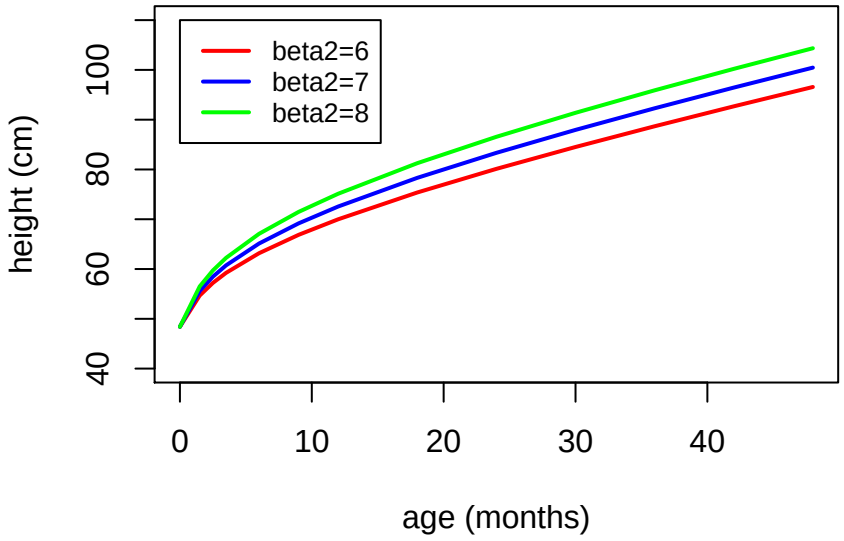


Figure 5.9: Count model beta2 parameter

	Estimate	Std.Error	DF	t-value	p-value
age	0.517	0.00539	7736	95.9	<0.001
I(log(age + 1))	6.98	0.0614	7736	114	<0.001

Figure 5.10 shows the standardised residuals of the Count model. We can see that the residuals are not completely randomly scattered around 0, indicating that the model may not be an appropriate fit to the data. Figures A.9 and A.10 show that the normality assumptions are satisfied.

Figure 5.11 shows that the Count model seems to adequately capture the rapid increase in height during the first few months following birth. However the Count model has a monotonic decrease in the growth velocity, while the actual growth curves appear to be more complex over time. The next model, Berkey-Reeds, improves on this lack of flexibility by allowing for an inflection point.

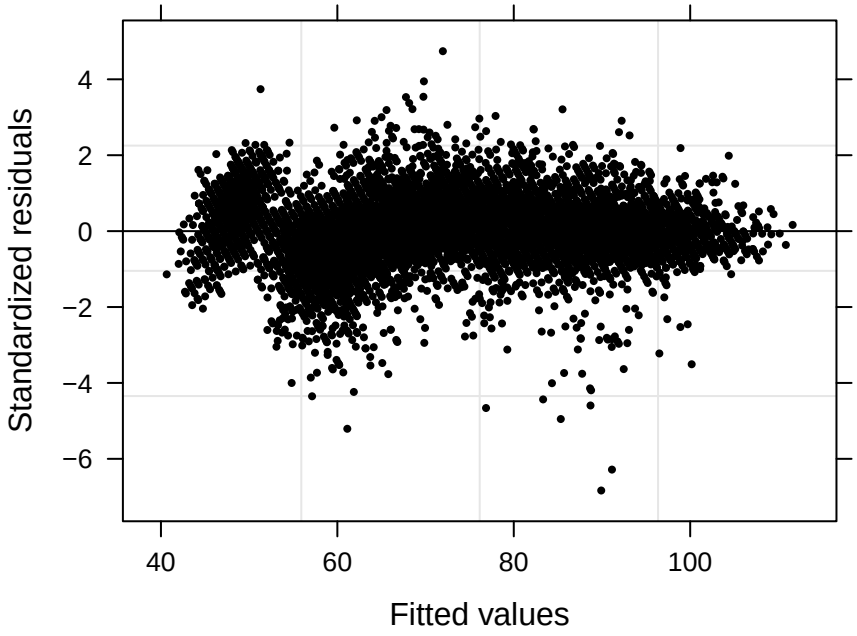


Figure 5.10: Count3 Residuals

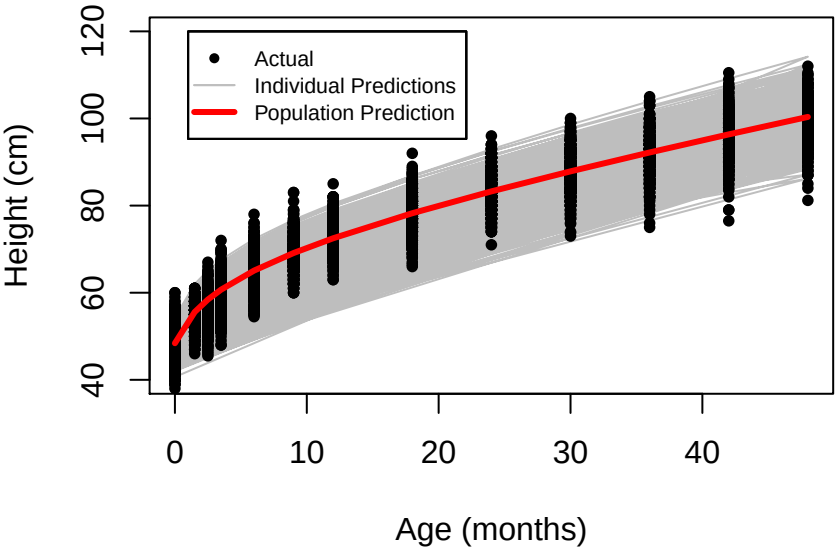


Figure 5.11: Count3 Estimated Profiles

5.2.6 Berkey-Reeds 1st order (Reed1) model

The Berkey-Reeds 1st and 2nd order models (1987) are extensions of the Count model, commonly used to describe normal and abnormal growth in early childhood (Chirwa et al. 2014). This four parameter model is more flexible than the three parameter Count model since it allows for an inflection point. The inflection point is where growth changes from deceleration to acceleration, or vice versa. This is where a minimum or maximum growth velocity occurs. Inflection points are useful for describing unusual patterns in growth resulting from sub-optimal nutritional or environmental factors. The Reed1 model has consistently proved to be the preferred structural model for early childhood growth compared to other common growth models (W. Johnson, Balakrishna, and Griffiths 2014, Pizzi et al. (2014), Chirwa et al. (2014), K. B. Simondon et al. (1992)).

The model for the height y_{ij} of child i at age t_{ij} is

$$y_{ij} = \beta_{0i} + \beta_{1i}t_{ij} + \beta_{2i}ln(t_{ij}) + \beta_{3i}\frac{1}{t_{ij}} + \epsilon_{ij}$$

where

$$\boldsymbol{\beta_i} = \begin{bmatrix} \beta_{0i} \\ \beta_{1i} \\ \beta_{2i} \\ \beta_{3i} \end{bmatrix} = \begin{bmatrix} \beta_0 \\ \beta_1 \\ \beta_2 \\ \beta_3 \end{bmatrix} + \begin{bmatrix} b_{0i} \\ b_{1i} \\ b_{2i} \\ b_{3i} \end{bmatrix} = \boldsymbol{\beta} + \boldsymbol{b_i}$$

$\boldsymbol{b_i} \sim \boldsymbol{N(0, \Psi)}, \quad \epsilon \sim N(0, \sigma^2)$

The model parameter interpretation is as follows:

β_{0i} is related to the baseline height at birth. β_{1i} is related to the linear component of the growth velocity. β_{2i} is related to the deceleration in growth velocity. β_{3i} represents an inflection point that allows growth velocity to peak after birth rather than at birth. (Chirwa et al. 2014)

Since the Berkey-Reeds model is not defined at age=0, the model was modified as follows:

$$y_{ij} = \beta_{0i} + \beta_{1i}t_{ij} + \beta_{2i}ln(t_{ij}+1) + \beta_{3i}\frac{1}{t_{ij}+1} + \epsilon_{ij}$$

The random effects were progressively added to each term. Table 5.11 shows the model fit parameters.

Table 5.11: Comparison of Reed1 models

	Random effects	df	AIC	BIC	Test	p-value
BerkeyReed1.1	Intercept	6	43761	43804		NA
BerkeyReed1.2	Intercept and age	8	42560	42617	1 vs 2	<0.001
BerkeyReed1.3	Intercept, age and log(age+1)	11	42346	42424	2 vs 3	<0.001
BerkeyReed1.4	Intercept, age, log(age+1) and 1/(age+1)	15	42244	42350	3 vs 4	<0.001

The AIC and likelihood ratio test statistics indicate that random effects on all terms are needed. Table 5.12 shows the fixed effect estimates. We can see that the baseline height at birth is $\beta_0 = 38cm$. The linear component of the growth velocity is $\beta_0 = 0.286cm$ per month. The deceleration in growth velocity is $\beta_2 = 12$ which is much higher than the count model. This is because the inflection point of $\beta_3 = 11.6cm$ allows for the age at which the growth changes from positive to negative.

A.11 and A.12 shows the random effect estimates and covariance estimates respectively.

Table 5.12: BerkeyReed1.4 Fixed Effects

	Estimate	Std.Error	DF	t-value	p-value
(Intercept)	38	0.335	7735	114	<0.001
age	0.286	0.00798	7735	35.8	<0.001
I(log(age + 1))	12	0.162	7735	74.2	<0.001
I(1/(age + 1))	11.6	0.348	7735	33.3	<0.001

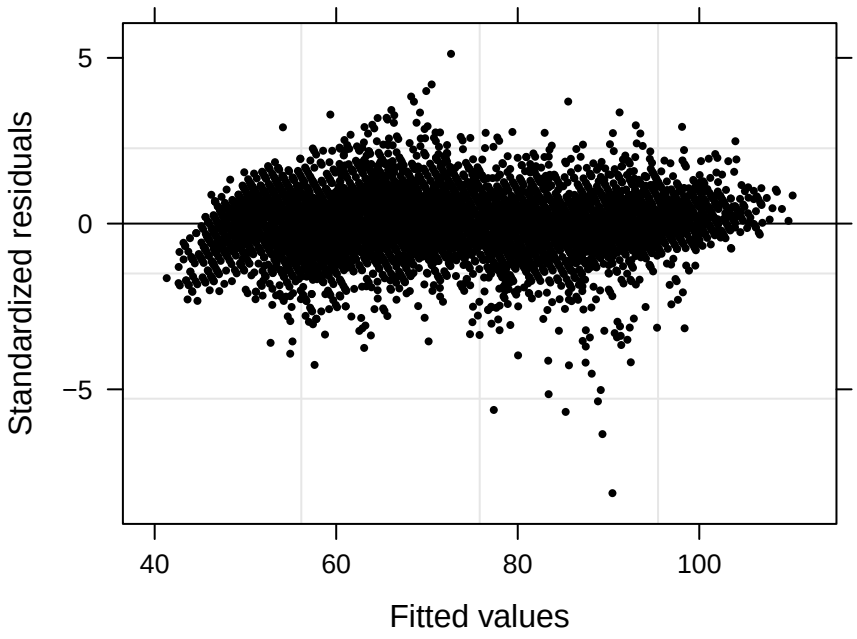


Figure 5.12: BerkeyReed1.4 Residuals

Figure 5.12 shows that the standarized residuals are randomly scattered around zero, indicating a good model fit. Figures A.11 and A.12 show that the normality assumptions are satisfied.

Figure 5.13 shows that the predictions are well aligned with the actual measurements. The model appears to capture the acceleration and deceleration in the actual height measurments.

5.2.7 Berkey-Reeds 2nd order (Reed2) model

The Berkey-Reeds 2nd order model is an extension of the Berkey-Reeds 1st order model that allows for a second inflection point. In most cases only a few children need the second order version (K. B. Simondon et al. 1992).

The model for the height y_{ij} of child i at age t_{ij} is

$$y_{ij} = \beta_{0i} + \beta_{1i}t_{ij} + \beta_{2i}ln(t_{ij}) + \beta_{3i}\frac{1}{t_{ij}} + \beta_{4i}\frac{1}{t_{ij}^2} + \epsilon_{ij}$$

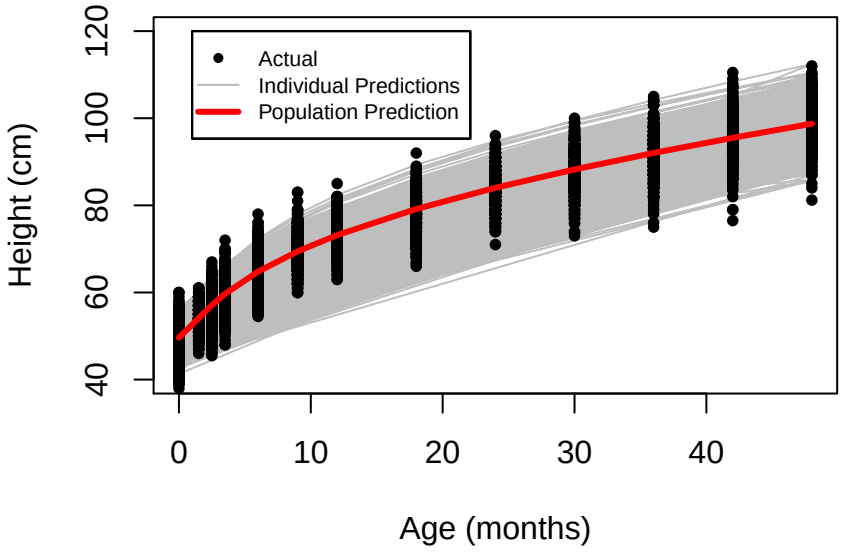


Figure 5.13: Model6.4 Estimated Profiles

where

$$\boldsymbol{\beta_i} = \begin{bmatrix} \beta_{0i} \\ \beta_{1i} \\ \beta_{2i} \\ \beta_{3i} \\ \beta_{4i} \end{bmatrix} = \begin{bmatrix} \beta_0 \\ \beta_1 \\ \beta_2 \\ \beta_3 \\ \beta_4 \end{bmatrix} + \begin{bmatrix} b_{0i} \\ b_{1i} \\ b_{2i} \\ b_{3i} \\ b_{4i} \end{bmatrix} = \boldsymbol{\beta} + \boldsymbol{b_i}$$

$\boldsymbol{b_i} \sim \boldsymbol{N(0, \Psi)}, \quad \epsilon \sim N(0, \sigma^2)$

The model parameters have the same interpretation as the Reed1 model however the 5th parameter allows a second inflection point. The formula was adjusted to allow for birth height as follows:

$$y_{ij} = \beta_{0i} + \beta_{1i} t_{ij} + \beta_{2i} ln(t_{ij} + 1) + \beta_{3i} \frac{1}{t_{ij} + 1} + \beta_{4i} \frac{1}{(t_{ij} + 1)^2} + \epsilon_{ij}$$

The random effects were progressively fit to the model terms. Table 5.13 shows the

model fit parameters.

Table 5.13: Comparison of Reed2 models (continued below)

	Random effects	df	AIC	BIC
BerkeyReed2.1	Intercept	7	43742	43791
BerkeyReed2.2	Intercept and age	9	42523	42586
BerkeyReed2.3	Intercept, age and log(age+1)	12	42307	42392
BerkeyReed2.4	Intercept, age, log(age+1)	16	42187	42301
BerkeyReed2.5	1/(age+1) and 1/((age+1)^2)	21	42198	42347

	Test	p-value
BerkeyReed2.1		NA
BerkeyReed2.2	1 vs 2	<0.001
BerkeyReed2.3	2 vs 3	<0.001
BerkeyReed2.4	3 vs 4	<0.001
BerkeyReed2.5	4 vs 5	0.9867

The AIC and likelihood ratio test statistics shows that the random component on the second inflection point is not needed, therefore model 2.4 is preferred.

Table 5.15 shows the fixed effect estimates. We can see that the baseline height, $\beta_0 = 45.7cm$, and linear growth, $\beta_1 = 0.372$, is higher than the Berkey-Reed 1st order model. The deceleration in growth velocity, $\beta_2 = 9.14$ is lower than the Berkey-Reed 1st order model. The negative value of the first inflection point, $\beta_3 = -6.92$ alows the acceleration in growth to increase at around 7 months, while the positive value of the second inflection point, $\beta_4 = 10.9$ allows the growth to change from acceleration to deceleration at around 11 months.

Tables A.13 and A.14 show random effect estimates and covariance estimates respectively.

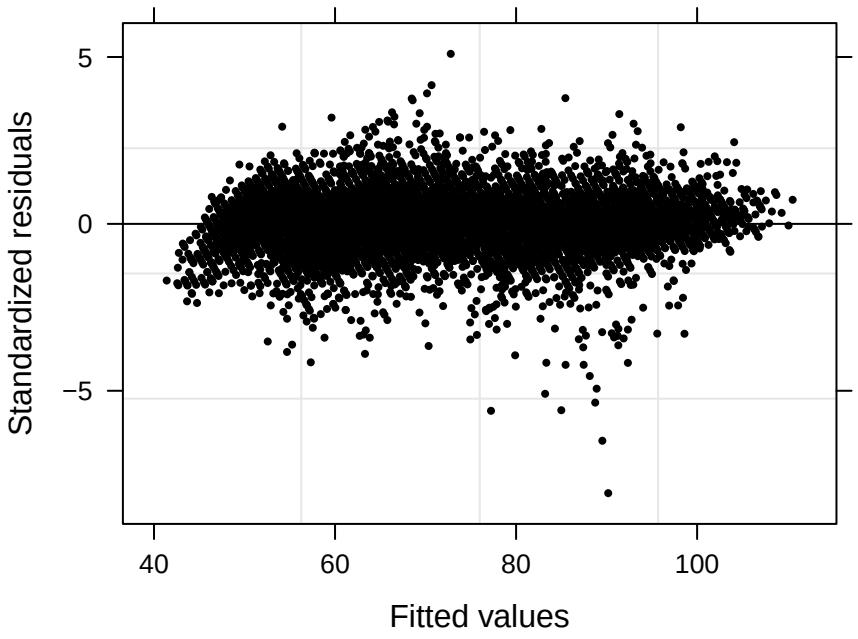


Figure 5.14: BerkeyReed2.4 Residuals

Table 5.15: BerkeyReed2.4 Fixed Effects

	Estimate	Std.Error	DF	t-value	p-value
(Intercept)	45.7	1.05	7734	43.4	<0.001
age	0.372	0.0138	7734	27	<0.001
I(log(age + 1))	9.14	0.406	7734	22.5	<0.001
I(1/(age + 1))	-6.92	2.42	7734	-2.86	0.004191
I(1/((age + 1)^2))	10.9	1.4	7734	7.75	<0.001

Figure 5.14 shows that the stardised residual plot is very similar to the Reed1 model.

Figure 5.15 shows that the predictions are very similar to the Reed1 model.

Table 5.16 shows a comparison of the Reed1 and Reed2 models. We can see that the Reed2 model has a lower AIC value. The likelihood ratio test statistic shows that the second inflection point is needed.

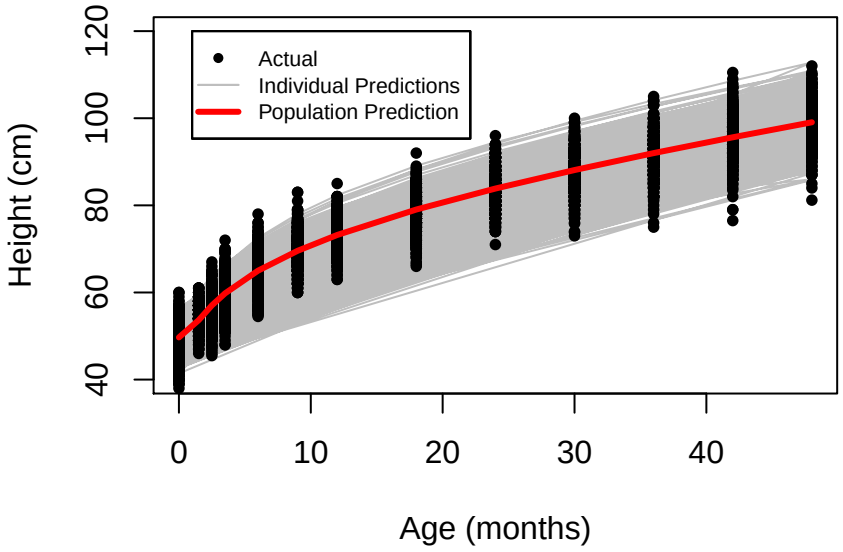


Figure 5.15: BerkeyReed2.4 Estimated Profiles

Table 5.16: Comparison of Reed models

	df	AIC	BIC	Test	p-value
BerkeyReed1.4	15	42244	42350		NA
BerkeyReed2.4	16	42187	42301	1 vs 2	<0.001

Finally, we compare the best fitting models from each linear model. Table 5.17 shows a comparison of the models that are linear in parameters. We can see that out of the linear mixed effects models, the Reed2 model has the smallest AIC and BIC test statistics, indicating that the Reed2 model is the preferred model.

Table 5.17: Comparison of linear mixed effect models

	df	AIC	BIC
Linear2	6	53053	53095
Quadratic3	10	46722	46793
Cubic3	11	43182	43260
Fractional3	11	42616	42694
Count3	10	43628	43699
BerkeyReed1.4	15	42244	42350
BerkeyReed2.4	16	42187	42301

Figure 5.16 shows all linear models superimposed on the observed profiles. We can see that the Fractional polynomial, Count and Berkey-Reed models all predict similarly.

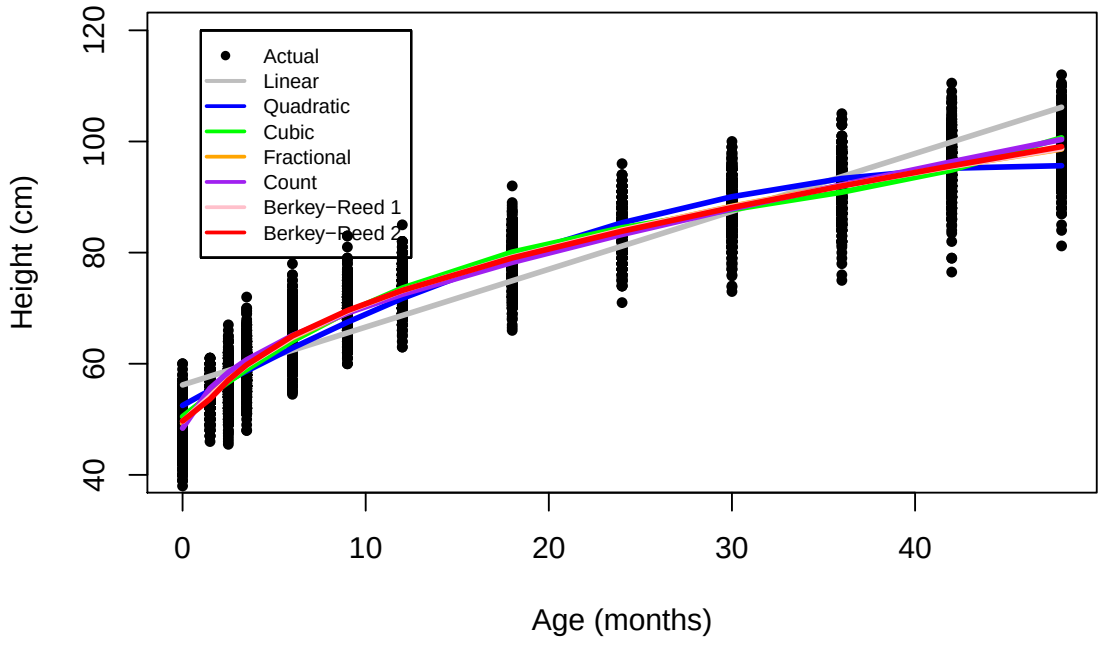


Figure 5.16: Comparison of linear model estimated profiles

5.3 Nonlinear Mixed Effects Models

Nonlinear mixed-effects (NLME) models are models in which the response is a nonlinear function of the parameters. The jth height measurement of the ith child is described as:

$$y_{ij} = f(\boldsymbol{\phi_{ij}}, \boldsymbol{x_{ij}}) + \epsilon_{ij}, \quad i = 1, ..., N, \;\; j = 1, ...n_i$$

where f is a nonlinear function of parameters $\boldsymbol{\phi_{ij}}$ and the covariates $\boldsymbol{x_{ij}}$, ϵ_{ij} are normally distributed within-group error terms. The group-specific parameters are modelled as:

$$\boldsymbol{\phi_{ij}} = \boldsymbol{A_{ij}}\boldsymbol{\beta} + \boldsymbol{B_{ij}}\boldsymbol{b_i}$$

where $\boldsymbol{\beta}$ represents the between group fixed effects, $\boldsymbol{b_i}$ the group-specific random effects with assumed distribution $\boldsymbol{N(0, \Psi)}$. $\boldsymbol{A_{ij}}$ and $\boldsymbol{B_{ij}}$ are design matrices for the fixed and random effects respectively, which may depend on the values of some covariates at the jth observation. It is further assumed that the b_i are independent of the ϵ_{ij}

5.3.1 Logistic growth model

The logistic growth curve is an S-shaped curve used to model fast-slow-fast growth.

The model for the height y_{ij} of child i at age t_{ij} is

$$y_i = \frac{\phi_{1i}}{1 + exp[-(t_{ij} - \phi_{2i})/\phi_{3i}]} + \epsilon_{ij}$$

$$\phi_i = \begin{bmatrix} \phi_{1i} \\ \phi_{2i} \\ \phi_{3i} \end{bmatrix} = \begin{bmatrix} \beta_1 \\ \beta_2 \\ \beta_3 \end{bmatrix} + \begin{bmatrix} b_{1i} \\ b_{2i} \\ b_{3i} \end{bmatrix} = \boldsymbol{\beta} + \boldsymbol{b_i}$$

$\boldsymbol{b_i \sim N(0, \Psi)}$,$\boldsymbol{\epsilon \sim N(0, \sigma^2)}$

ϕ_1 represents the upper asymptote, i.e. the height at the end of the growth cycle. ϕ_2 represents the age at the inflection point of the growth curve. ϕ_3 represents a scale

parameters on the x-axis (Aggrey 2009).

The random effects were progressively added to the terms of the model. Table 5.18 shows the model fit parameters of the various logistic models that were fit.

Table 5.18: Comparison of logistic models

	Random effects	df	AIC	BIC	Test	p-value
Logistic1	phi1	5	46543	46578		NA
Logistic2	phi1 and phi2	7	46311	46361	1 vs 2	<0.001
Logistic3	phi1, phi2 and phi3	10	46329	46400	2 vs 3	0.0057
Logistic4	phi1 and phi3	7	46538	46588	3 vs 4	<0.001

The AIC test statistic is smallest for model 2. However the likelihood ratio test statistic shows that all random effects are needed.

Table 5.19 and A.15 show the fixed effect estimates and random effect estimates for model 2.

Table 5.19: Three parameter logistic model Fixed Effects

	Estimate	Std.Error	DF	t-value	p-value
phi1	96.6	0.193	7736	502	<0.001
phi2	-1.82	0.0528	7736	-34.5	<0.001
phi3	12.3	0.0791	7736	156	<0.001

Figure 5.17 shows the standardised residuals of the logistic model. We can see that they are not randomly scattered around 0, indicating a poor fit. Figures A.15 and A.16 show that the normality assumptions are satisfied.

Figure 5.18 shows the estimated profiles of the logistic model, overlayed with the actual measurements. The model does not appear to adequately capture the trend. This was to be expected since the height profiles did not show an s-shape.

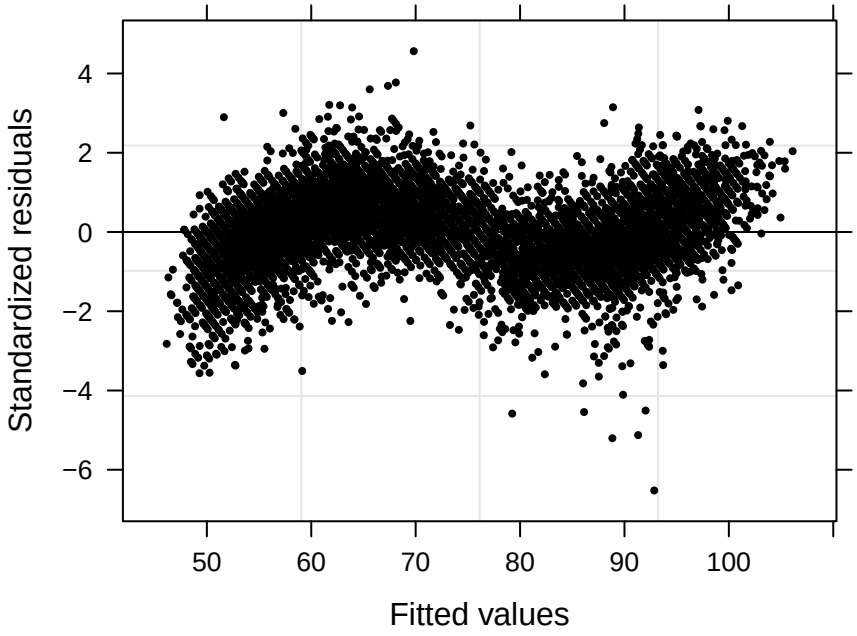


Figure 5.17: Three parameter logistic model Residuals

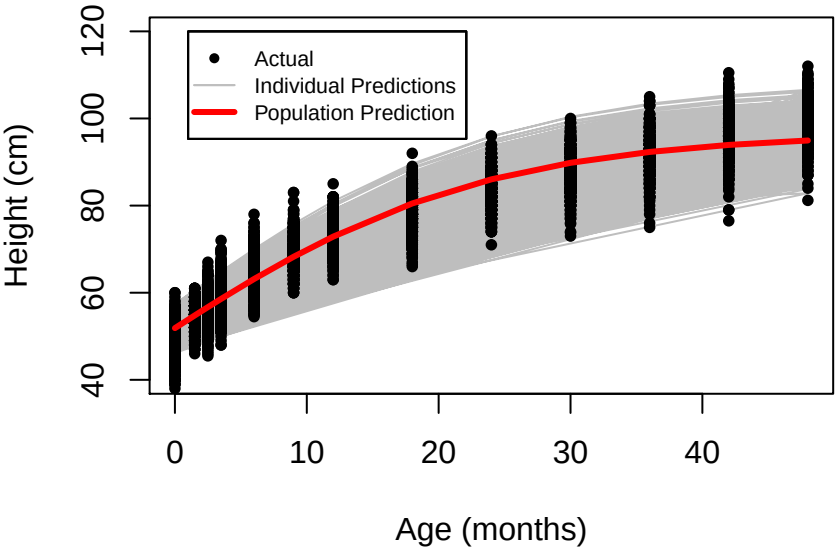


Figure 5.18: Three parameter logistic model Estimated Profiles

5.3.2 Gompertz growth model

Similarly to the Logistic curve, the Gompertz curve also follows a s-shape. However the Gompertz curve allows for faster early growth, with a slower approach to the asymptote and an extended period around the inflection point (Karkach 2006).

The model for the height y_{ij} of child i at age t_{ij} is

$$y_{ij} = \alpha exp(\beta_i(1 - exp(-\mu_i t_{ij}))) + \epsilon_{ij}$$

where α is the theoretical value of the upper asymptote, β is the initial instantaneous growth rate, and μ is the rate at which the growth rate β decreases.

Table 5.20 shows the required statistics for the various Gompertz models that were fit. The random effects were progressively added to the curve parameters.

Table 5.20: Comparison of Gompertz models

	Random effects	df	AIC	BIC	Test	p-value
Gompertz1	alpha	5	45808	45843		NA
Gompertz2	alpha and beta	7	45485	45534	1 vs 2	<0.001
Gompertz4	alpha, beta and mu	10	45392	45463	2 vs 3	<0.001
Gompertz3	alpha and mu	7	45771	45820	3 vs 4	<0.001

The AIC and likelihood ratio test statistic show that all random effects are needed.

Table 5.21 and A.16 show the fixed effect estimates and random effect estimates for model 2. We see that the residual error is smaller than the Logistic model, indicating an improved fit.

Table 5.21: Gompertz Model Fixed Effects

	Estimate	Std.Error	DF	t-value	p-value
alpha	98	0.259	7736	378	<0.001
beta	0.645	0.00239	7736	269	<0.001

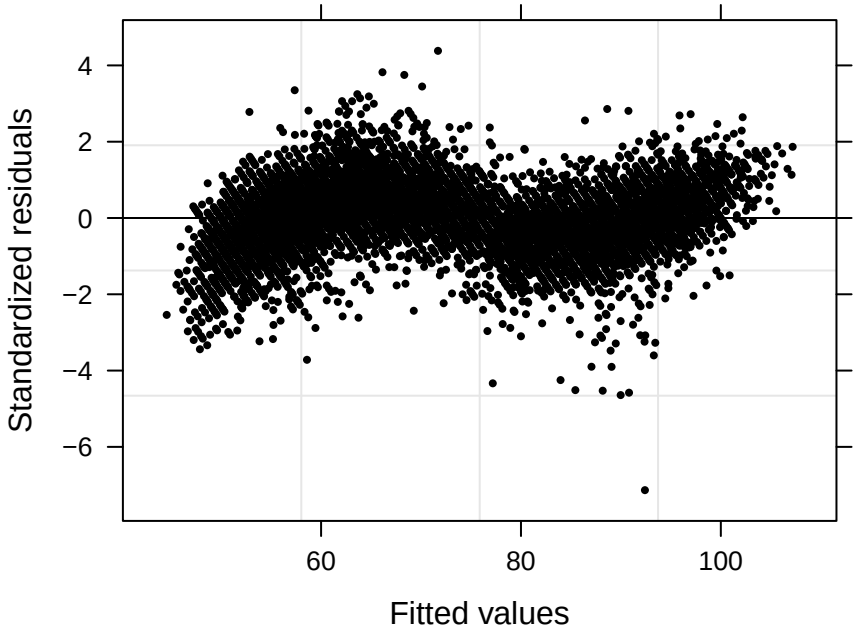


Figure 5.19: Gompertz Model Residuals

	Estimate	Std.Error	DF	t-value	p-value
mu	0.935	0.000603	7736	1550	<0.001

Figure 5.19 shows the standardised residuals of the Gompertz model. We can see that they are not randomly scattered around 0, indicating a poor fit. Figures A.17 and A.18 show that the normality assumptions are satisfied.

Figure 5.20 shows the estimated profiles of the Gompertz model. Although there is an improvement from the Logistic model, it is clear that an s-shape curve is not appropriate for modelling the height profile.

5.3.3 Exponential growth model

The exponential growth curve is a concave curve where the height y_{ij} of child i at age t_{ij} is

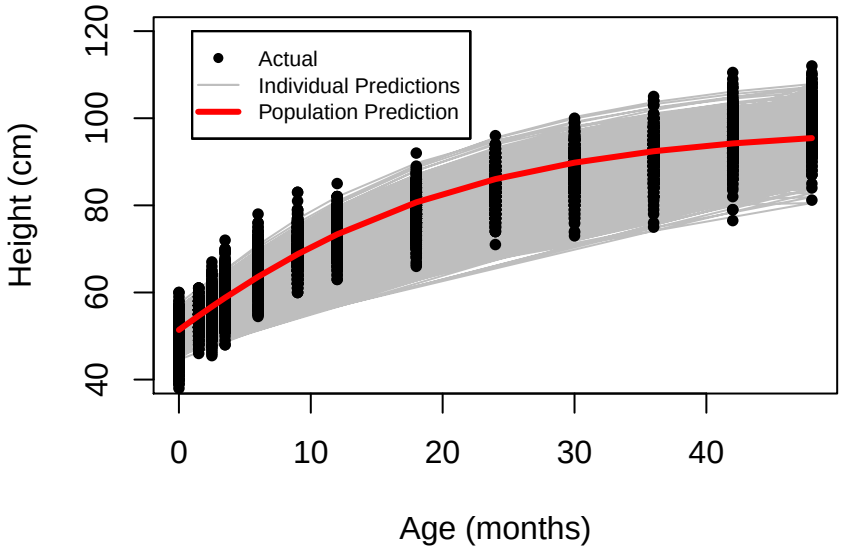


Figure 5.20: Gompertz Model Estimated Profiles

$$y_{ij} = \beta_{0i} exp(\beta_{1i} t_{ij}) + \epsilon_{ij}$$

where β_{0i} represents the initial height at birth.

Table 5.22 shows the AIC and likelihood ratio test statistics for the exponential models.

Table 5.22: Comparison of Exponential models

	Random effects	df	AIC	BIC	Test	p-value
Exponential1	beta0	4	56082	56111		NA
Exponential2	beta0 and beta1	6	55988	56030	1 vs 2	<0.001

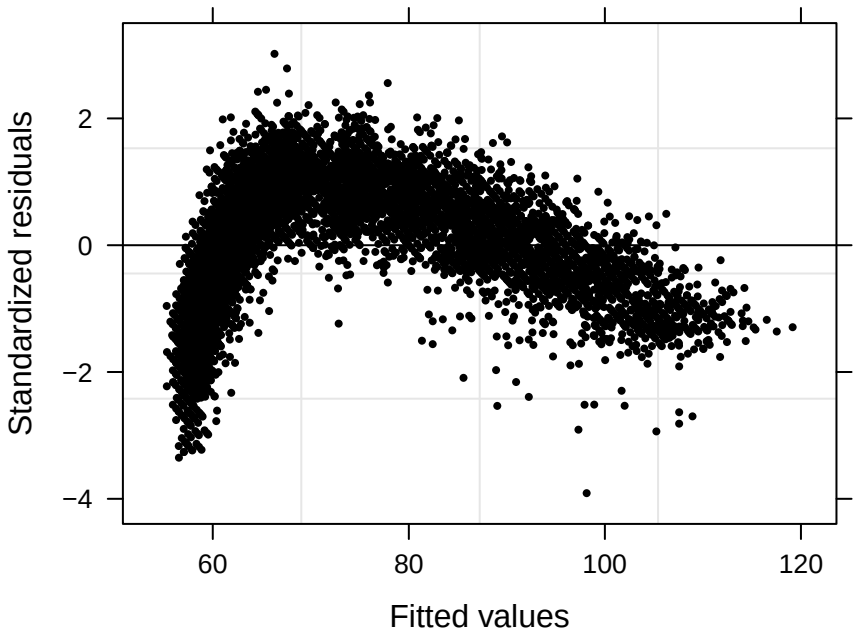


Figure 5.21: Exponential Model Residuals

We can see that the random effect on β_0 as well as the random effect on β_1 is needed in the model.

Table 5.23 and A.17 show the fixed effect estimates and random effect estimates for model 2. The residual error for this model is very high.

Table 5.23: Exponential Model Fixed Effects

	Estimate	Std.Error	DF	t-value	p-value
beta0	58.1	0.0875	7737	664	<0.001
beta1	0.0129	6.03e-05	7737	215	<0.001

Figure 5.21 shows the standardised residuals of the Exponential model. We can see that they are not randomly scattered around 0, indicating a poor fit. Figures A.19 and A.20 show that the normality assumptions are satisfied.

Figure 5.22 shows the estimated profiles of the Exponential model. The model allows height to increase exponentially with age and is not capturing the rapid initial increase

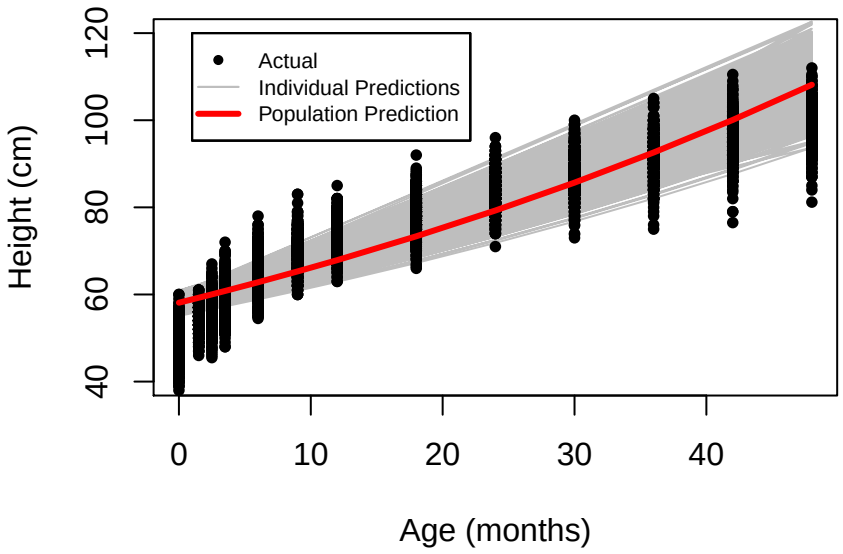


Figure 5.22: Exponential Model Estimated Profiles

in height following birth.

5.3.4 Jenss-Bayley growth model

Jenss and Bayley (1937) proposed a four-parameter nonlinear model to describe the growth of children from birth to 8 years of age. It has successfully been used to model child growth in various studies such as Dwyer et al. (1983) and Dommelen et al. (2005).

The curve is a negatively accelerated exponential and approaches a linear asymptote with positive slope. The height y_{ij} of child i at age t_{ij} is given as

$$y_{ij} = \beta_{0i} + \beta_{1i}t_{ij} - exp(\beta_{2i} + \beta_{3i}t_{ij}) + \epsilon_{ij}$$

The height at birth is represented by $\beta_0 - exp(\beta_2)$. β_1 represents the linear component of the growth velocity. The exponential component, $exp(\beta_2 + \beta_3 t)$, represents the decreasing growth rate shortly after birth (Dommelen et al. 2005).

Table 5.24 shows the AIC and likelihood ratio test statistics for the various Jenss Bayley models that were fit. A model with a random effect on β_3 did not coverge.

Table 5.24: Comparison of Jenss Bayley models

	Random effects	df	AIC	BIC	Test	p-value
JenssBayley1	beta0	6	43671	43714		NA
JenssBayley2	beta0 and beta1	8	42426	42483	1 vs 2	<0.001
JenssBayley3	beta0, beta1 and beta2	11	42211	42289	2 vs 3	<0.001

We can see that the random effects on β_0, β_1 and β_2 are needed in the model.

Table 5.25 and A.18 show the fixed effect estimates and random effect estimates for model 3.

Table 5.25: Jenss Bayley Model Fixed Effects

	Estimate	Std.Error	DF	t-value	p-value
beta0	69.5	0.229	7735	304	<0.001
beta1	0.621	0.00654	7735	94.9	<0.001
beta2	3	0.0112	7735	268	<0.001
beta3	-0.147	0.00229	7735	-64.1	<0.001

Figure 5.23 shows the standardised residuals of the Jenss Bayley model. We can see that they are randomly scattered around 0, indicating a good fit. Figures A.21 and A.22 show that the normality assumptions are satisfied.

Figure 5.24 shows the estimated profiles of the Jenss Bayley model. We can see that the Jenss Bayley model has an initial rapid growth that fits the data well.

5.3.5 Karlberg model

The infancy part of the Karlberg model (1987) was used to study healthy Swedish children and malnourished Pakistani children (Karlberg 1987). The height y_{ij} of child

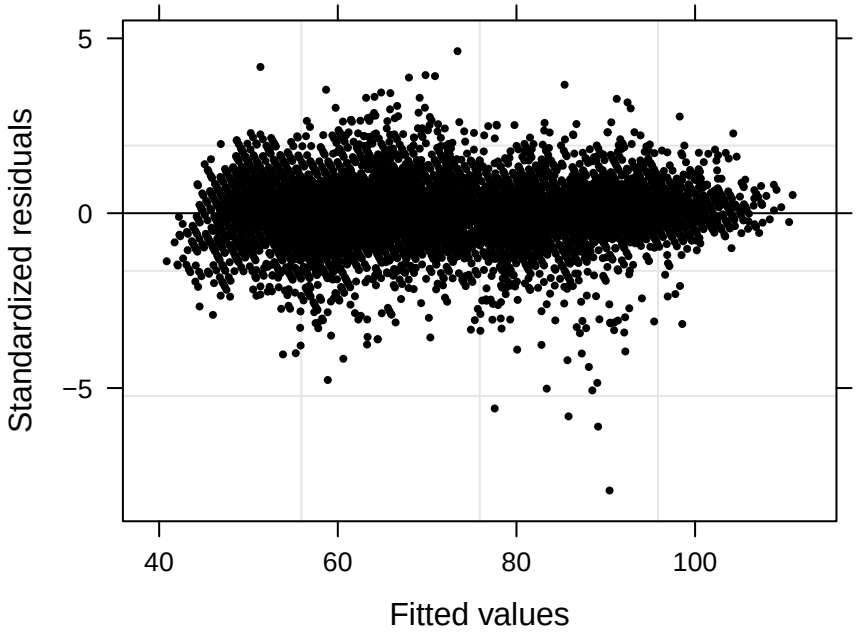


Figure 5.23: Jenss Bayley Model Residuals

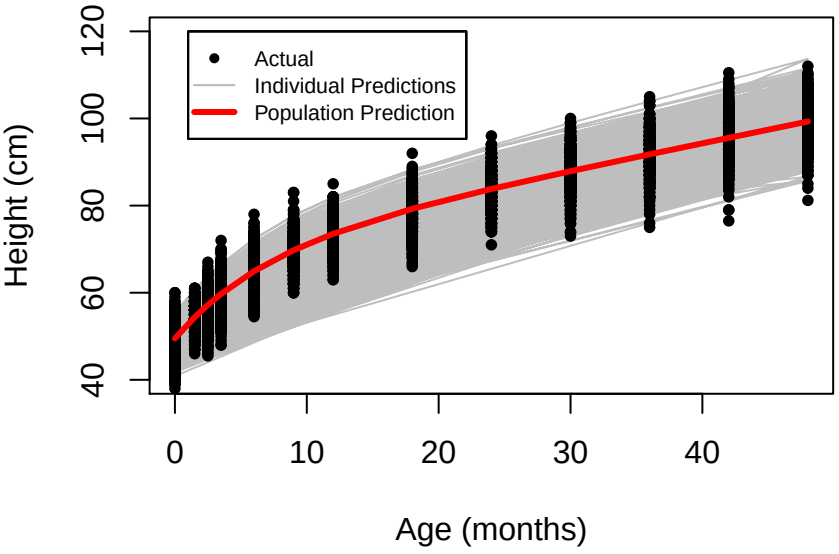


Figure 5.24: Jenss Bayley Model Estimated Profiles

i at age t_{ij} is given as

$$y_{ij} = \beta_{0i} + \beta_{1i}[1 - exp(-\beta_{2i}t_{ij})] + \epsilon_{ij}$$

where β_0 represents the birth height and $\beta_1[1 - exp(-\beta_2 t)]$ represents a rapidly decelerating growth.

Table 5.26: Comparison of Karlberg models

	Random effects	df	AIC	BIC	Test	p-value
Karlberg1	beta0	5	45481	45516		NA
Karlberg2	beta0 and beta1	7	44642	44691	1 vs 2	<0.001
Karlberg3	beta0, beta1 and beta2	10	44554	44625	2 vs 3	<0.001

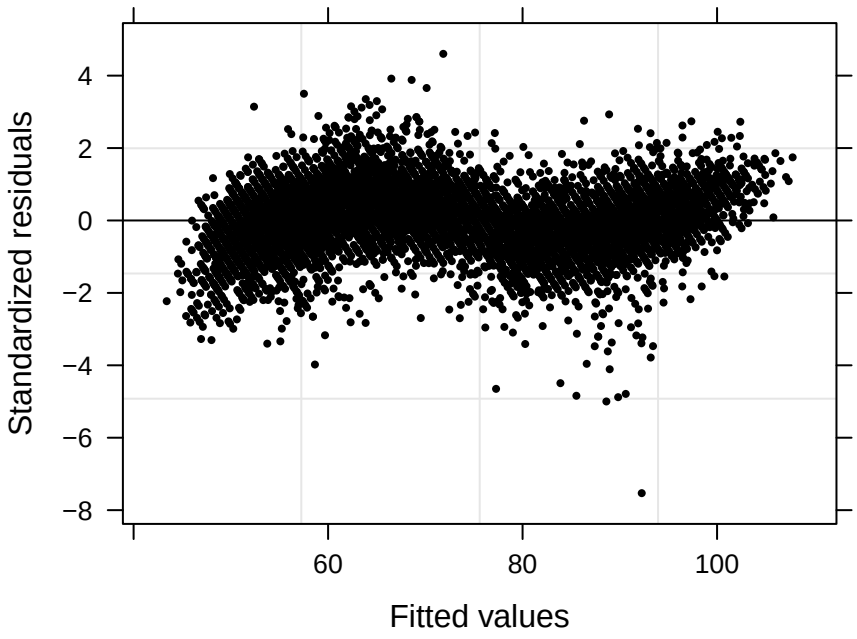


Figure 5.25: Karlberg Model Residuals

We can see that the random effects on β_0, β_1 and β_2 are needed in the model.

Table 5.27 and A.19 show the fixed effect estimates and random effect estimates for model 3.

Table 5.27: Karlberg Model Fixed Effects

	Estimate	Std.Error	DF	t-value	p-value
beta0	50.9	0.0963	7736	528	<0.001
beta1	49.9	0.275	7736	181	<0.001
beta2	0.0506	0.000557	7736	90.9	<0.001

Figure 5.25 shows the standardised residuals of the Karlberg model. We can see that they are not randomly scattered around 0, indicating a poor fit. Figures A.23 and A.24 show that the normality assumptions are satisfied.

Figure 5.26 shows the estimated profiles of the Karlberg model. We can see that the inital rapid increase in growth is not captured. The deceleration in growth is also not

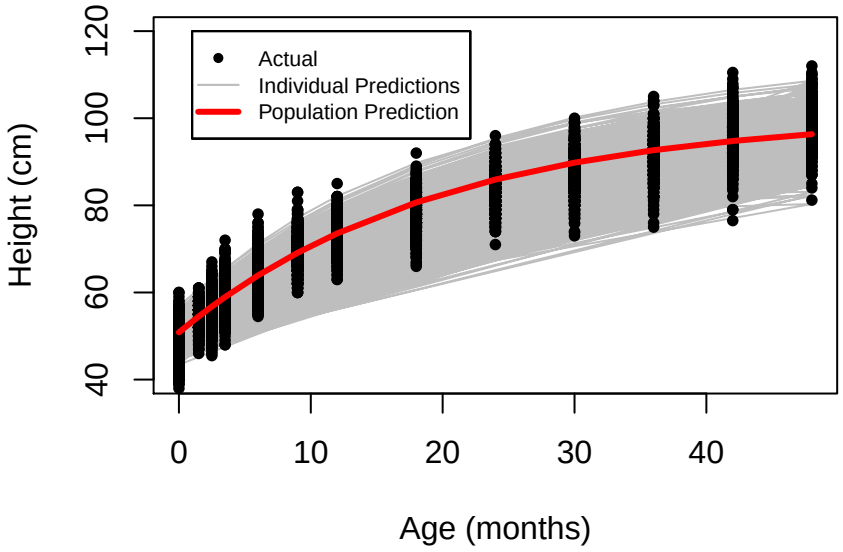


Figure 5.26: Karlberg Model Estimated Profiles

correctly captured.

5.3.6 Comparison of nonlinear models

Table 5.28 shows a comparison of the models that are nonlinear in parameters.

Table 5.28: Comparison of nonlinear mixed effect models

	df	AIC	BIC
Logistic2	7	46311	46361
Gompertz4	10	45392	45463
Exponential2	6	55988	56030
JenssBayley2	8	42426	42483
Karlberg3	10	44554	44625

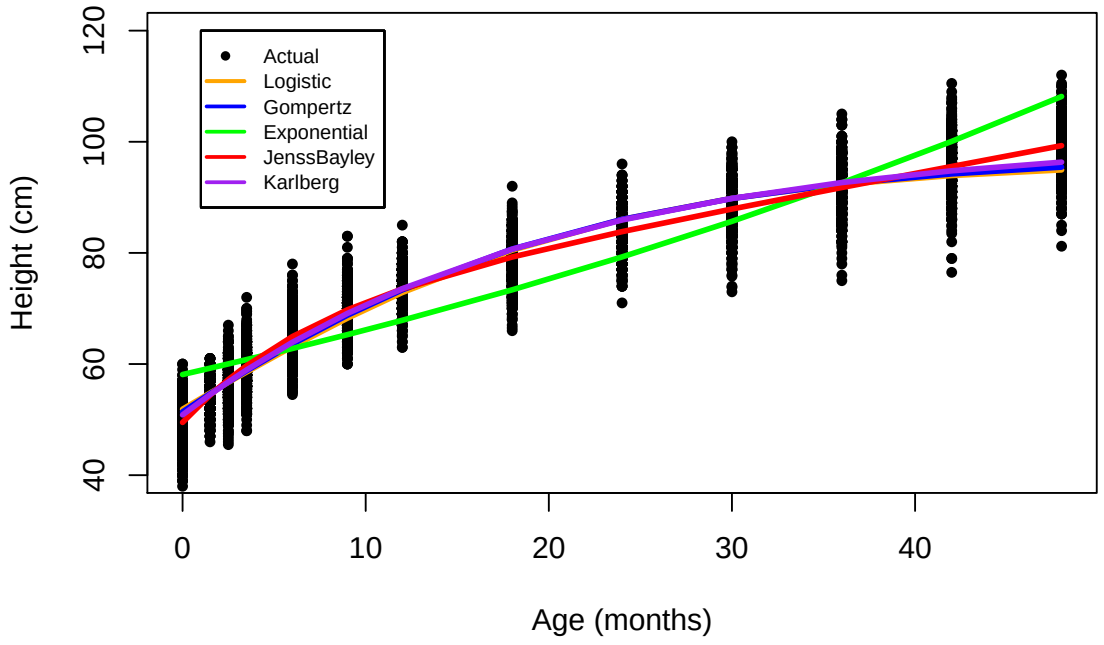


Figure 5.27: Comparison of nonlinear model estimated profiles

The AIC and BIC test statistics are smallest for the Jenss Bayley model, indicating that the Jenss Bayley is the preferred nonlinear mixed effect model. Figure 5.27 shows all nonlinear models superimposed on the observed profiles. We can see that the Jenss Bayley model is the most appropriate for capturing the fast-slow-fast growth trajectory.

5.4 Final mixed effect model choice

Next we compare the best linear (Reed2) and best nonlinear (Jenss Bayley) mixed effect models. Table 5.29 shows the AIC and BIC test statistics of these two models. We can see that the Reed2 model has the smallest AIC test statistic, but the Jenss Bayley model has a smaller BIC. Analysis of the effects of covariates upon growth is simpler if linear models are used. The Berkey-Reeds 2nd order model is therefore chosen as the model to use in order to fit the conditional mixed-effects model. Section 5.6 investigates the covariates to use in fitting a conditional model.

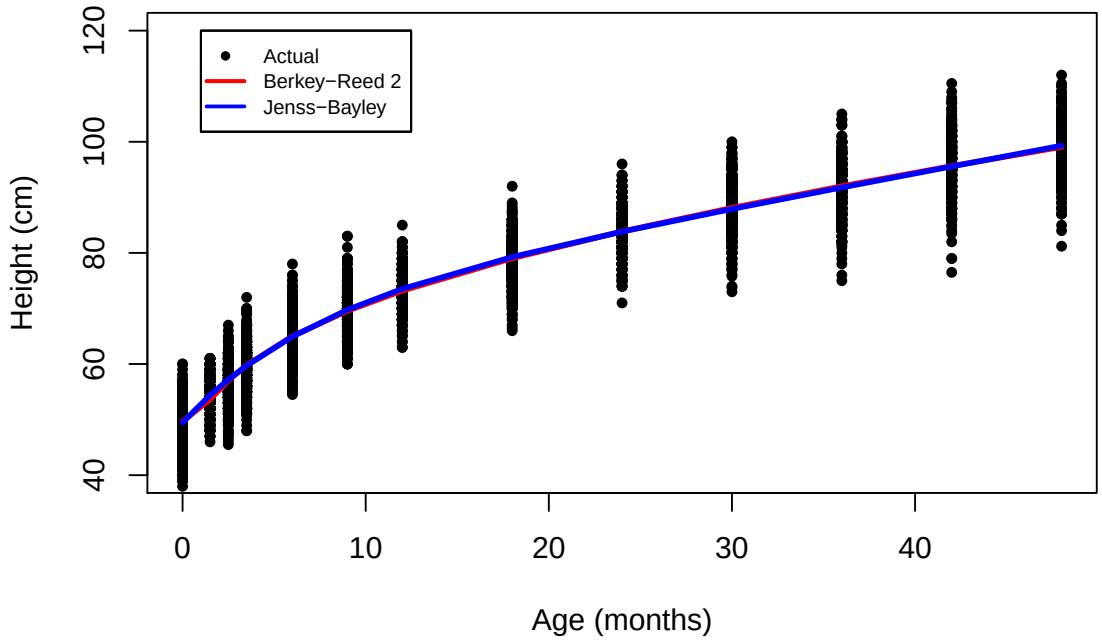


Figure 5.28: Comparison of best linear and best nonlinear model estimated profiles

Table 5.29: Comparison of linear and nonlinear mixed effect models

	df	AIC	BIC
JenssBayley3	11	42211	42289
BerkeyReed2	16	42187	42301

Figure 5.28 shows the Berkey-Reeds 2nd order model and Jenss Bayley model superimposed on the actual measurements. We can see that the two models predict very similarly. In order to further improve the model fit, correlation functions are investigated in the next section.

5.5 Correlation functions

The inclusion of subject-specific effects induces constant correlation between the within-subject measurements. However, specifying further correlation structures can improve the model fit. Correlation structures model dependence among the within-group errors. With no subject-specific effects, the within-subject errors, ϵ_{ij}, are assumed to be uncorrelated, i.e. $corr(\epsilon_{ij}, \epsilon'_{ij}) = I$, where I is the identity matrix. However, in practice, measurements that are closer together in time will have similar departures from that child's growth trajectory (Steele 2008). The variability in height remains constant regardless of when it is measured.

We can extend the Berkey-Reeds 2nd order model by adding a model for the ϵ_{ij}. A general within-subject model for subject i at time j:

$$corr(\epsilon_{ij}, \epsilon'_{ij}) = h[d(\boldsymbol{p_{ij}}, \boldsymbol{p'_{ij}}), \boldsymbol{\rho}]$$

where $\boldsymbol{p_{ij}}$ is a position vector, $d(\boldsymbol{p_{ij}}, \boldsymbol{p'_{ij}})$ is some distance between two positions, $\boldsymbol{\rho}$ is a vector of correlation parameters, and $h()$ is a correlation function with values between -1 and 1.

The following correlation structures were fit and the resulting models compared to the unconditional Berkey Reed 2nd order model:

- Constant correlation.Up to now we have used this correlation structure. This assumes an equal correlation among all within-group errors, corresponding to the correlation model:

$$corr(\epsilon_{ij}, \epsilon_{ij'}) = \rho \ \forall \ j \neq j', \quad i.e. \ h(k, \rho) = \rho, \ \ k = 1, 2, ...,$$

- General correlation structure. This allows each correlation in the data to be represented by a different parameter, corresponding to the correlation function:

$$h(k, \rho) = \rho_k, \ \ k = 1, 2, ...,$$

- Autoregressive correlation structure (AR1). The AR1 structure has homogeneous correlations that decline exponentially with distance. This means that

two measurements next to each other are more highly correlated than height measurements further apart.

The correlation function, where ϕ represents the lag-1 correlation, is:

$$h(k, \phi) = \phi^k, \;\; k = 1, 2, ...,$$

It is realistic to assume that the correlation between two observations decreases, in absolute value, with distance. Therefore the constant correlation is often too simplistic. On the other extreme, the general correlation structure can lead to overparameterized models. The autoregressive structure of order 1 (AR1) model was found to be the most appropriate. The resulting correlation parameter was $\phi = 0.0569$.

Table 5.30 compares the Berkey-Reeds 2nd order model with and without the AR1 correlation structure. We can see that inclusion of the correlation structure slightly improves the AIC values but not the BIC value. The likelihood ratio test statistic shows that the additional correlation parameter is not required. We therefore choose to continue with the conditional model without the correlation parameter.

Table 5.30: Comparison of model with and without correlation structure

	Model	df	AIC	BIC	Test	p-value
BerkeyReed2	1	16	42187	42301		NA
BerkeyReed2.AR_corr	2	17	42180	42301	1 vs 2	0.00278

5.6 Conditional Mixed effect models

The objectives of analysing important covariates is two-fold:

- To assess the effects of predictors of height.
- To build a flexible yet parsimonious model for prediction in a longitudinal data setting.

The random effects in the mixed-effects model account for individual deviations in the parameters among groups. These deviations may partially be explained by differences in covariate values among groups. Including covariates in the model to explain inter-group variation can lead to simplications in the random effects model and allows for a better understanding of the system producing the response (W. Johnson, Balakrishna, and Griffiths 2014).

Inclusion of covariates is primarily done through the fixed effects design matrix. Covariates are screened for selection using plots of the estimated random effects against available covariates. The chosen covariates are incorporated into the model by adding columns to the fixed effects design matrix, with resulting estimated fixed effects being tested for significance. We can then determine if certain random effects should be included or eliminated from the model after allowing for covariates.

Some children had missing baseline covariates and were therefore excluded from the covariate analysis. This resulted in 841 children in the conditional model. To show that the children with missing covariates did not exhibit different growth trajectories to those with no missing covariates, figure 5.29 shows the height profiles of all babies with the colour indicating if covariates are missing. Figure 5.30 shows the mean height profile comparison of these two groups.

5.6.1 Investigation of covariates

The model building approach to include covariates in the model is to start with the unconditional model to explain inter-group variability, and use plots of the estimated random effects b_i versus the potential covariates to identify patterns. Since the random effects allow for individual departures from the population mean, plotting the estimated random effects against potential covariates provides meaningful information in the

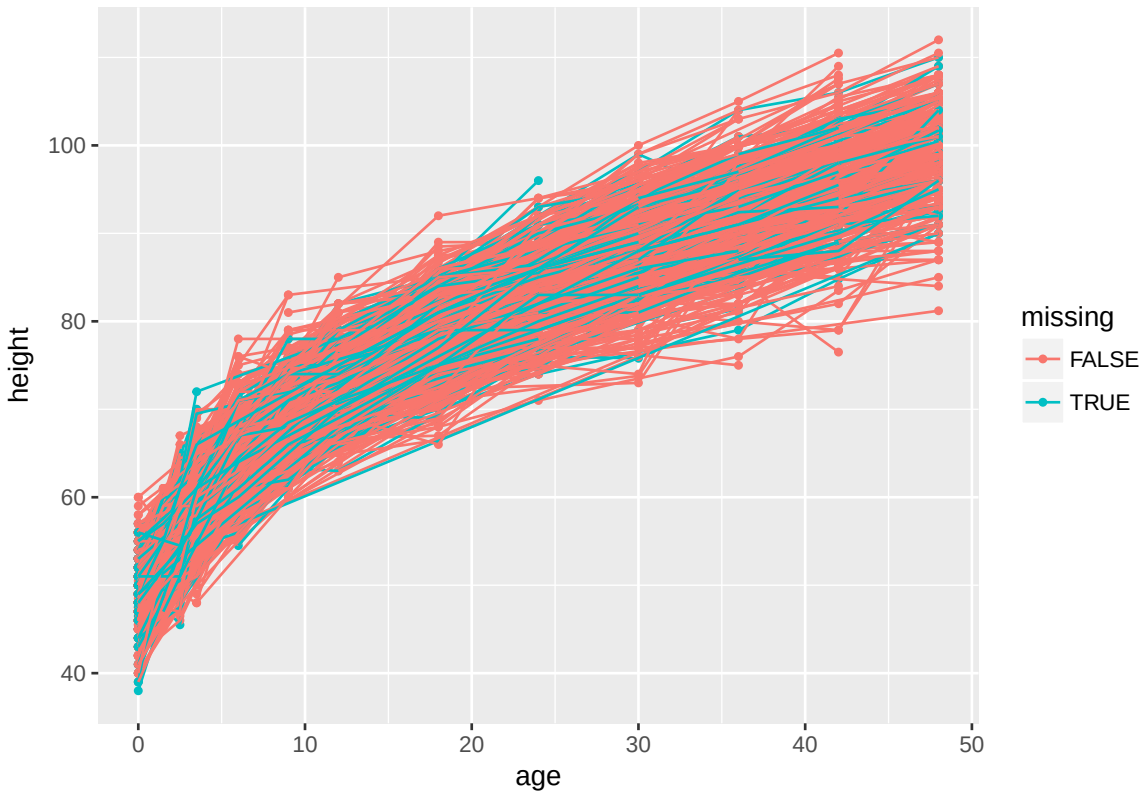


Figure 5.29: Height profile comparison of children with missing covariates

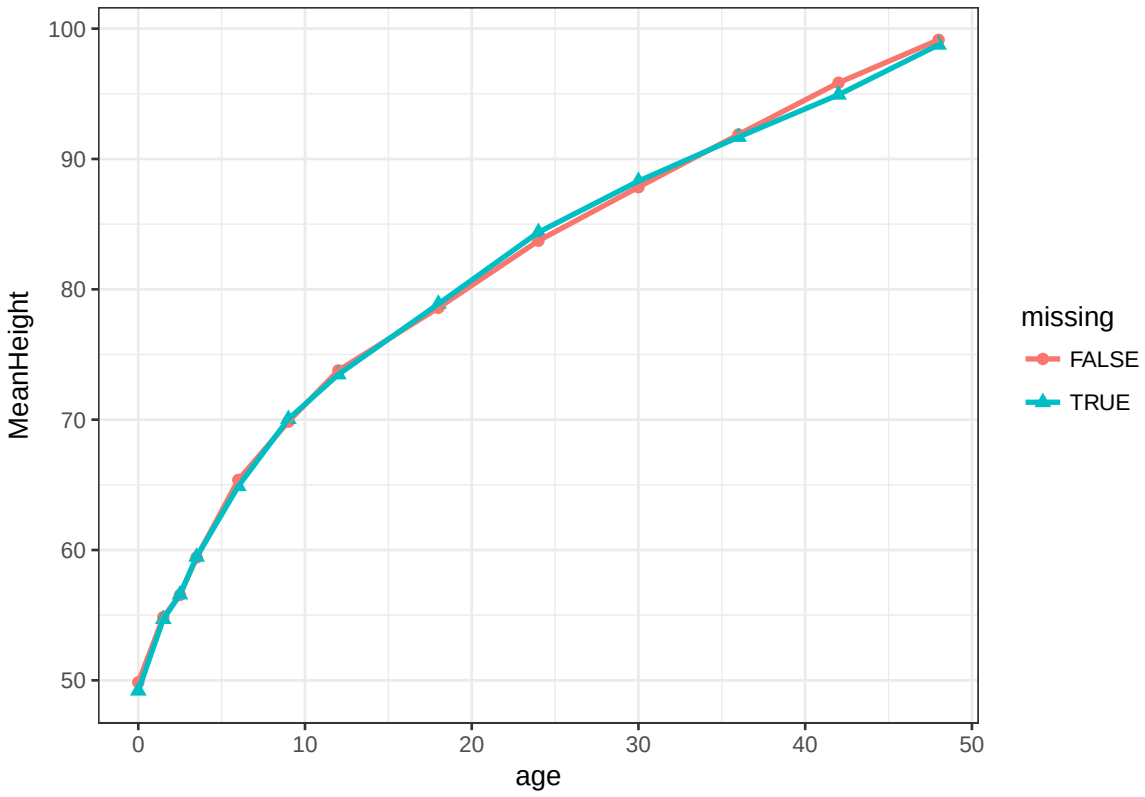


Figure 5.30: Mean height profile comparison of children with missing covariates

model building process. A systematic pattern will indicate that the covariate should be included in the model. The covariate is then tested for inclusion in the model.

The unconditional Berkey-Reed 2nd order model for the height y_{ij} of child i at age t_{ij} is

$$y_{ij} = \beta_{0i} + \beta_{1i}t_{ij} + \beta_{2i}ln(t_{ij}+1) + \beta_{3i}\frac{1}{t_{ij}+1} + \beta_{4i}\frac{1}{(t_{ij}+1)^2} + \epsilon_{ij}$$

where $\beta_{ki} = \beta_k + b_{ki}$ for $k = 1, ...4$

b_{ki} are the subject effects

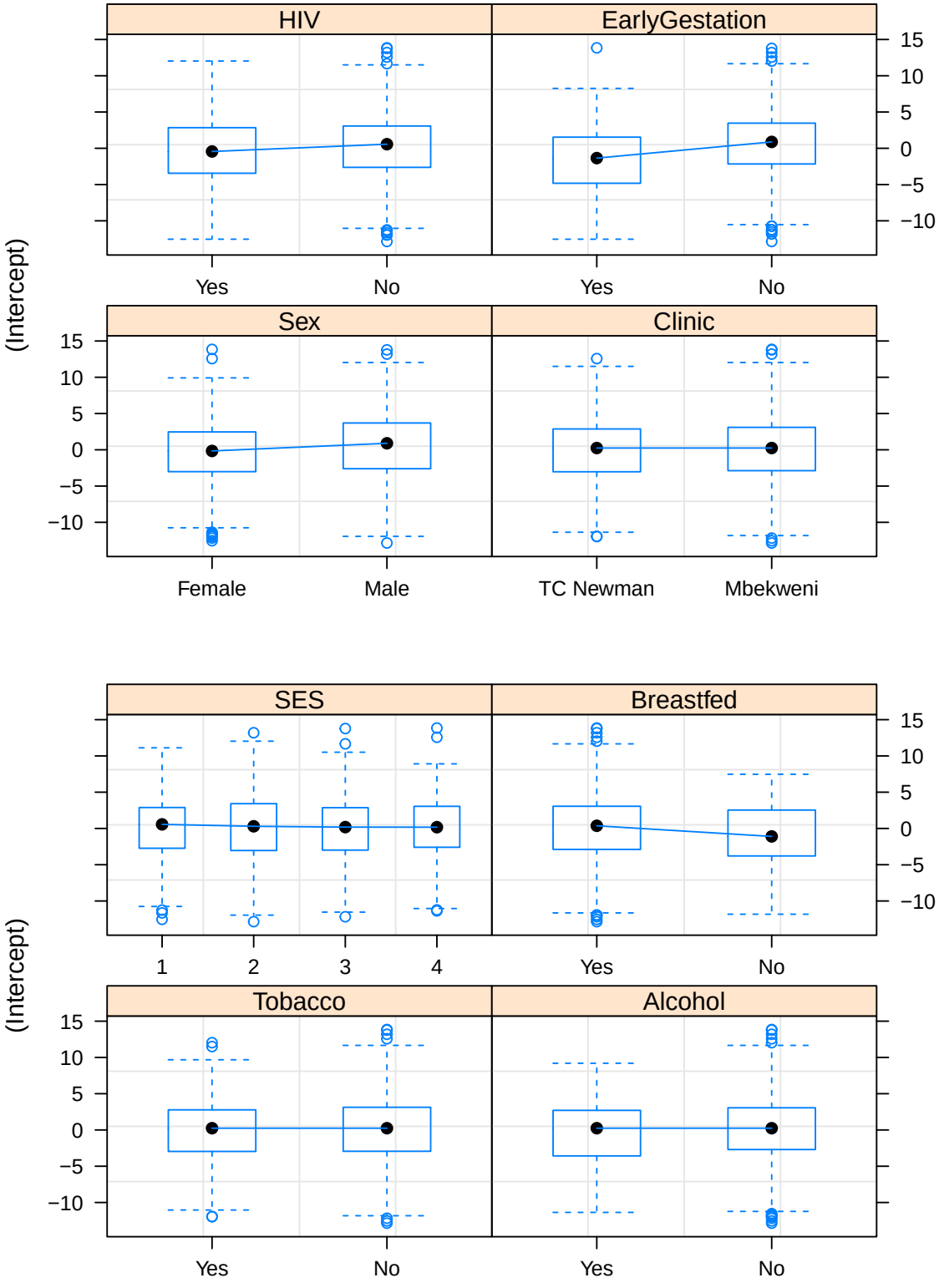
$\boldsymbol{\epsilon_i \sim N(0, \sigma_\epsilon^2 I_n)}$

Relationships between the estimated random effects and categorical covariates are displayed using boxplots.

5.6.1.1 Intercept random effects

Figure 5.31 shows the relationship between the covariates and the intercept random effect. We can see that differences appear small however the following observations are made:

- An early gestational age (<38 weeks) is associated with a smaller birth height. In particular, the intercept is low for very early gestational ages. It then increases with gestational age and decreases after 40 weeks.
- Females have a smaller birth height.
- A positive HIV status results in a smaller birth height.
- Maternal age does not seem to affect the birth height.
- Maternal height does not seem to have an effect on the intercept term for mother's below 170cm tall. For mothers above 170cm, the birth height of their babies is generally larger.


HIV
EarlyGestation
Sex
Clinic
SES
Breastfed
Tobacco
Alcohol
(Intercept)
15
10
5
0
−5
−10
Yes
No
Female
Male
TC Newman
Mbekweni
1
2
3
4

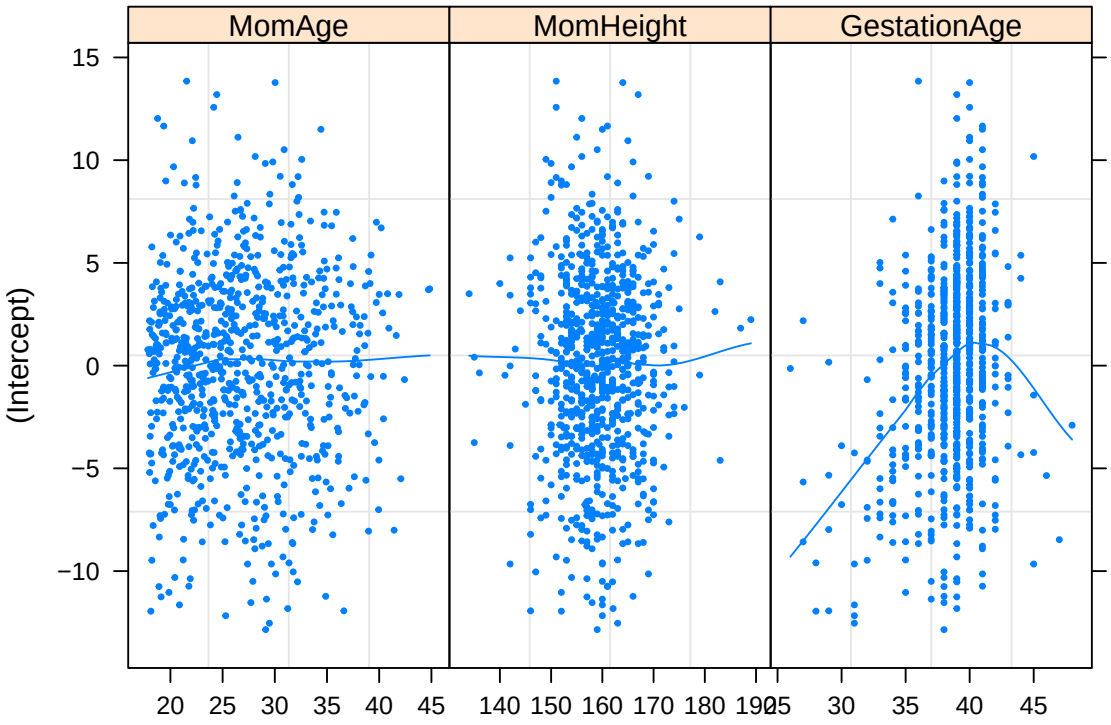
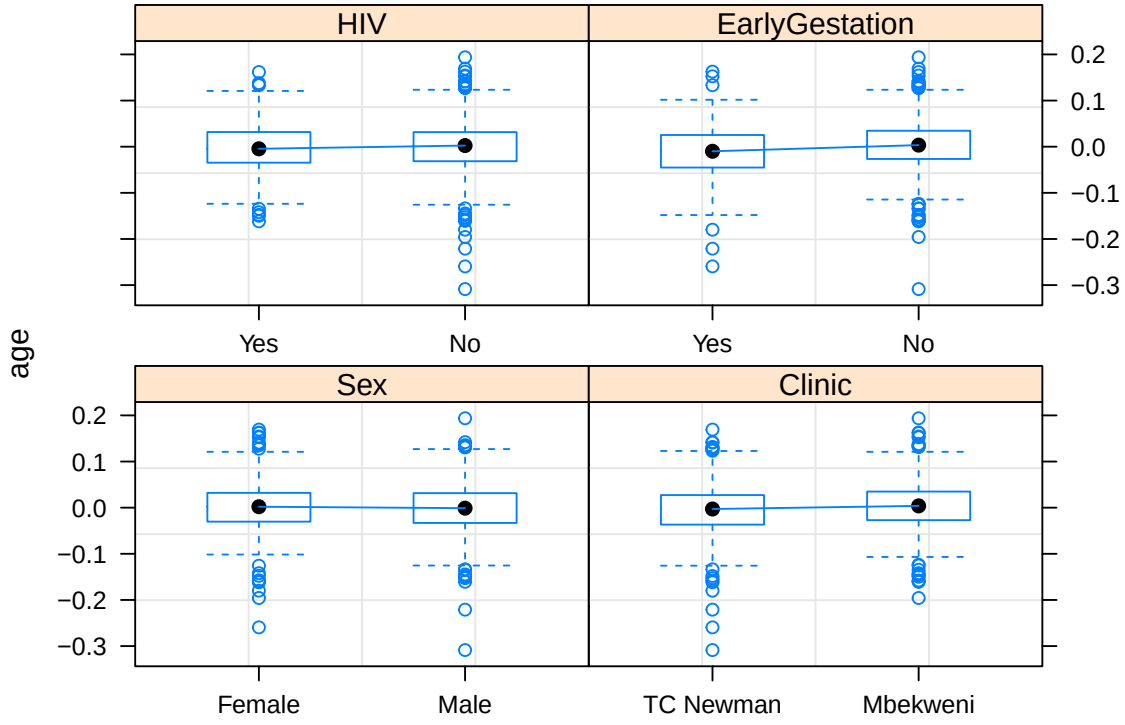


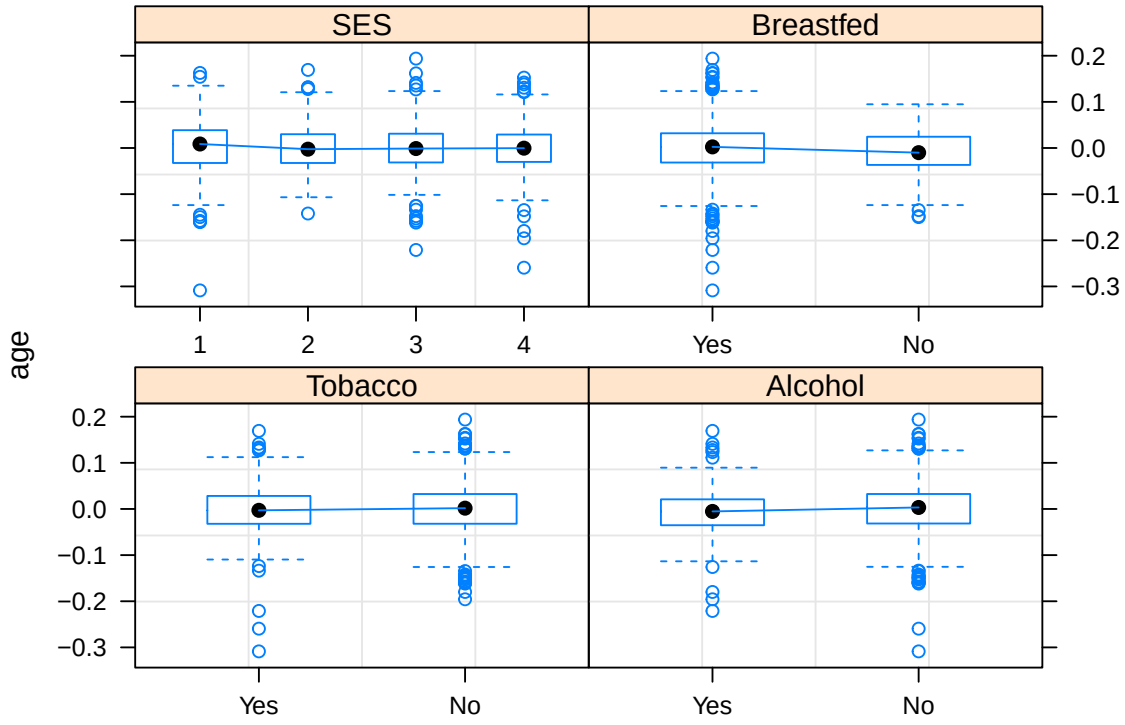
Figure 5.31: Intercept random effects

5.6.1.2 Linear age random effects

Figure 5.32 shows the relationship between the linear growth and covariates. We can see that the linear component of growth (the age term) is not strongly affected by the categorical covariates. However we note the following small differences:

- An early gestational age is associated with a smaller linear increase in height.
- TC Newman babies seem to have a smaller linear growth rate.
- Alcohol and tobacco during pregnancy may be associated with a smaller linear growth rate.
- Males have a lower linear growth than females.
- A positive HIV status is associated with a smaller linear growth.
- The lowest socio-economic quartile has the greatest linear growth.
- Linear growth increases with maternal age, indicating that older mothers have babies that grow at a faster linear rate than younger mothers.
- The linear growth in children increases with gestational age but decreases again after 40 weeks.
- Maternal height does not seem to have an effect on the linear growth in child height.


HIV
EarlyGestation
Sex
Clinic
age
0.2
0.1
0.0
−0.1
−0.2
−0.3
Yes
No
Female
Male
TC Newman
Mbekweni


SES
Breastfed
Tobacco
Alcohol
age
0.2
0.1
0.0
−0.1
−0.2
−0.3
1
2
3
4
Yes
No

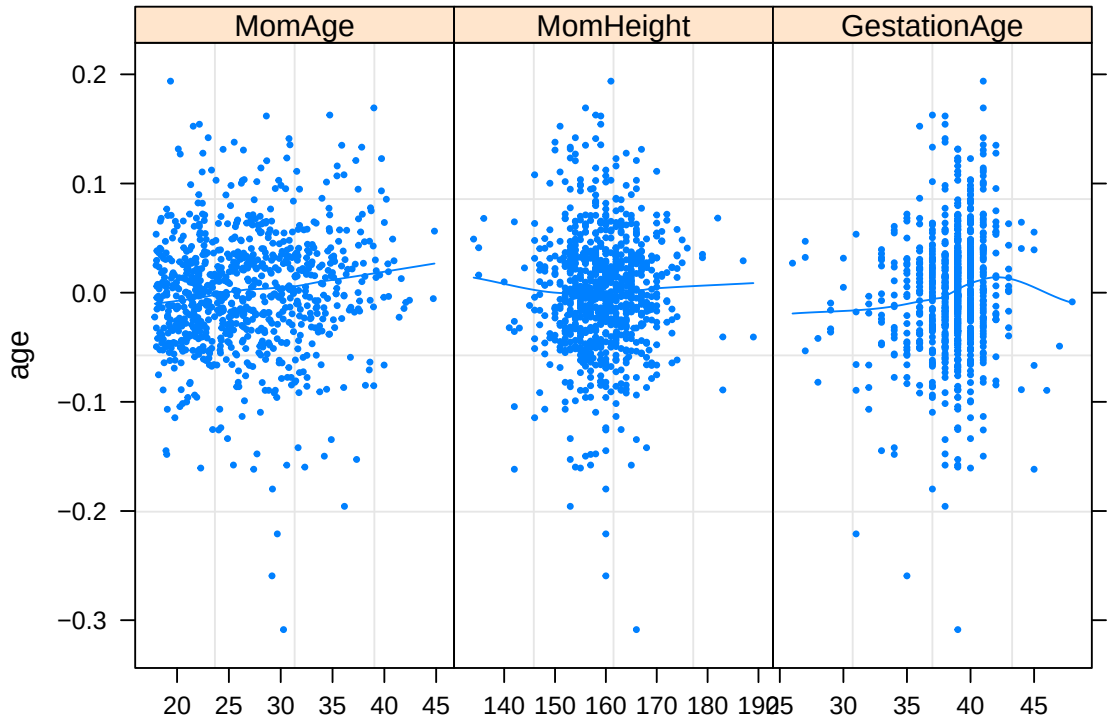
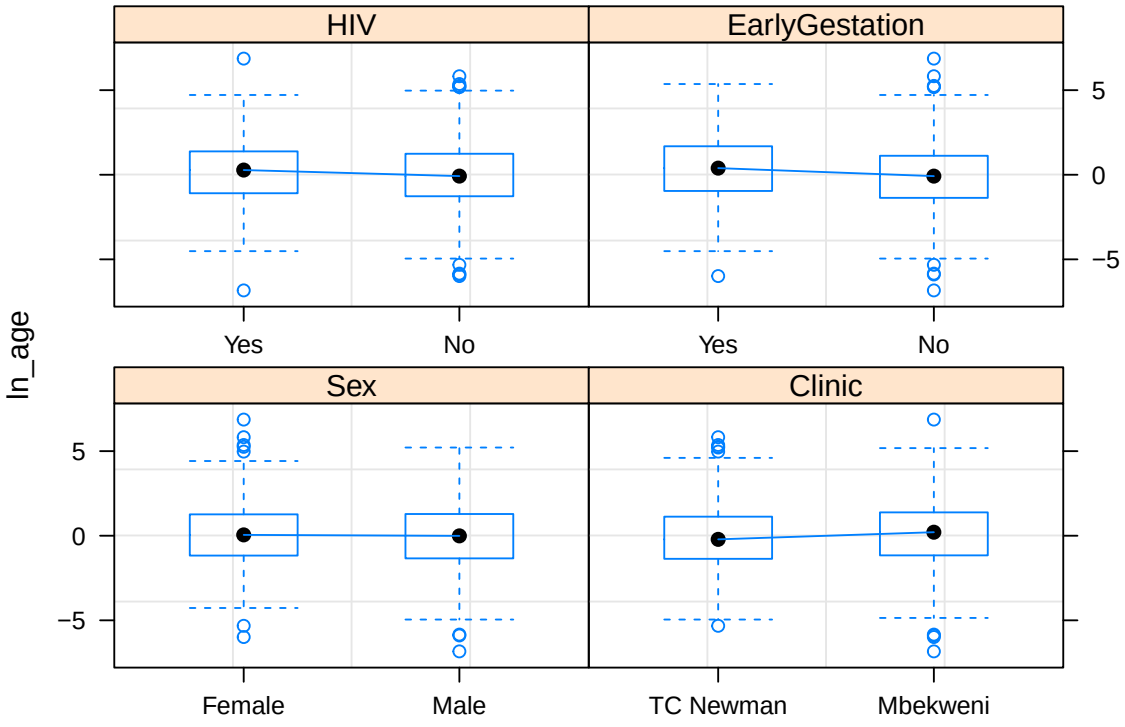


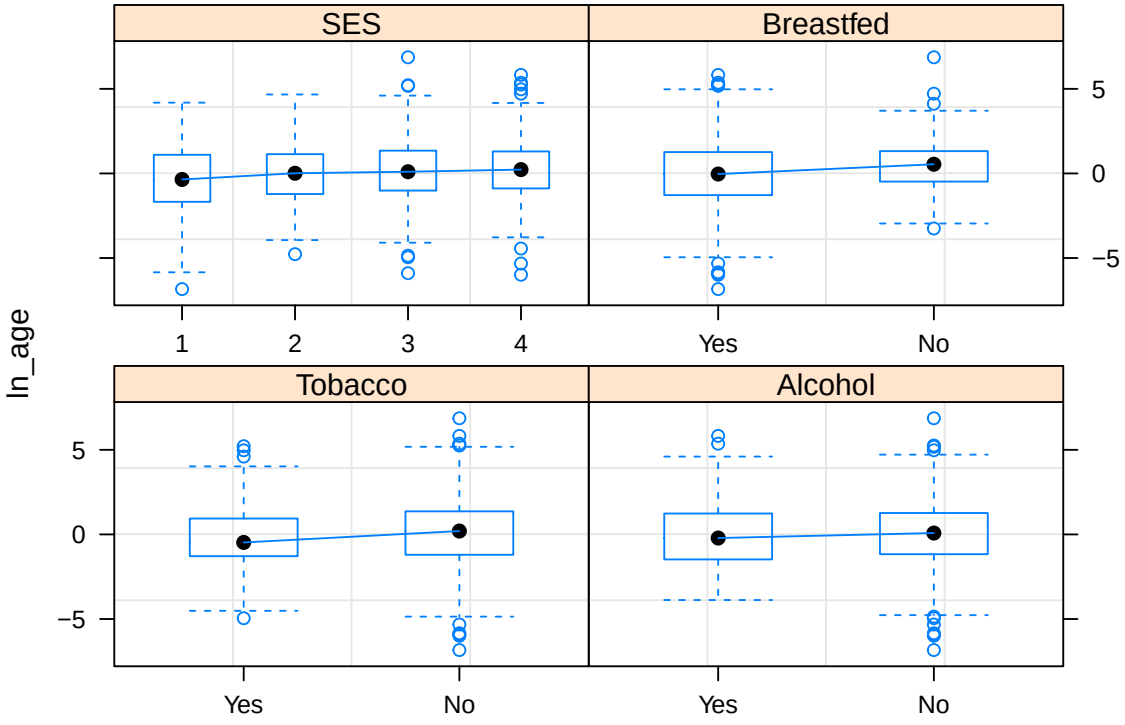
Figure 5.32: Linear growth random effects

5.6.1.3 ln(age) random effects

Figure 5.33 shows the relationship between the deceleration in growth velocity and the covariates. We note the following:

- An early gestational age is associated with a greater deceleration in growth.
- Alchol and tobacco during preganancy results in a smaller deceleration in growth.
- A positive HIV status seems to have a greater deceleration in growth.
- Babies born in Mbekweni have a greater deceleration in growth.
- The lowest socio-economic quartile has the smallest deceleration in growth.
- The deceleration in growth decreases with maternal age.
- The deceleration in growth increases with maternal height until 170cm. This indicates that the babies of shorter mothers have a smaller deceleration in growth.
- The deceleration in growth is higher for earlier gestational ages, up to 40 weeks and then increases.





5.6.1.4 1/age random effects

Figure 5.34 shows the relationship between the inflection point in growth and the covariates. We observe the following:

- An early gestational age is associated with a later inflection point.
- Alcohol and tobacco during pregnancy seems to result in an earlier inflection point.
- HIV may also delay the inflection point.
- Male babies have a later inflection point than female babies.
- TC Newman babies may have a later inflection point than Mbekweni babies.
- The lowest socio-economic quartile has the smallest inflection point.
- The inflection point decreases with gestational age up to 40 weeks and increases after
- Children with taller mothers have larger inflection points in their growth curve, up to 170cm. Thereafter the inflection point decreases.

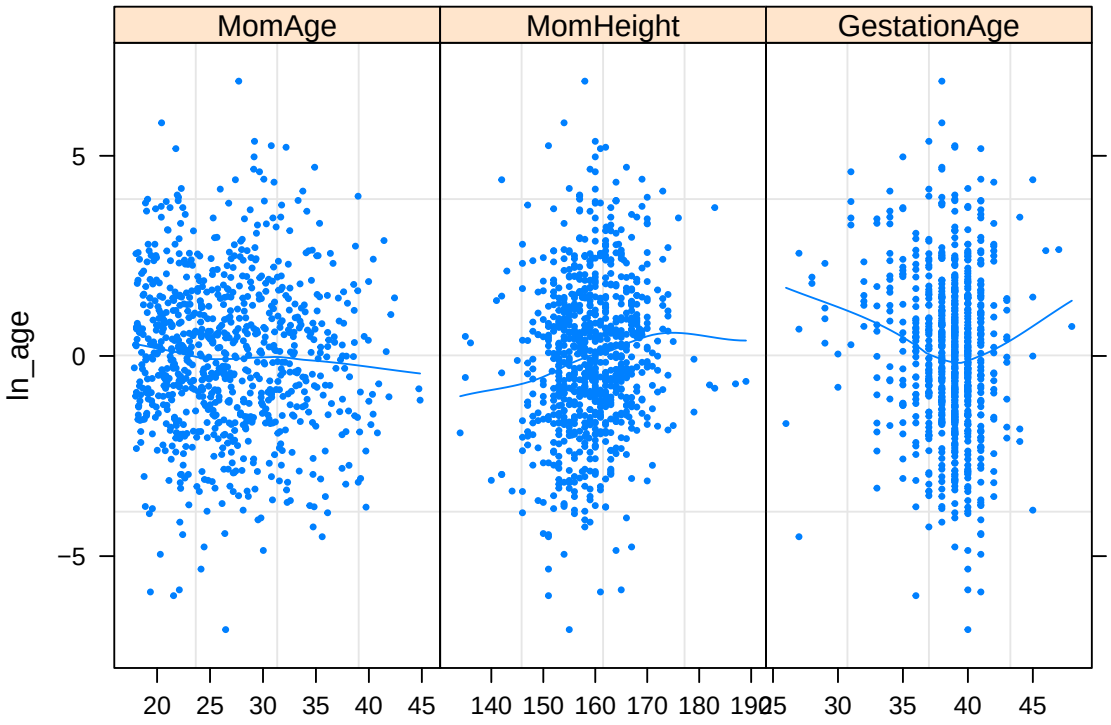
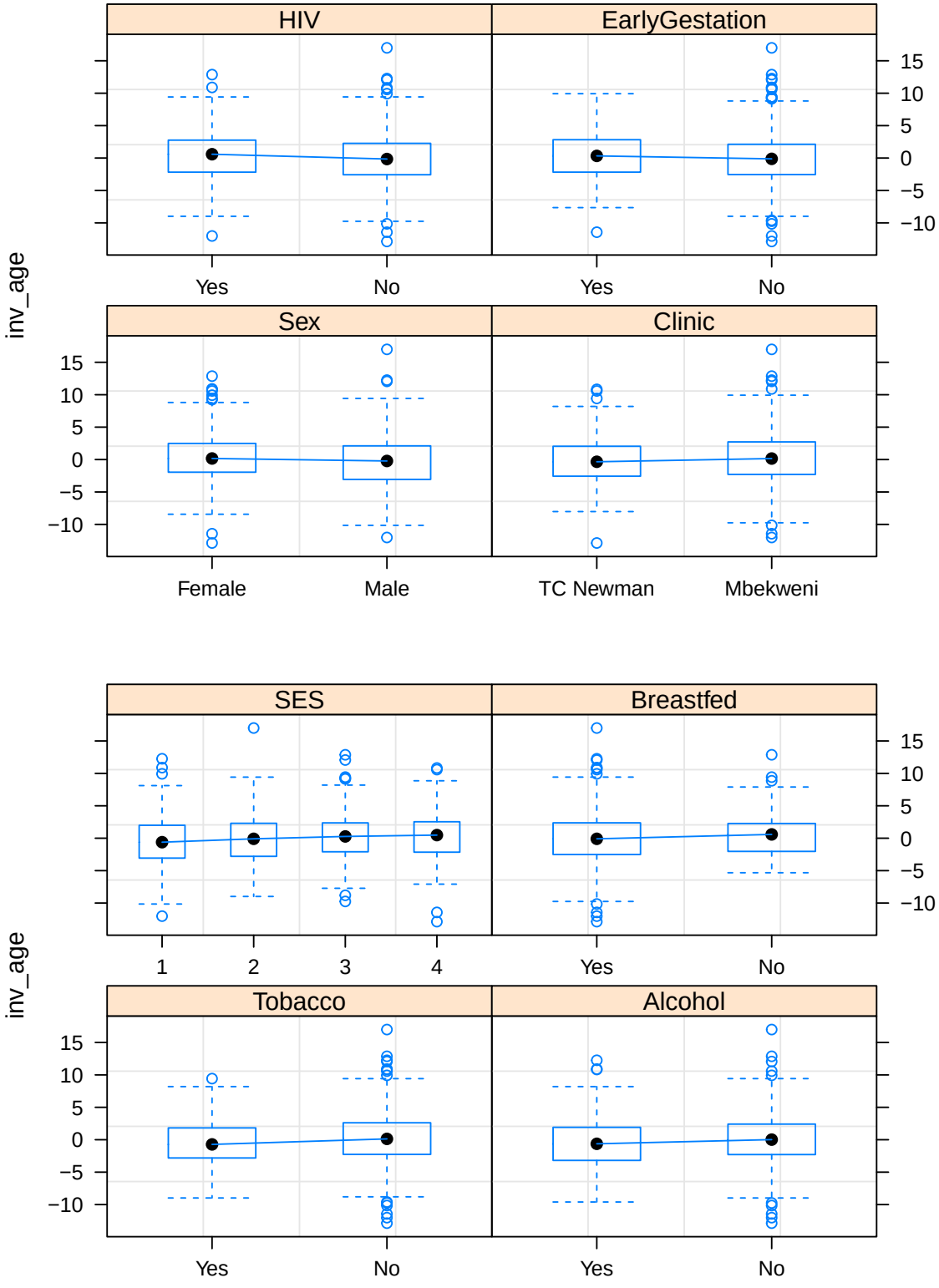


Figure 5.33: Growth deceleration random effects


HIV
EarlyGestation
Sex
Clinic
SES
Breastfed
Tobacco
Alcohol
inv_age
15
10
5
0
−5
−10
Yes
No
Female
Male
TC Newman
Mbekweni
1
2
3
4

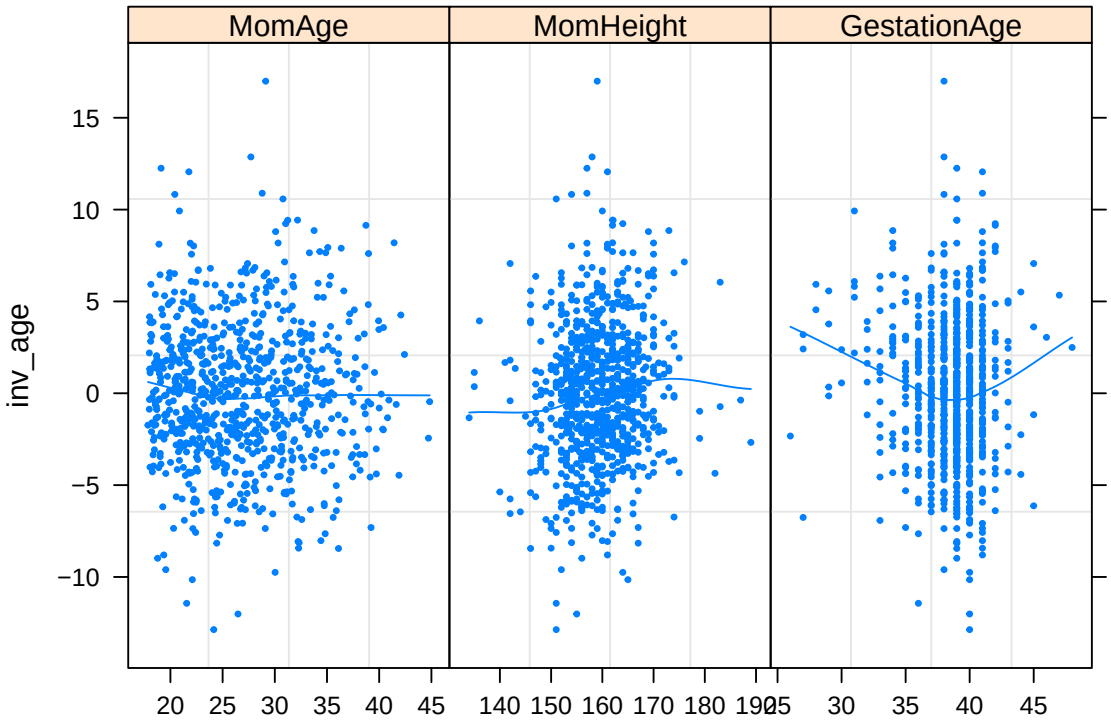


Figure 5.34: Inflection point random effects

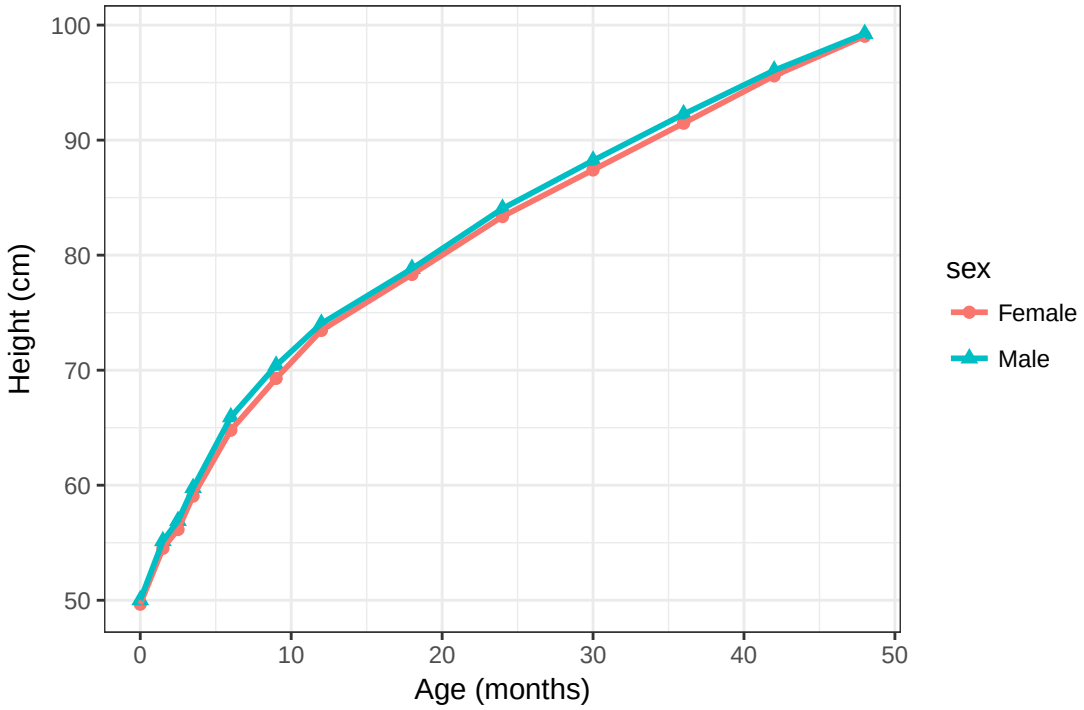


Figure 5.35: Height profiles by sex

5.6.1.5 Estimated profiles per covariates

Figures 5.35 to 5.42 show the mean height profiles per categorical covariate. We observe the following:

- The mean female height profile is below that of males.
- The mean TC Newman height profile is below that of Mbekweni.
- The lowest socio-economic quartile has the smallest height measurements throughout the ages. The highest socio-economic quartile has the largest values throughout the ages.
- Babies born before 38 weeks gestation have a lower height trajectory.
- Tobacco and alcohol during pregnancy does not affect the birth height or inital growth however the growth rate is lower for children exposed to tobacco and alcohol during pregnancy.

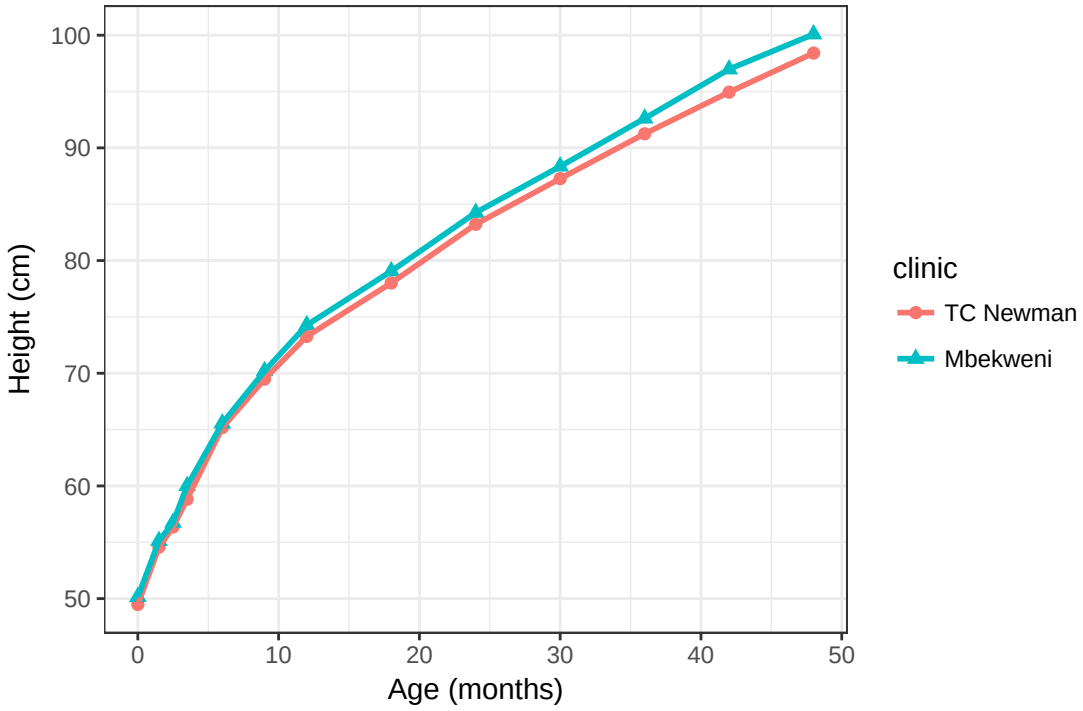


Figure 5.36: Height profiles by clinic

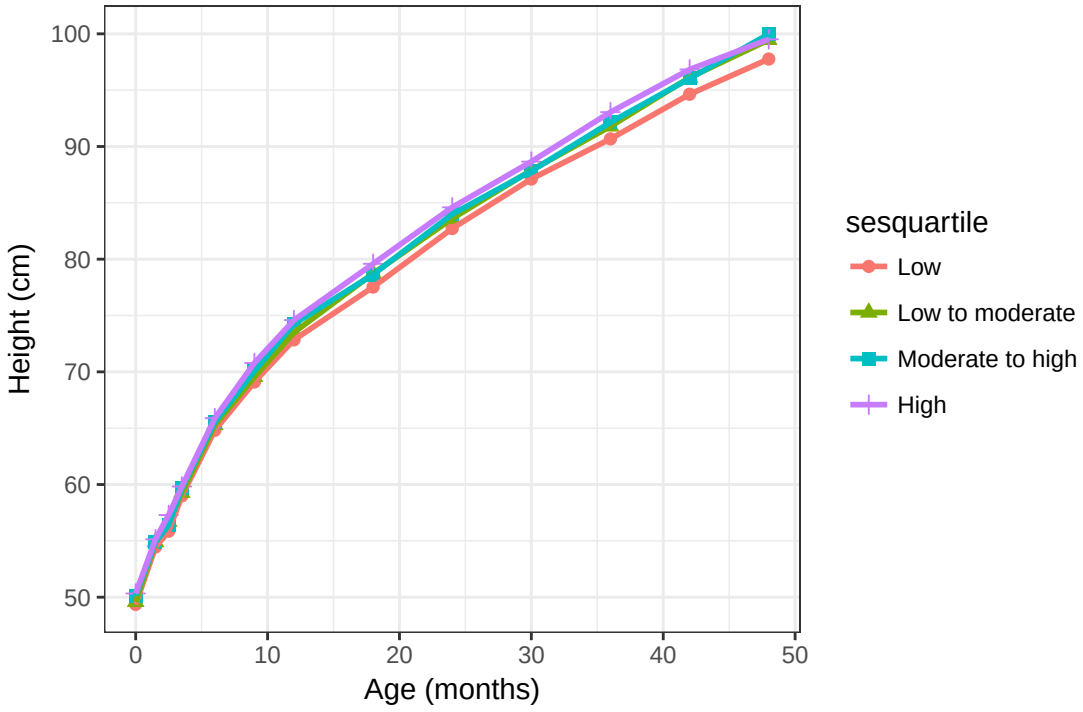


Figure 5.37: Height profiles by socio-economic quartile

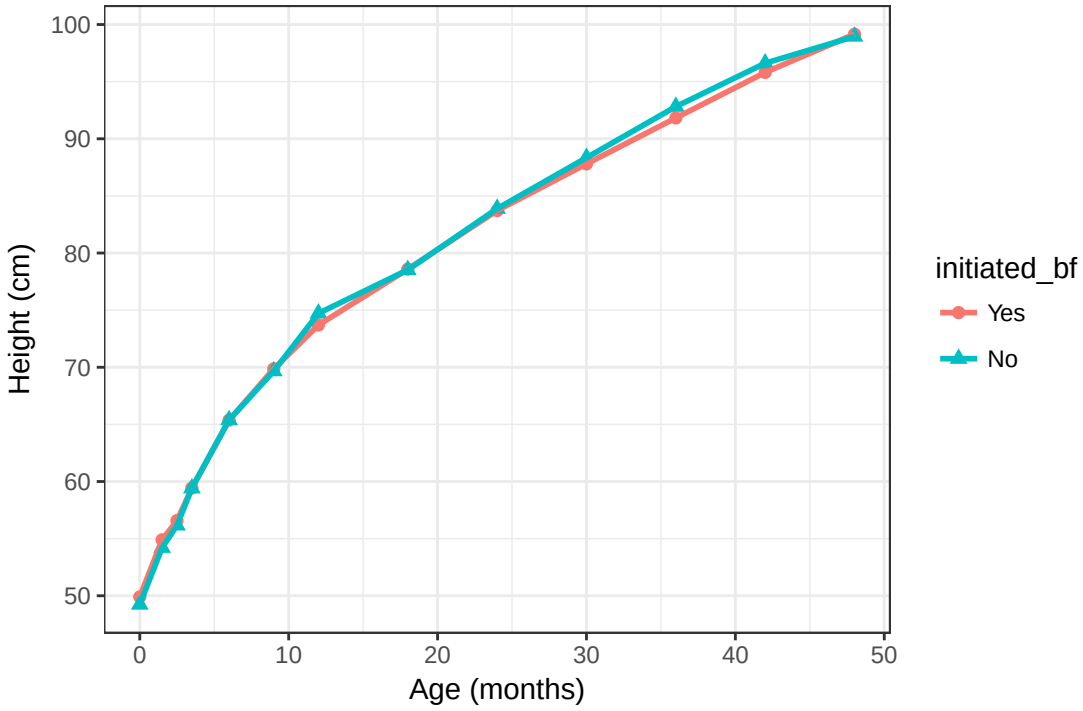


Figure 5.38: Height profiles by Breastfeeding initiation

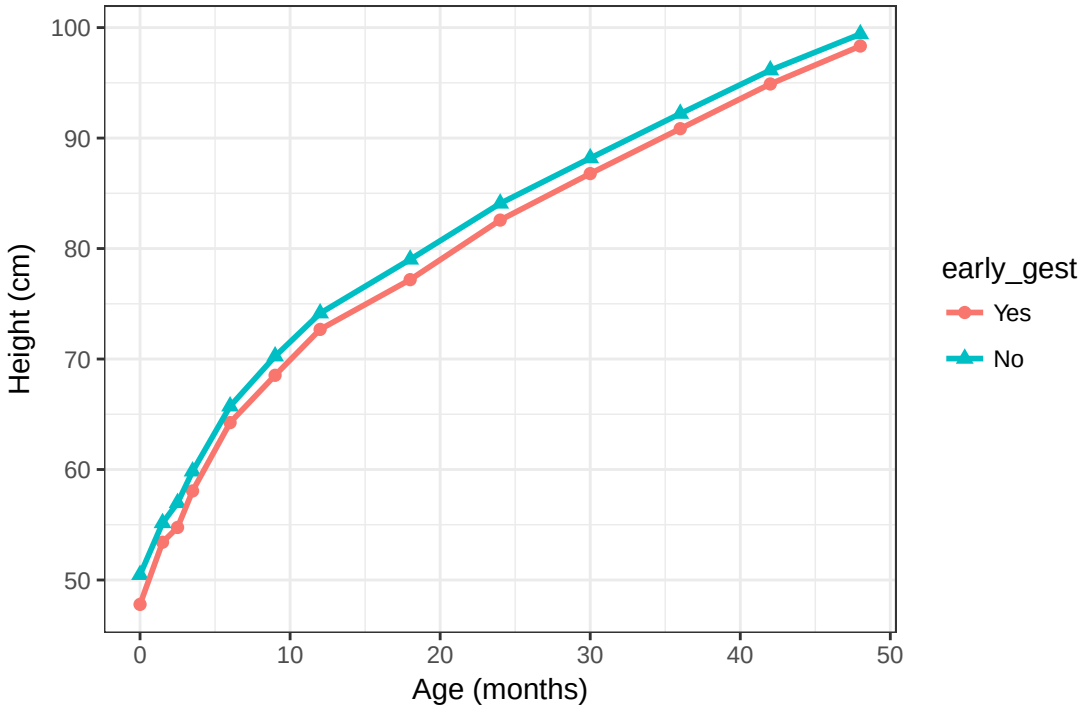


Figure 5.39: Height profiles by early gestational age

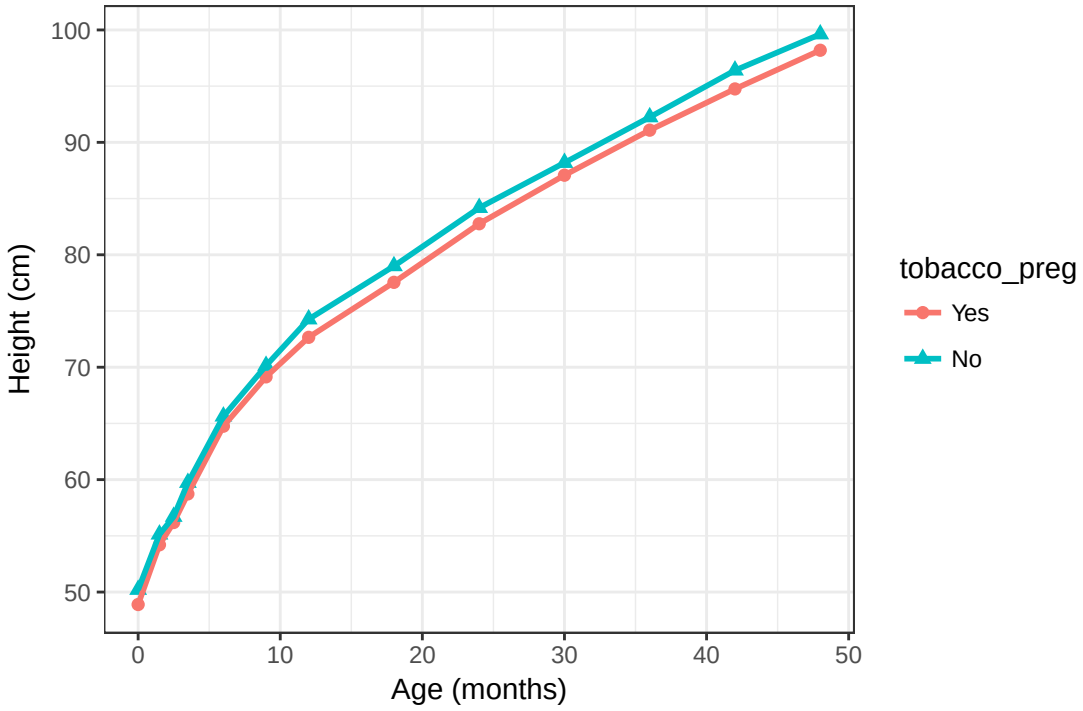


Figure 5.40: Height profiles by tobacco during pregnancy

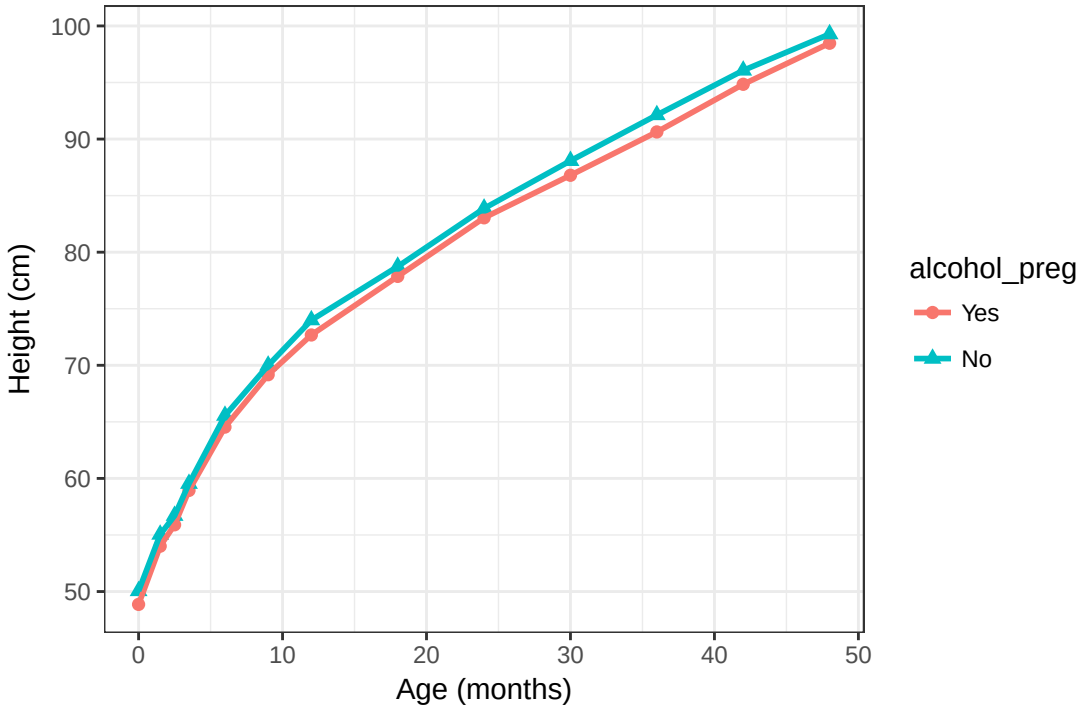


Figure 5.41: Height profiles by alcohol during pregancy

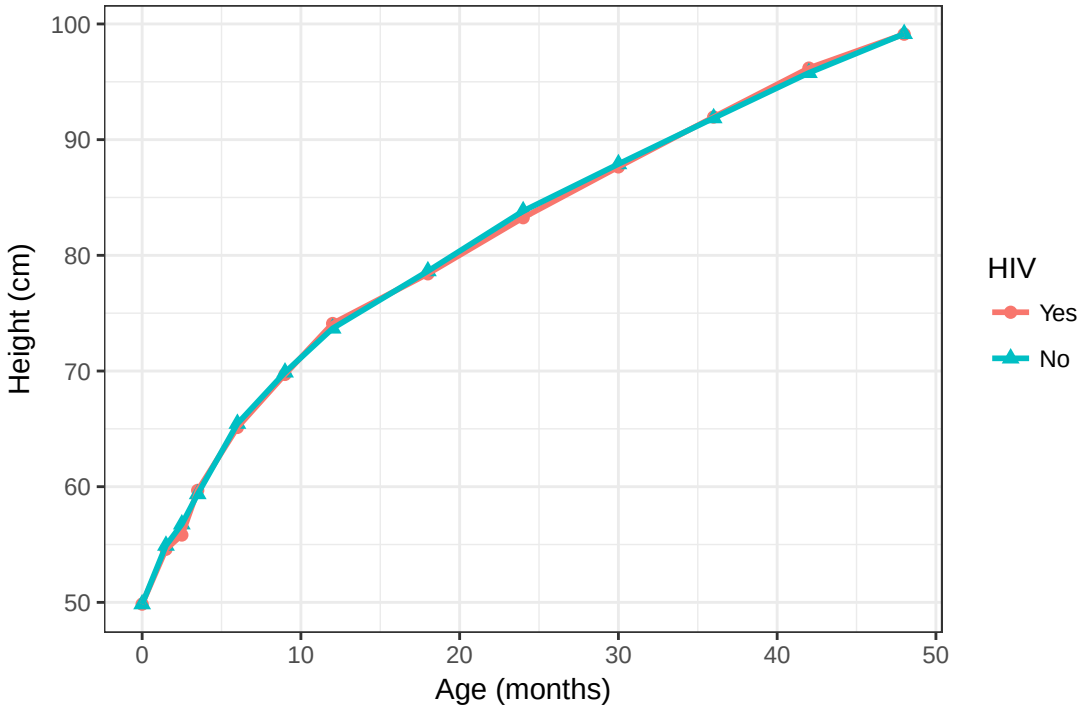


Figure 5.42: Height profiles by maternal HIV status

5.6.2 Fitting conditional models

In this section, each covariate is introduced into the Berkey-Reeds 2nd order model and tested for significance. Based on the strongest associations from the previous section, a forward stepwise approach is followed to introduce variables into the model. Covariates are added to the model one at a time. The effect of all covariates on each feature of the model are examined successively, starting with the intercept and moving to other parameters. The significance of the association between the fixed-effects and a covariate is assessed using the Wald-type tests.

$$y_{ij} = \beta_{0i} + \beta_{1i}t_{ij} + \beta_{2i}ln(t_{ij}+1) + \beta_{3i}\frac{1}{t_{ij}+1} + \beta_{4i}\frac{1}{(t_{ij}+1)^2} + \epsilon_{ij}$$

where

$$\beta_{0i} = \beta_0 + \beta_5(X_{early}) + b_{0i} \quad \beta_{1i} = \beta_1 + b_{1i} \quad \beta_{2i} = \beta_2 + b_{2i} \quad \beta_{3i} = \beta_3 + b_{3i}$$

and X_{early} is a dummy variable coded 1 for <38 weeks gestation and 0 otherwise

This model allows a different intercept for each gestation category. β_5 represents an upward or downward shift in the entire curve for children born early. Table 5.31 shows the fixed effect estimates of this model. We see that the early gestation intercept parameter p-value is very small, indicating that it is needed in the model. Table 5.32 shows the AIC comparison between the unconditional model and this model. The inclusion of the early gestation intercept parameter has decreased the AIC value. The likelihood ratio test statistic shows that the early gestation intercept parameter is needed in the model.

Table 5.31: Model1.1 Fixed Effects

	Estimate	Std.Error	DF	t-value	p-value
Intercept	46.6	1.14	6514	40.8	<0.0001
age	0.381	0.0143	6514	26.6	<0.0001
Early gestation (<38 weeks)	-2.16	0.2	840	-10.8	<0.0001
ln(age)	8.97	0.437	6514	20.5	<0.0001
1/age	-7.33	2.65	6514	-2.77	0.005598
1/(age^2)	11	1.54	6514	7.17	<0.0001

Table 5.32: Comparison of models

	df	AIC	BIC	Test	p-value
Unconditional model	16	35428	35538		NA
Early gestation intercept model	17	35342	35460	1 vs 2	<0.001

Next we progressively add covariate intercept terms and keep them in the model as needed. The graphs in the previous section serve as a guide to assess which variable to test for next. When the inclusion of a variable resulted in a lower AIC value for the model, as well as being significant at a 5% level of significance, it was kept in the model, otherwise it was removed. This approach resulted in intercept parameters for sex, clinic, tobacco during pregnancy and maternal height being included in the model. Table 5.33 shows the estimated fixed effects of this model.

Table 5.33: Model with intercept covariates fixed effects

	Estimate	Std.Error	DF	t-value	p-value
Intercept	34	2.2	6514	15.5	<0.0001
age	0.383	0.0148	6514	25.9	<0.0001
Early gestation (<38 weeks)	-1.98	0.188	836	-10.5	<0.0001
Sex: Male	0.844	0.159	836	5.31	<0.0001
Tobacco during pregnancy	-0.997	0.178	836	-5.6	<0.0001
Maternal height	0.0806	0.0117	836	6.91	<0.0001
Socio-economic quartile: Low	-0.699	0.183	836	-3.82	0.00014
ln(age)	8.89	0.441	6514	20.1	<0.0001
1/age	-8.15	2.64	6514	-3.09	0.00201
1/(age^2)	11.6	1.53	6514	7.57	<0.0001

Next we progressively add covariates through interactions with the age, $ln(age)$, $\frac{1}{age}$

and $\frac{1}{age^2}$ terms. Once again, when the inclusion of a variable resulted in a significant p-value and a lower AIC value than the previous step, it was kept in the model.

This resulted in a model with the following parameters:

$$\beta_{0i} = \beta_0 + \beta_5(X_{early}) + \beta_6(X_{sex}) + \beta_7(X_{tobacco}) + \beta_8(X_{maternalHeight}) + \beta_9(X_{HIV}) + b_{0i}$$

$$\beta_{1i} = \beta_1 + \beta_{10}(X_{clinic}) + \beta_{11}(X_{sex}) + b_{1i}$$

$$\beta_{2i} = \beta_2 + \beta_{12}(X_{early}) + \beta_{13}(X_{ses1}) + \beta_{14}(X_{maternalHeight}) + b_{2i}$$

$$\beta_{3i} = \beta_3 + \beta_{15}(X_{alcohol}) + b_{3i}$$

$$\beta_{4i} = \beta_4 + \beta_{16}(X_{sex}) + b_{4i}$$

where

$$X_{early} = \begin{cases} 1 & \text{for} \quad gestation < 38 weeks \\ 0 & \text{for} \quad gestation >= 38 weeks \end{cases}$$

$$X_{sex} = \begin{cases} 1 & \text{for} \quad males \\ 0 & \text{for} \quad females \end{cases}$$

$$X_{tobacco} = \begin{cases} 1 & \text{for} \quad tobacco\ during\ pregnancy \\ 0 & \text{for} \quad no\ tobacco\ during\ pregnancy \end{cases}$$

$$X_{alcohol} = \begin{cases} 1 & \text{for} \quad alcohol\ during\ pregnancy \\ 0 & \text{for} \quad no\ alcohol\ during\ pregnancy \end{cases}$$

$$X_{clinic} = \begin{cases} 1 & \text{for} \quad TC \ \ Newman \\ 0 & \text{for} \quad Mbekweni \end{cases}$$

$$X_{ses1} = \begin{cases} 1 & \text{for} \quad socio-economic \ \ class \ \ low \\ 0 & \text{for} \quad otherwise \end{cases}$$

This model allows the shape of the height curve to be different for different combinations of the covariates. Table 5.34, A.20 and A.21 show the fixed effects, random effects and covariance estimates of the final conditional model respectively.

Table 5.34: Final conditional model fixed effects

Parameter	Covariate	Estimate	Std.Error	p-value
Intercept		40.2	2.58	<0.0001
Intercept	Early gestation (<38 weeks)	-2.65	0.234	<0.0001
Intercept	Sex: Male	1.11	0.183	<0.0001
Intercept	Tobacco during pregnancy	-0.926	0.186	<0.0001
Intercept	Maternal height	0.0412	0.0145	0.00444
Intercept	HIV: Yes	-0.57	0.197	0.00393
age		0.404	0.0154	<0.0001
age	Clinic:TC Newman	-0.0212	0.00681	0.00191
age	Sex:Male	-0.0144	0.00725	0.04766
ln(age)		4.36	1.05	<0.0001
ln(age)	Early gestation	0.467	0.0964	<0.0001
ln(age)	Socio-economic quartile: Low	-0.356	0.0759	<0.0001
ln(age)	Maternal height	0.0281	0.00599	<0.0001
1/age		-7.82	2.65	0.00314
1/(age)	Alcohol during pregnancy	-0.738	0.289	0.01059
1/(age^2)		11.7	1.54	<0.0001
1/(age^2)	Sex: Male	-0.62	0.231	0.00726

This model can be interpretated as follows:

- Having an early gestation age decreases the growth curve intercept (birth height)

by an estimated 2.65cm on average.

- Male babies have a 1.11cm higher birth height on average.
- Tobacco during pregnancy lowers the birth height by an estimated 0.926cm on average.
- Taller mothers have babies with a higher birth height, with an estimated 0.0412cm increase per 1cm increase in maternal height.
- A positive HIV status lowers the birth height by an estimated 0.57cm on average.
- TC Newman babies have a smaller linear growth rate.
- Male babies have a smaller linear growth rate.
- Having an early gestation age increases the deceleration in growth velocity.
- Belonging to the lowest socio-economic class decreases the deceleration in growth velocity.
- Taller mothers have babies with a higher deceleration in growth velocity.
- Alcohol during pregnancy is associated with a smaller inflection point of the growth curve.
- Male babies have a larger second inflection point.

To assess the fit of this model, we examine plots of residuals. Figure 5.43 shows the standardised residuals versus the fitted values. We can see that the residuals are randomly scattered around 0, indicating a good fit. Figure 5.44 shows a histogram of the residuals. We can see that the residuals are normally distributed. Figure 5.45 shows qq-plots of the random effects. We can see that the normality assumptions are satisfied.

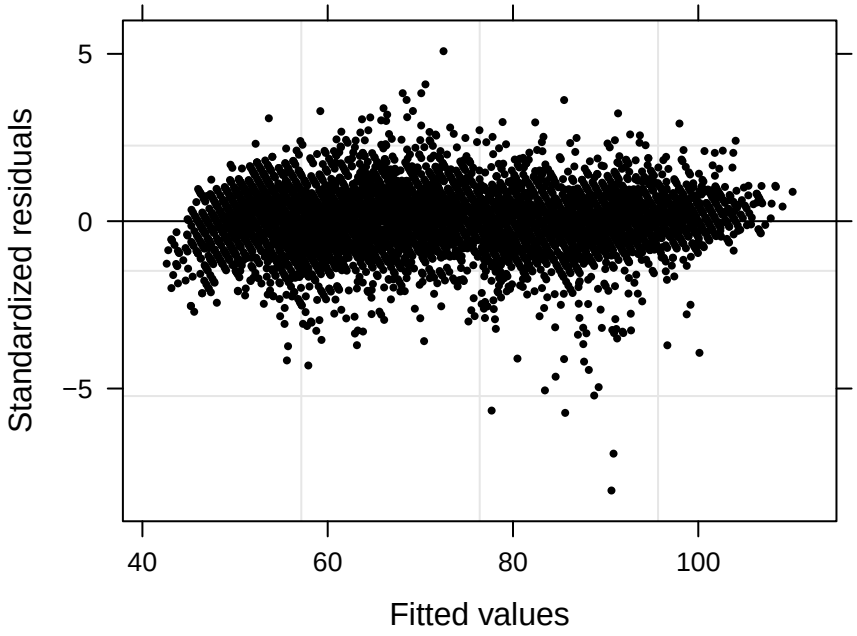


Figure 5.43: Conditional Reed2 model Residuals

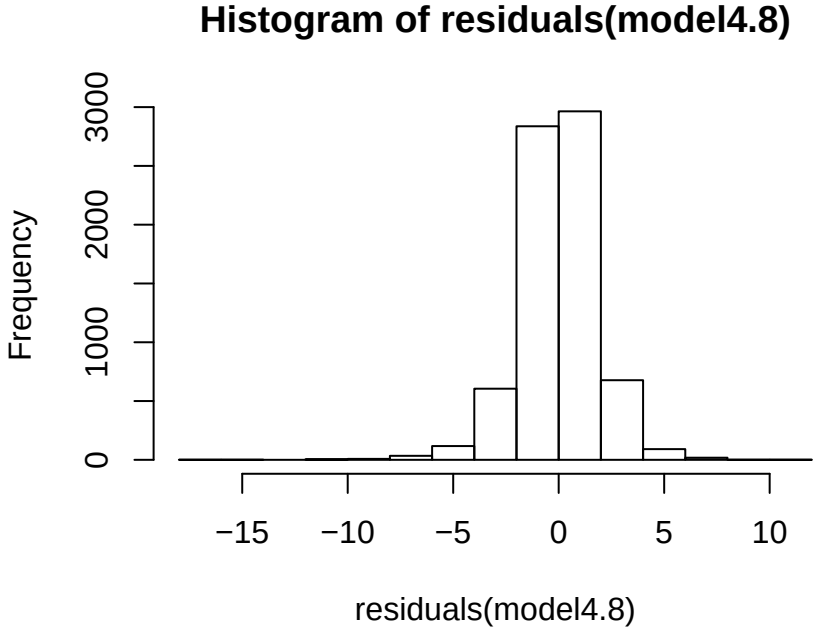


Figure 5.44: Conditional model residuals

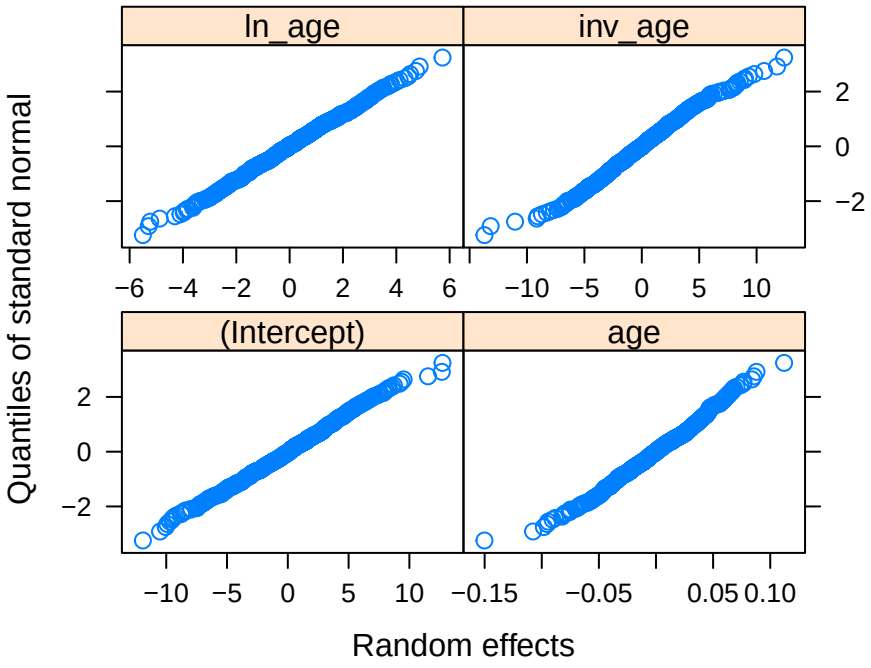


Figure 5.45: Conditional model random effects normal plot

6 Neural Networks

6.1 Introduction

Neural networks offer a more flexible modeling approach compared to mixed-effects models. They allow nonlinear models to be fit without requiring knowledge about the inter-relationship between variables. The model learns automatically from the training data to estimate the parameters (weights), unlike statistical models where the actual form of the model has to be provided (Maity and Pal 2013).

However a standard neural network does not allow for any correlation between observations. For longitudinal data, the number of data points for any given subject will usually not be sufficient to learn the underlying model. Even if it were, achieving a specific model for each subject may not serve the purpose of understanding the common characteristics of the problem. Therefore the main question is how a neural network can take account for the time or subject correlation in the data.

A typical neural network is made up of 3 or more layers of neurons (input, hidden and

output layers), each connected to the neuron in the next layer by weighted connections. See figure 6.3 for an illustration of the neural network architecture. The network architecture can also contain bias nodes in the input and hidden layers. The connecting weights are adjusted to best fit the series of input variables to the outcome. When the weights are updated, the network is learning.

The inclusion of a greater number of hidden layers increases the modelling flexibility. One hidden layer is sufficient to model any piecewise continuous function. When selecting the number of neurons in the hidden layer, we note that if too few neurons are used the network will not be able to model the data and the resulting fit will be poor. However if too many are used, the network will overfit the data and begin to model random noise in the data. Therefore we need to do a cross validation exercise to determine the appropriate number of neurons in the hidden layer.

The output of the j^{th} neuron in the hidden layer is given by:

$$H_j = \varphi(\sum_i w_{ij} y_i + \theta_{1j})$$

where $\varphi(x) = \frac{1}{1+e^{-x}}$, w_{ij} are the weights between the input layer and hidden layer, y_i is the value of the i^{th} input and θ_{1j} is the bias value. Similarily, the output of the j^{th} neuron in the output layer is given by

$$O_j = \varphi(\sum_k w_{kj} H_k + \theta_{2j})$$

where w_{kj} are the weights between the hidden layer and output layer, H_k is the output from the k^{th} hidden neuron and θ_{2j} is the bias value. Illustrations of this can be seen in the network architecture plots to follow.

The type of training that was used in this analysis was Resilient Propagation (RPROP+) (Riedmiller and Braun 1993), with a sigmoid activation function. Backpropagation training optimises the weights of the network by minimising the loss function through an iterative gradient descent process. Backpropagation requires a learning rate, usually a very small constant, in order to force the weights to be updated smoothly and slowly. The effectiveness of the backpropagation is highly sensitive to the learning rate. Resilient backpropagation was developed to improve

upon the backpropagation algorithm by only using the sign of the gradient to update weights, and not the magnitude. RPROP does not require parameter tuning of the learning rate since it does not use a single learning rate.

Scaling of the data is essential because otherwise a single variable may have a large impact on the outcome only because of its scale. Therefore using unscaled data may lead to meaningless results. The numerical covariates were normalized in intervals of 0 to 1, using the min-max normalization method:

$$x_{norm} = \frac{x - x_{min}}{x_{max} - x_{min}}$$

This transforms the data into a common range, removing the scaling effect from all variables. This method also retains the original distribution of the variables, unlike the z-score normalisation method. The categorical covariates were binary coded, with 1 representing presence of the variable and 0 otherwise.

Further to fitting an adequate model to the height measurements, interpreting variable significance is also of interest as we wish to detect factors that contribute to malnutrition. Therefore interpretability of the model is important. The relative importance of covariates in a neural network can be assessed by deconstructing the model weights, as seen in (Garson 1991). This method is explored and compared to the variable importance identified in the conditional mixed-effects model.

6.2 Neural network models

The following approaches were taken to fit a neural network to the height data:

- Age-specific intercepts. Here the age variable has been scaled along with the other variables and used as an input into the neural network. i.e. age in months is associated with a numerical value between 0 and 1, therefore identifying age-specific intercepts for time. This structure takes no special account for subject-specific effects. This corresponds to approach 1 discussed on page 12, where Ganesan et al. (2014) used neural networks to model growth data of sheep.
- Age-specific neural networks. Here a separate neural network is trained for height at each age. This allows the effects of each covariate on the outcome variable to

be different for each age. This may result in a more meaningful analysis of the neural network since factors affecting height at certain ages may differ to those affecting height at other ages.

- Subject-dummy model. Here a subject dummy variable is fed into the network along with the other variables. This allows for a subject-specific effect in the model. This corresponds to approach 3 discussed on page 12, where Maity and Pal (2013) uses a subject-specific treatment in neural networks.

6.2.1 Age-specific intercepts

In order to effectively select the appropriate number of neurons in the hidden layer, 10 fold cross validation was used to obtain the configuration that minimises the test Mean Squared Error (MSE). This was done by, starting with 1 hidden neuron, splitting the data into training and testing data 10 times and calculating the average MSE. The number of hidden neurons is then increased and the process repeated. The optimal number of hidden neurons is the number that resulted in the smallest average MSE over 10 folds.

Figure 6.1 shows the testing dataset MSE (averaged across 10 folds) for various numbers of hidden neurons. We can see that 6 neurons resulted in the smallest MSE. Therefore 6 neurons were used in the hidden layer.

The stopping criteria of the network training is either reaching 1 000 000 steps or having a mean square error of less than 0.01. This is to avoid excessive steps or insufficient cycles, resulting in non-convergence.

The first time weights used in training the neural network are usually randomly assigned. However this can result in convergence at a local minimum. It was noted that when different seeds were set before fitting the network, the results were different. Therefore it is better to provide initial weights that work well for the problem. This can be done based on prior knowledge, e.g. we know that age will be the most important predictor of height. The network was fit using various initial weights and the variable importance was assessed each time. The chosen seed is one that resulted in the most common order of variable importance across various seeds. Table 6.1 shows the variable importance obtained from a few different choices of inital seeds. We can

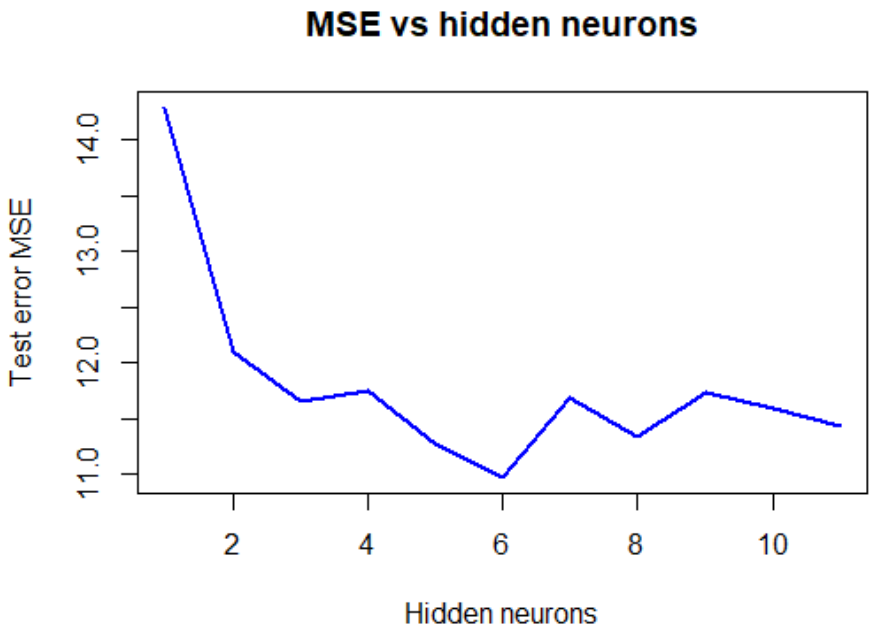


Figure 6.1: MSE vs hidden neurons

see that in most instances, the order of importance is consistent. However there can be abnormal behaviour due to poorly initialised weights.

Figure 6.2 shows the correlation between the variable importance of various seeds. We can see that they are all highly correlated. The choice of seed that has the most correlation in variable importance with the other seeds is seed 5 and therefore this was chosen to use in the modelling process.

Table 6.1: Variable Importance for different choice of seeds (continued below)

	seed.3	seed.4	seed.5	seed.8	seed.15	seed.22
Age	0.3323	0.2486	0.346	0.4436	0.4194	0.2209
Early	0.041	0.0867	0.1087	0.0682	0.0487	0.0862
Sex	0.1468	0.0448	0.0397	0.009	0.0498	0.043
Clinic	0.195	0.0889	0.1191	0.0289	0.1496	0.0974
SES 1	0.0179	0.0673	0.0656	0.0694	0.0263	0.0532
Mom Height	0.1007	0.0906	0.1112	0.121	0.1165	0.0995

	seed.3	seed.4	seed.5	seed.8	seed.15	seed.22
Tobacco	0.0705	0.102	0.0639	0.0282	0.0781	0.1142
Alcohol	0.0161	0.0823	0.0507	0.0701	0.0305	0.0838
Mom age	0.0456	0.0443	0.0434	0.0813	0.0447	0.0445
Breastfeed	0.0189	0.0865	0.0151	0.0587	0.0173	0.0611
HIV	0.0153	0.0579	0.0366	0.0216	0.0193	0.0962

	seed.31	seed.36	seed.38	seed.44
Age	0.2414	0.2982	0.2945	0.2852
Early	0.1122	0.0433	0.0835	0.0966
Sex	0.0994	0.0428	0.0428	0.0461
Clinic	0.0895	0.1428	0.1184	0.0775
SES 1	0.0482	0.036	0.0799	0.0815
Mom Height	0.1332	0.1467	0.0907	0.1211
Tobacco	0.1044	0.0734	0.0616	0.0736
Alcohol	0.0382	0.0854	0.036	0.0485
Mom age	0.0615	0.0486	0.1405	0.0815
Breastfeed	0.0461	0.0506	0.0308	0.0496
HIV	0.0258	0.0324	0.0211	0.0389

To visualize the Neural Networks, figure 6.3 shows the neural network architecture. The following provides a key to interpreting the plot:

- The first layer consists of the model inputs, indicated by I1, I2, etc
- The hidden layer nodes are indicated by H1, H2, etc
- The output layer, O1, corresponds to height
- The bias is indicated by B1, B2, etc
- Larger weights have thicker connecting lines
- The colour of the line indicates the sign of the weight: (+ black, - grey)

Similarly to how the parameter coefficients in a mixed-effects model can be used to describe the relationship between variables, so can the weights in a neural network model. The weights indicate the relative importance of the input variables in their relation to the outcome. The plot can be analysed to assess these relationships. For ex-

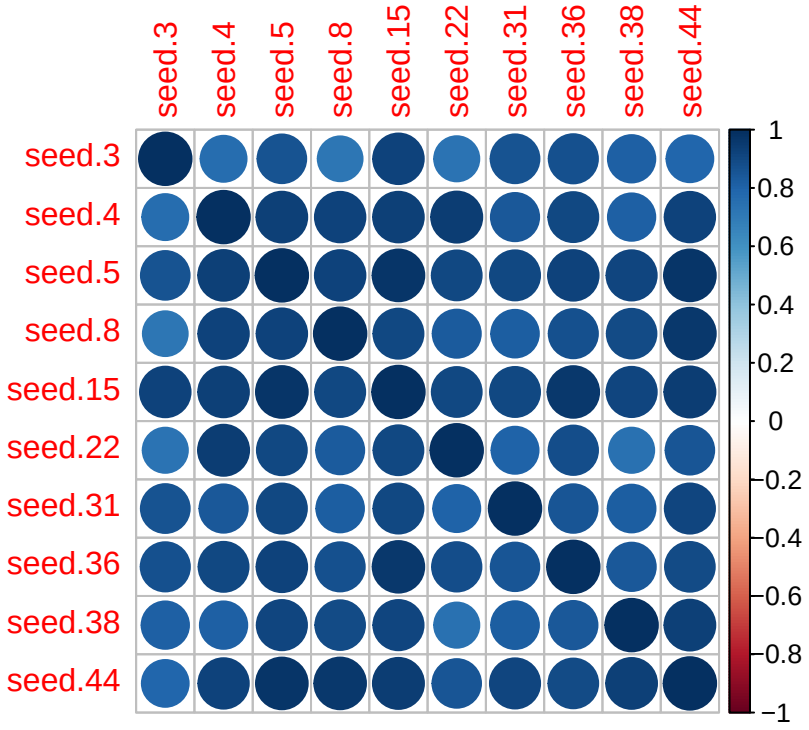


Figure 6.2: Correlation between variable importance of various seeds

ample, input variables that have a strong positive association with height are expected to have many thick black connections between the layers.

However the hidden layer can make this interpretation challenging, particularly where the sign of the weight changes after the hidden layer. The number of weights in a neural network can be far greater than parameter coefficients in a mixed effect model. This characteristic is positive in that it allows neural networks to be very flexible in modelling non-linear functions with multiple interactions. However interpretation complexity increases with an increasing number of hidden layers and nodes within each layer. For example, we know that age is an important predictor of height, however the connecting lines do not highlight this. Instead, we can investigate the relative importance of the input variables.

Relative importance of the input variables can be assessed by using the garson function in the NeuralNetTools package (MW. Beck). This function uses the method proposed by Garson (1991), which calculates the relative importance of input variables in a neural network by deconstructing the model weights. All weights connecting a specific input that pass through the hidden layer to the outcome variable are identified. This is

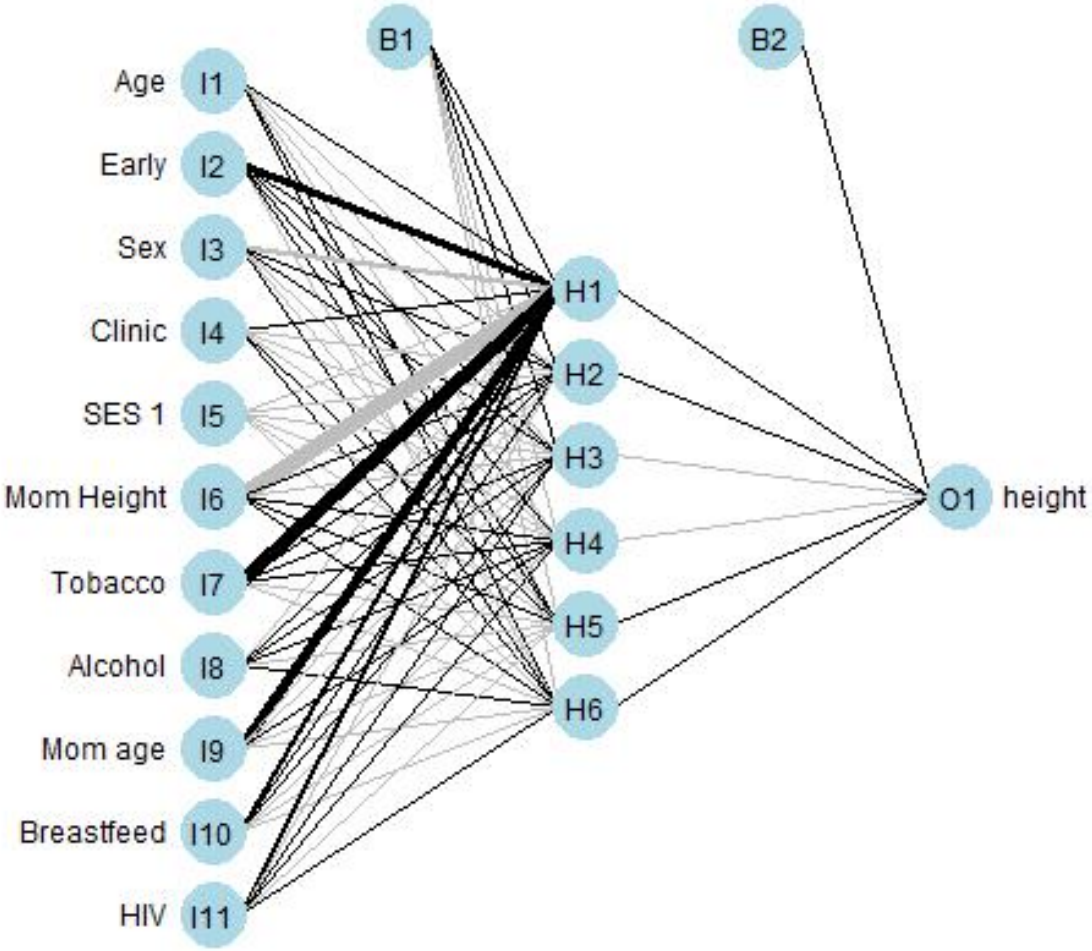


Figure 6.3: Neural Network Architecture

done for all input variables until a list of all weights associated with each input variable is obtained. The weights are then tallied for each variable and scaled relative to the other variables. This single value then describes the relationship with the outcome variable.

Figure 6.4 shows the variable importance in the neural network architecture.

Comparing this with the variable significance investigated in the mixed effect model, we see a similar result of the most influential covariates. We would expect age to be the major contributing covariate, as we know that height increases with age. Clinic, maternal height and early gestation are the next most important predictors. This was also evident in the mixed-effect model where clinic affected the linear growth rate, maternal height affected the birth height and growth velocity, and early gestation affected the birth height and growth velocity. The remaining variables are of less importance in the network but still contribute to the determination of the height value. (Note: SES1 refers to socio-economic class: Low).

The downside to this approach of the neural network is that the dataset used to train the network is in long format. i.e. each child has multiple rows for the height measurements at each age and the baseline covariates are repeated. Since the data is unbalanced, this may result in undue emphasis on certain covariates if some children exhibiting those risk factors had a greater number of measurements. We also cannot see how the covariate affects growth at different stages of the growth trajectory. An alternate method is to use age-specific neural networks, as discussed below.

6.2.2 Age-specific neural networks

An alternative approach to the age-specific intercepts is to train a separate neural network for each age. This will allow the effects of the covariates on the outcome to be different for each age. Using this approach may result in a more meaningful analysis of the neural network variable significance, since factors affecting height at certain ages may differ to those affecting height at other ages. Neural networks were fit for height at birth and 24 months to illustrate this approach. Cross validation was again used to choose the number of hidden of neurons, which was 5 for these networks.

Figures 6.5 and 6.6 shows the neural network architecture and variable importance for the birth height model. We can see that the variable importance differs from the

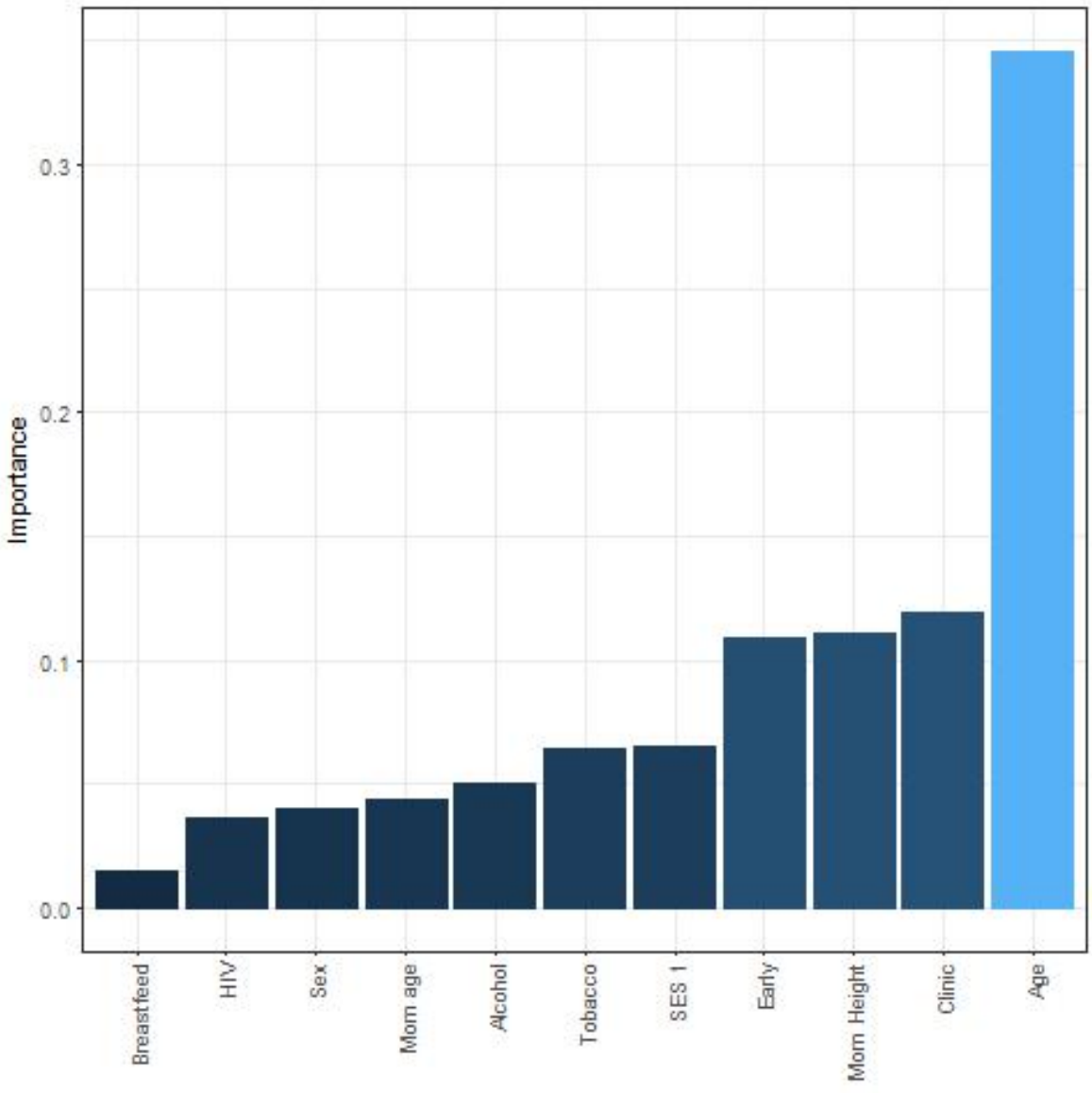


Figure 6.4: Variable Importance

overal height neural network. HIV, early gestation and maternal height are the most important predictors in the model at birth. This is in line with the Berkey-Reeds model where the intercept (birth height) parameter was

$$\beta_{0i} = \beta_0 + \beta_5(X_{early}) + \beta_6(X_{sex}) + \beta_7(X_{tobacco}) + \beta_8(X_{maternalHeight}) + \beta_9(X_{HIV}) + b_{0i}$$

Figures 6.7 and 6.8 shows the neural network architecture and variable importance for the height model at 24 months. The most important variables are socio-economic class: Low, early gestation, tobacco during pregnancy and sex. Socio-economic class has moved from being the 5th most important predictor in the birth height model, to the most important predictor in the 24 month model. This illustrates how lifestyle factors, such as living conditions, impact height more significantly post birth compared to at birth.

6.2.3 Subject dummy variables

In this approach, we investigate how a subject-specific effect can be incorporated into a neural network. As with mixed-effect models, we wish to distinguish between common effects and subject-specific effects to allow us to understand how certain risk factors affect all child growth. Each subject was given a dummy variable. These variables were then fed into the neural network along with age and the other variables. The number of neurons used in the hidden layer was 6 so that the effect of each covariate can be compared with the age-specific neural network where 6 hidden neurons were also used. Since this results in 853 input variables, the network architecture is difficult to plot. In the variables importance plot, the subject dummy variables were omitted and only the other input variables were plotted. Figure 6.9 shows the importance assigned to each of the input variables. We see that age is the most important variable, followed by maternal height, socio-economic class: Low, tobacco during pregnancy and sex.

A specific child's height will then be the combination of their individual dummy input, along with the other variables. If we select a specific child, we can plot the variable importance taking into account their dummy input. Figure 6.10 shows the variable importance plot with one child's dummy input included. We can see that the subject

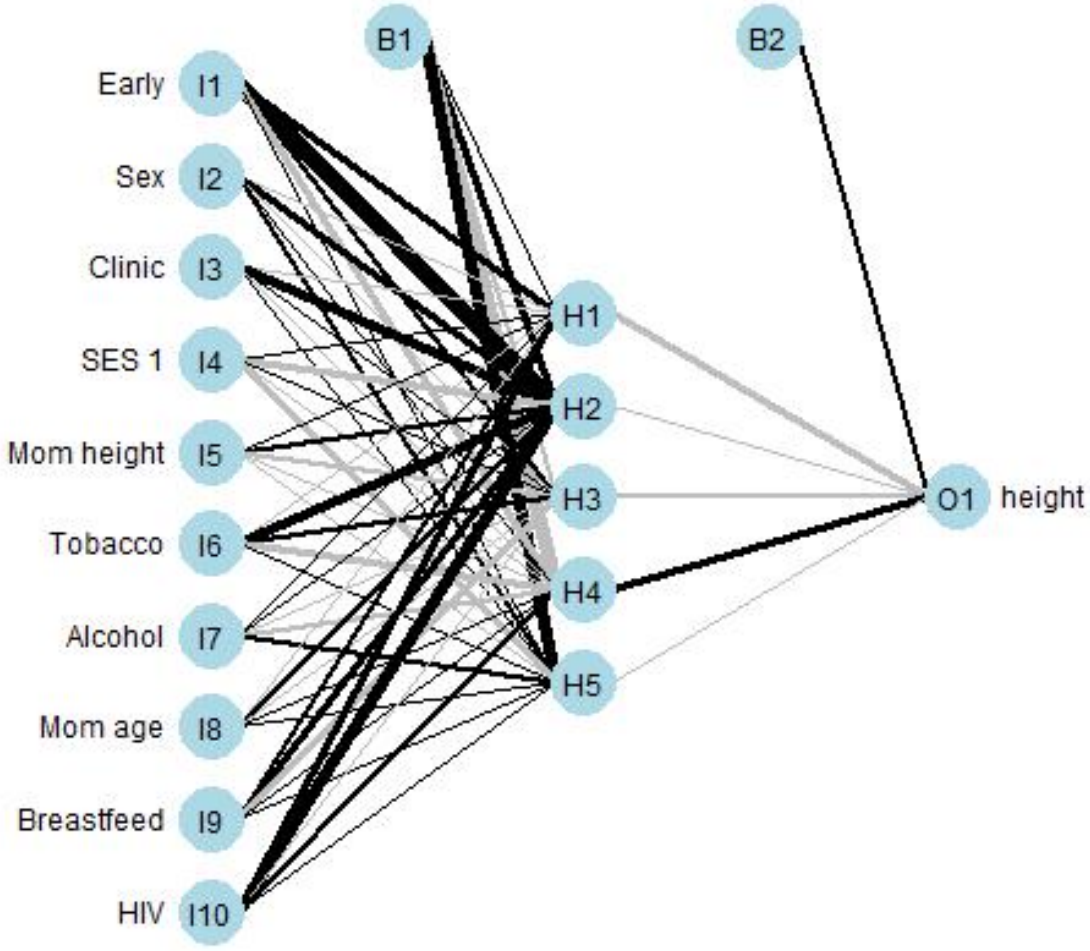


Figure 6.5: Birth Height Neural Network Architecture

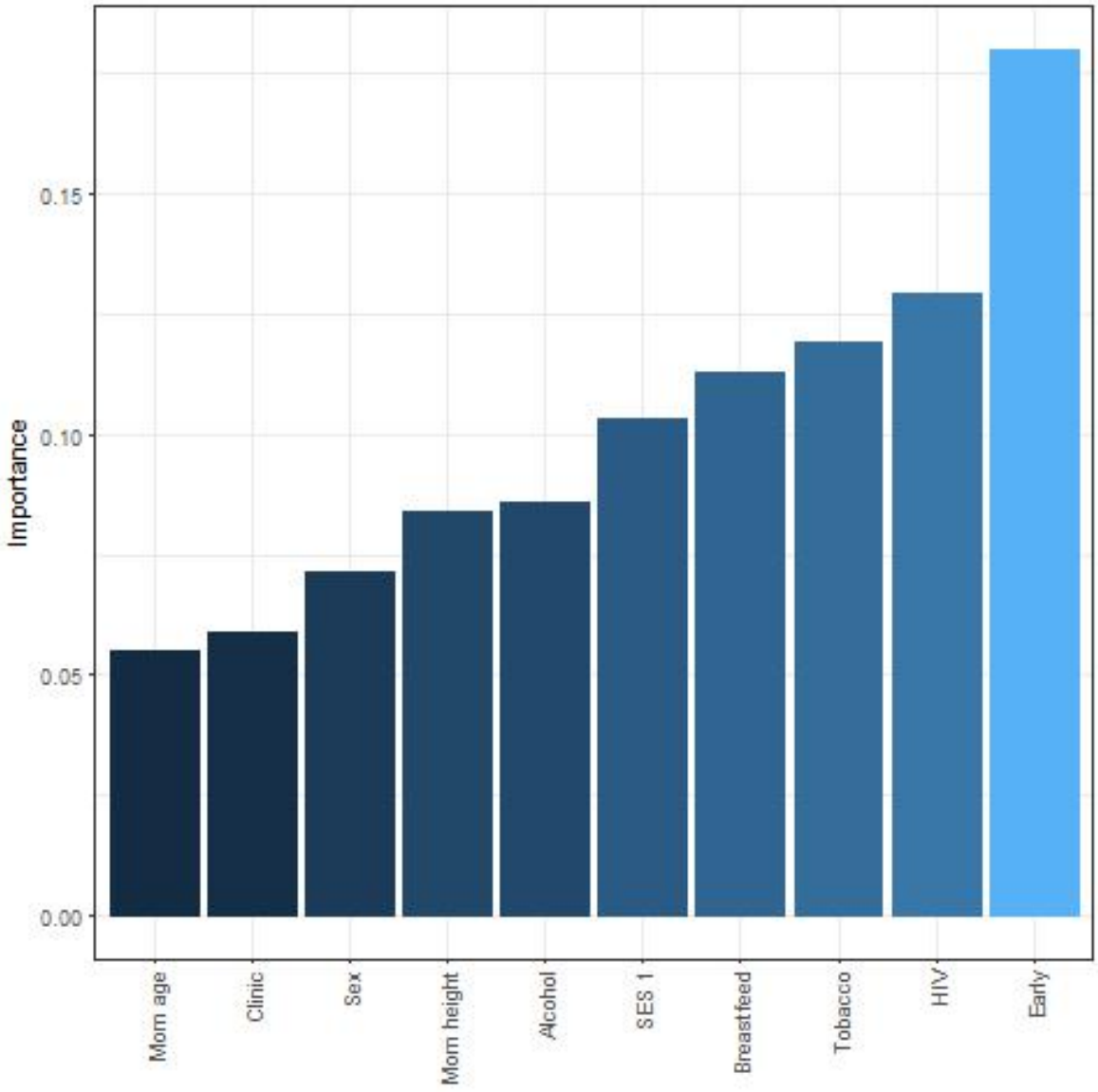


Figure 6.6: Variable importance for birth height

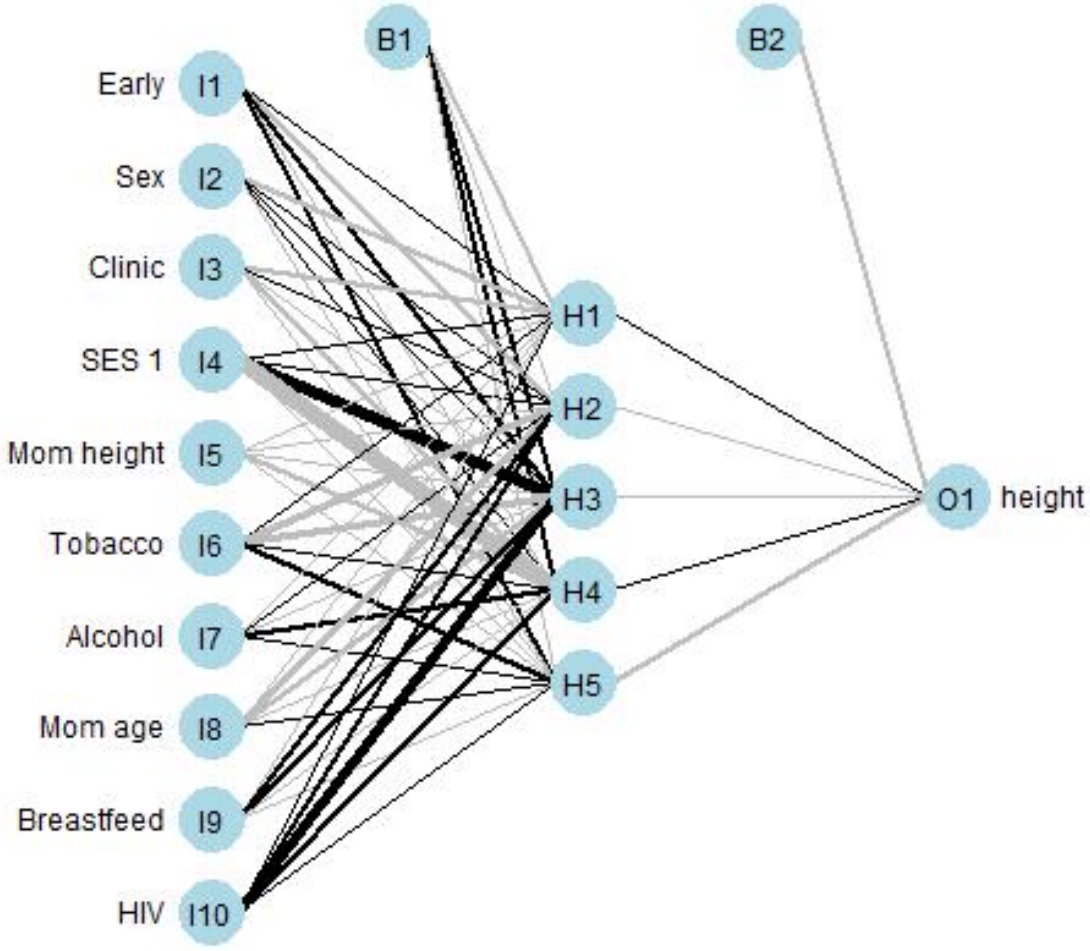


Figure 6.7: 24 month Height Neural Network Architecture

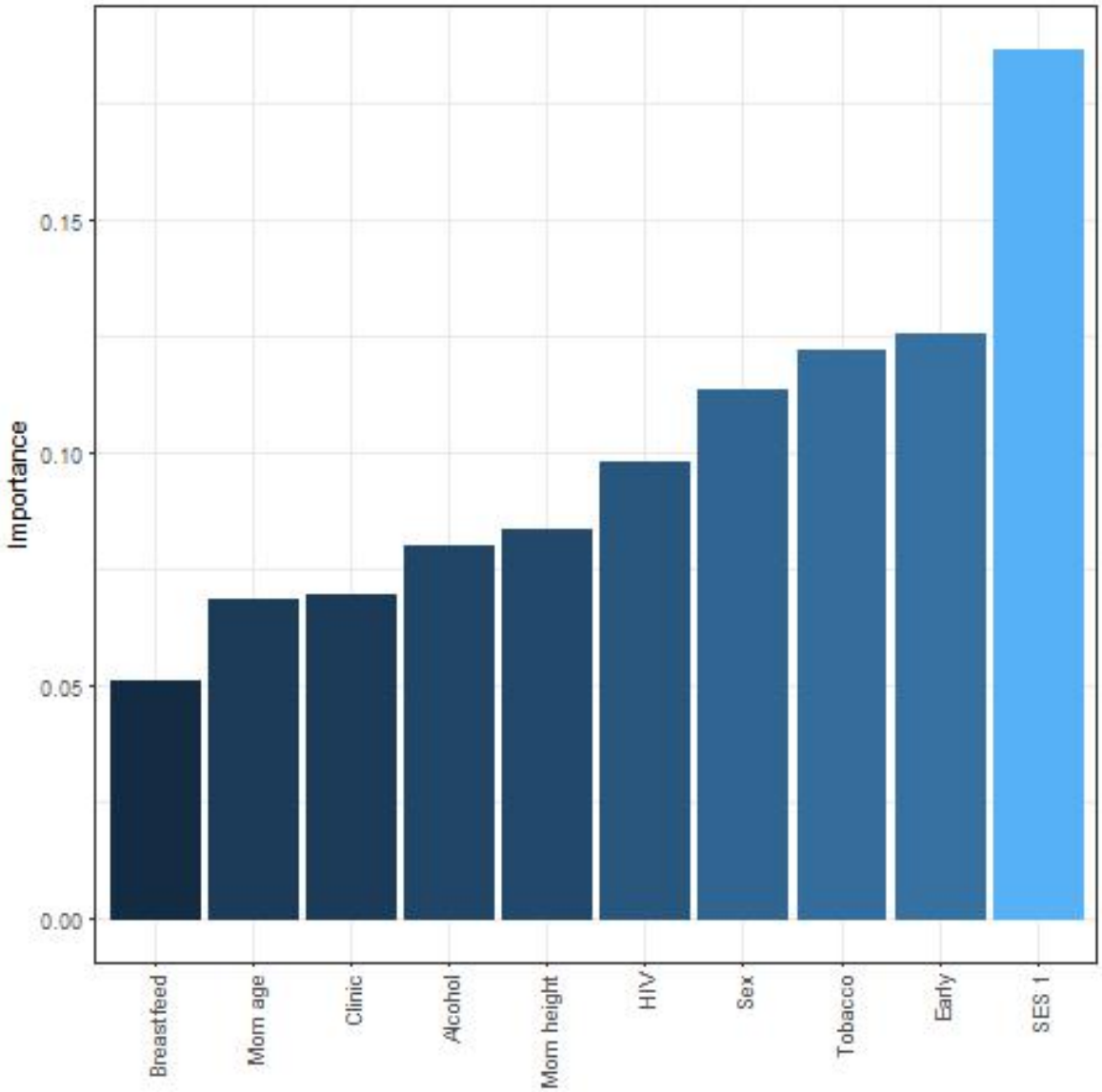


Figure 6.8: Variable importance for 24 month height

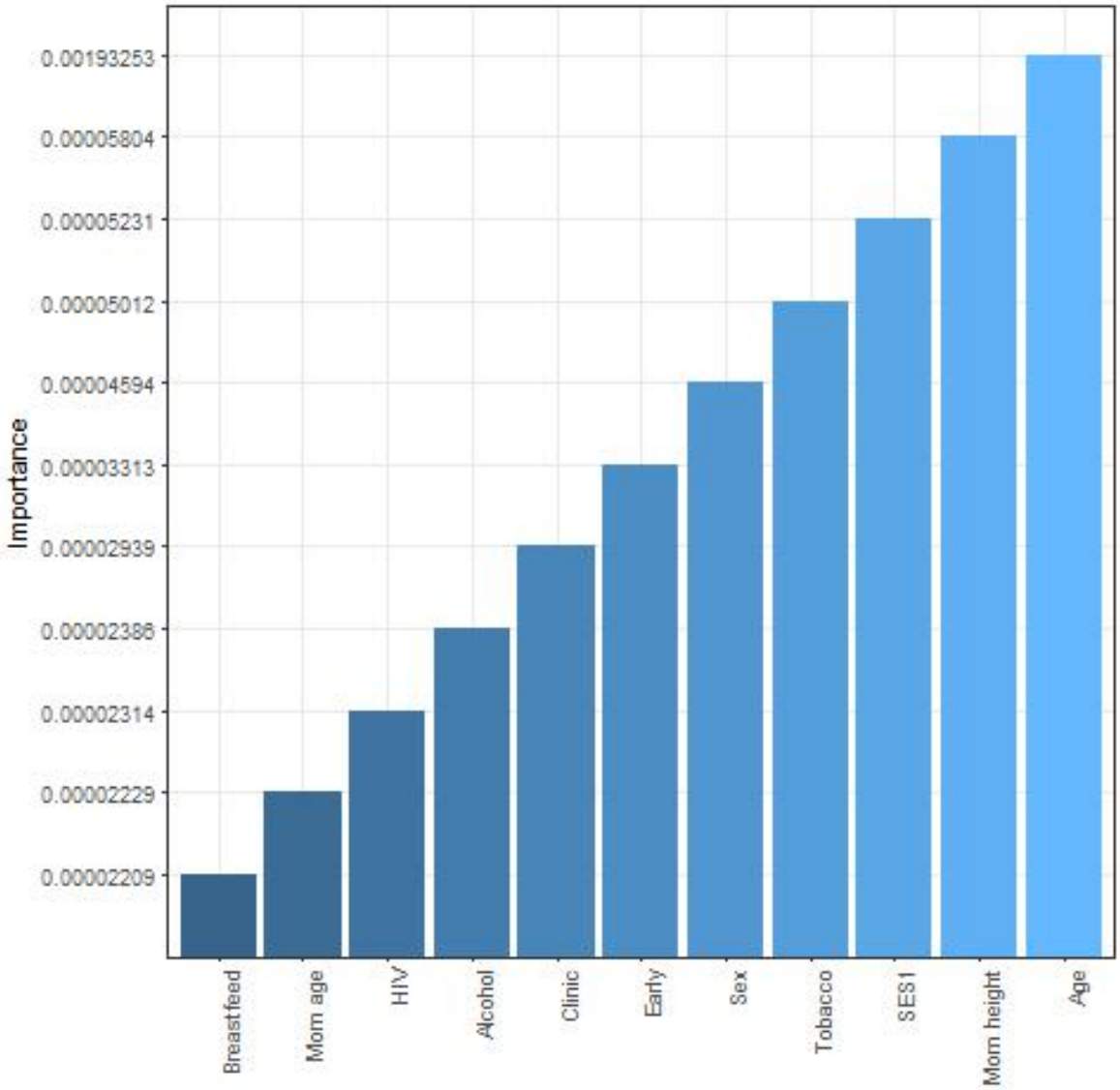


Figure 6.9: Variable importance for subject dummy neural network

dummy is the most important input.

6.2.4 Evaluation of neural networks

The previous section illustrated that neural networks can be useful in providing insight into the variables affecting child growth. The first approach, age-specific intercepts, showed us which variables were important in determining child growth across all ages. This model allows us to use one neural network to predict height at any age. Due to the nonlinear modelling capability of neural networks, complex relationships between variables can be captured in one model. The drawback of this approach is that we cannot distinguish variable importance at different ages.

The next approach, age-specific neural networks, allowed us to investigate the variable importance at each age. This very nicely illustrates the effect of different covariates at different stages of the growth profile. In the mixed-effect model, we were able to estimate this using interactions between age and the covariates. Comparing the birth-height neural network to the intercept covariate terms in the Berkey-Reeds model, we saw that the variables of high relative importance in the neural network were the same as the significant intercept variables in the Berkey-Reeds model.

The structural nature of the Berkey-Reeds model allows us to investigate how certain covariates affect aspects of child growth such as linear growth, deceleration in growth velocity and growth inflection points. The neural network, on the other hand, does not allow this biological interpretation of covariates. However we can determine which covariates are the most influencial predictors of height at each age. The two interpretations can be complimentary. For example, in the 24 month neural network, we saw that a low socio-economic class and early gestation were the most important predictors of height. This corresponds to both of these variables having significant interactions with the linear growth term of the Berkey-Reeds model.

The last neural network approach was the subject dummy variable model. This model allowed a subject-specific effect to be incorporated into the neural network. This enables us to understand the common characteristics that apply to child growth by seperating the common effects and subject-specific effects. The results showed that the individual subject variables were the most important predictors relative to the other inputs. Thereafter age, maternal height, a low socio-economic class and tobacco

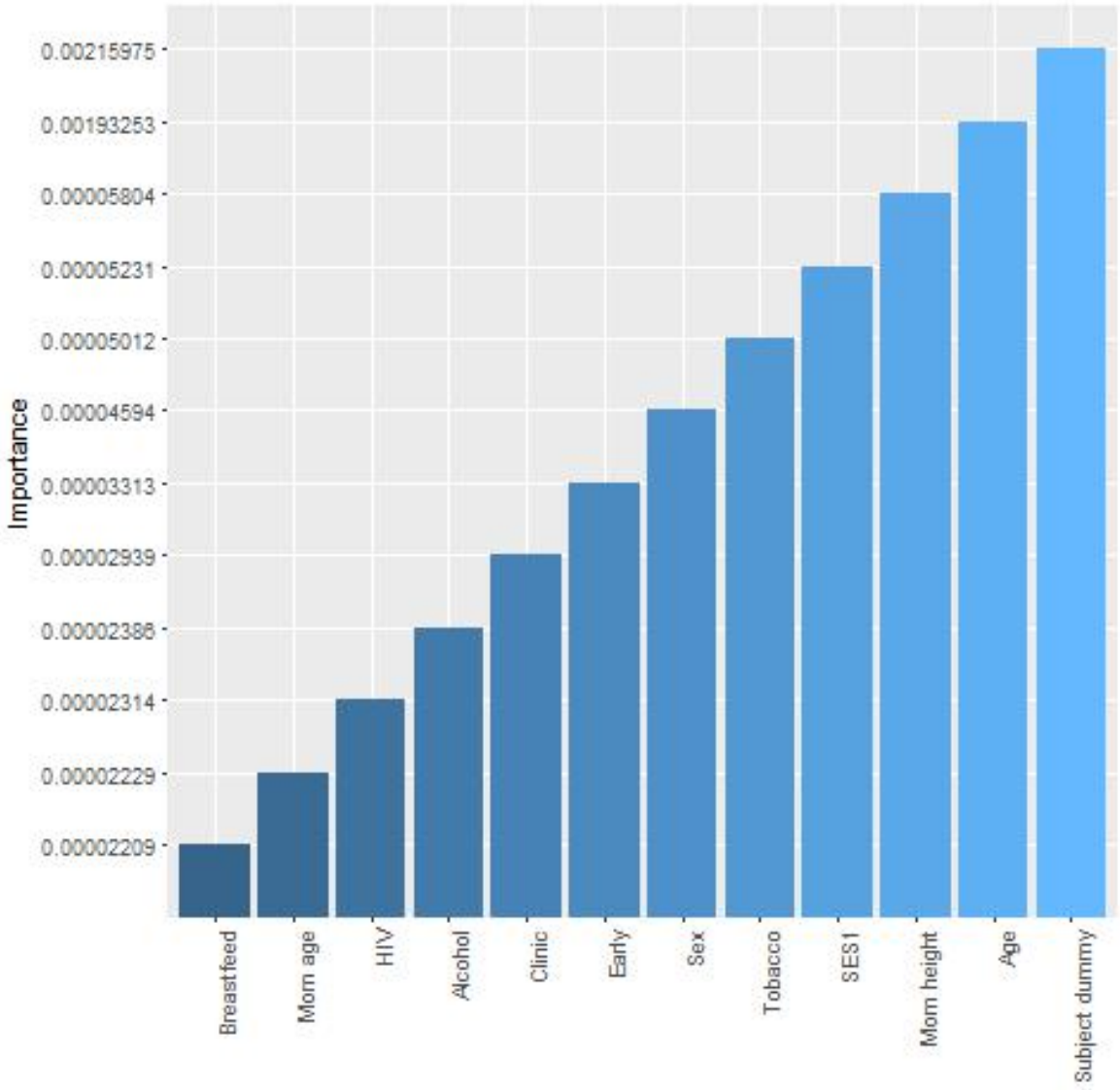


Figure 6.10: Variable importance for subject 1 dummy neural network

during pregnancy were the most important variables that applied to all children.

In the next section we investigate the predictive power of the mixed effect model compared to the neural network model. We then compare the two methods based on interpretability, predictive power and model construction.

7 Predictions

In this section, the potential of using an Artificial Neural Network as an alternative modelling strategy to the more coventional mixed effect model was assessed for the purpose of predicting child growth using covariates available at birth.

7.1 Training and testing datasets

The data was divided into a training and test set. The training set is used to find the relationship between the height measurements and covariates, and hence determine estimates of the parameter values. The test set is used to assess the performance of the models and compare the predictive power of the mixed effect model with the NN model. Random sampling was performed to split the data. 70% of the children are in the training dataset and 30% are in the testing dataset.

7.2 Mixed modelling prediction

The conditional 2nd order Berkey-Reeds model was fit on the training dataset and predictions done on the testing dataset. Figure 7.1 shows the actual versus predicted values. We can see that the points lie close to the diagonal line, indicating that the model is performing well. The labeled points in the scatter plot indicate where the predicted height greatly under- or over-predicted the actual height. This identifies cases where the model predicted very poorly. We compare these predicted measurements with the neural network predictions to determine if the neural network is better at modelling unusual cases.

The Root Mean Square Error (RMSE) was calculated to compare the prediction accuracy of the mixed effect model to the NN model.

$$RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(actual_i - predicted_i)^2}$$

The RMSE for the conditional Reed2 model is 3.476628

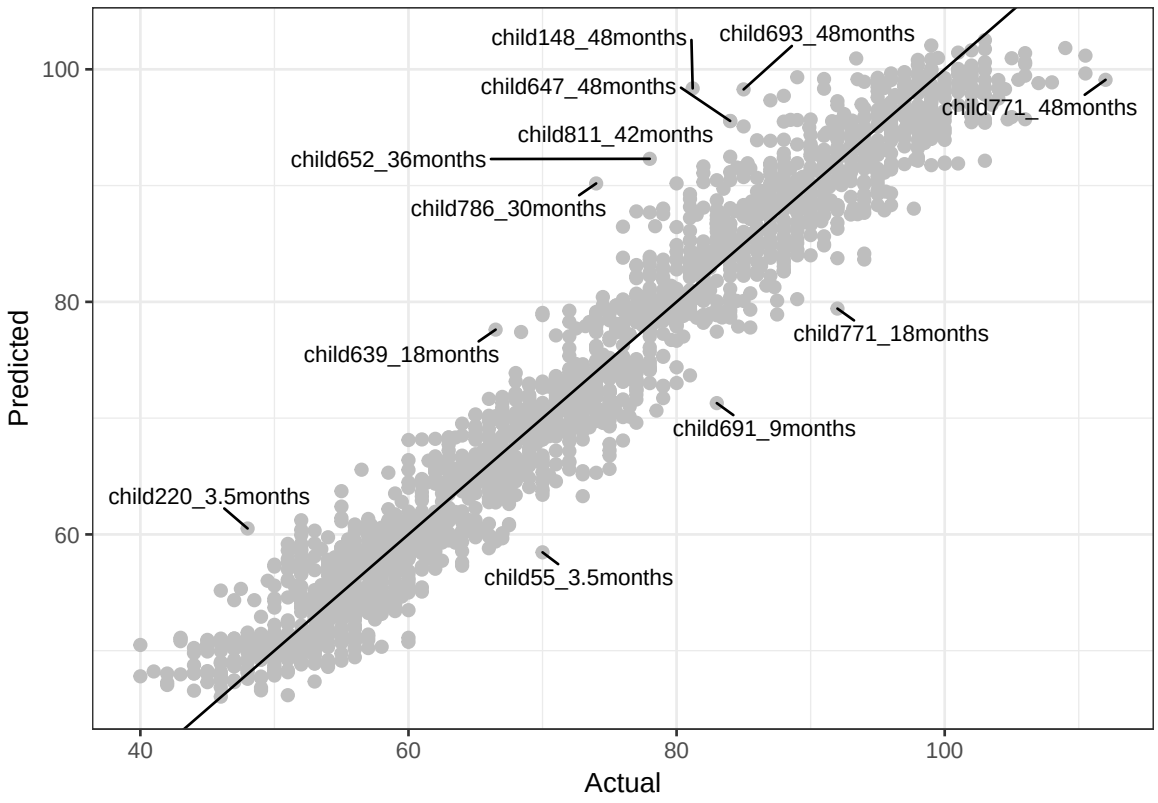


Figure 7.1: Berkey-Reeds model: Actual versus predicted heights

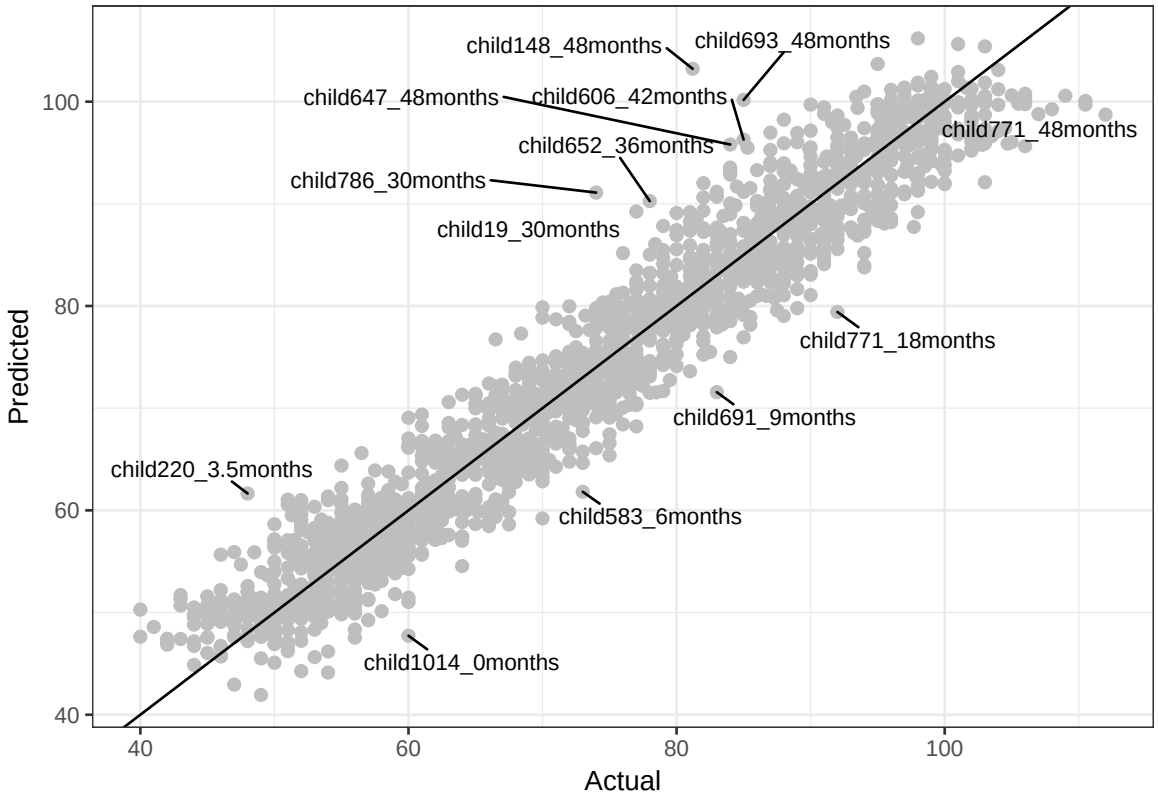


Figure 7.2: Neural Network: Actual versus predicted heights

7.3 Neural network prediction

The 3 layer age-specific intercept neural network with 6 neurons in the hidden layer was used to train and test the model. Figure 7.2 shows the actual versus predicted values. The points lie close to the diagonal line, indicating good fit. The scatter of points is similar to that of the mixed effect model. However we see that there are a greater number of cases where the model severely under- or over-predicted the height (12cm difference in actual and predicted values). The same measurements have been poorly predicted as in the mixed-effect model, indicating that these cases may be difficult to capture adequately in any model.

The RMSE using this model is 3.751524. This model performs quite well considering that no specific allowance has been made for the longitudinal nature of the data. Comparing the RMSE of the mixed effect model and neural network model shows that the mixed effect model predicted heights better.

When comparing which method (mixed model or neural network) is best to model height curves in order to assess malnutrition, there are various considerations to take into account. These include:

- Interpretability of the model
- Predictive power
- Ease of model construction (i.e. parameter choice)

In terms of model interpretation, the structural nature of the Berkey-Reeds model allows for a more in-depth understanding of how certain risk factors affect various aspects of child growth. The neural network has no associated mathematical formula and therefore interpretation can be difficult. However, with the use of the Garson method, variable importance can be determined from the neural network, allowing us to derive insights into the risk factors affecting child growth. This is particularly useful when we use a seperate network per age, so that risk factors at each age can be analysed.

As we saw when comparing the RMSE of the two models, the Berkey-Reeds model was able to make better predictions of child height than the neural network. Since the Berkey-Reeds model is a well-defined and studied curve for modelling child curve, we would expect this model to perform well. However the predictive performance of the neural network showed promising results and did not greatly under-perform relative to the mixed model.

The advantage of the neural network is that the parameter selection process may be much easier due to a defined formula for the relationship between height and age not needing to be defined. During the mixed-model process, we were required to test every interaction of each variable with each age term in the formula. In building a neural network, interactions between variables do not need to be defined and the model handles nonlinear relationships well with no need to specify the specific form. However, in the neural network, the choice of hidden layers and neurons may become lengthly. Although a grid search approach can be applied to this problem to determine the optimal network architecture by testing all possible combinations.

8 Classification

8.1 Stunting indicator

In this section we investigate classification models for childhood stunting. Stunting is defined as a shortfall in height relative to a child's age. At each age, the height measurement is coverted into a height-for-age z-score. Z-scores below -2 are classified as stunted. The critical window, wherein 70% of stunting occurs, is from birth to 2 years of age (0-23 months) (World Health Organization 2010). This growth deficiency continues after 2 years of age due to continued exposure to sub-optimal care.

The analysis was therefore done in two phases: 0-23 months and 24-48 months. Children who had a height-for-age z-score of below -2 on two visits within the period were classified as stunted. The proportion of children who were stunted during the first 2 years of life was 0.27 and between 24 and 48 months was 0.14, indicating that stunting is more prevelant during the first 2 years of life. Some children had missing covariates and were therefore excluded from the analysis.

Conventional statistical methods were compared with tree-based methods for evaluating stunting. Logistic regression was first performed to determine the pre- and post-natal risk factors that were associated with stunting. Decision trees were also investigated as a graphical tool for understanding covariates. Prediction models were then constructed to model the probability of stunting based on risk factors available at birth. In particular, we compare the predicitive performance of a logistic regression model with a random forest model, which is an ensamble of decision trees.

8.1.1 Investigation of covariates

Table 8.1 shows number and proportion stunted (0-23 months) for each level of the categorical covariates. The following observations were made:

- A greater proportion of males are stunted than females.
- A greater proportion of children born at TC Newman are stunted than those born at Mbekweni.
- Exposure to tobacco during pregnancy has a greater proportion of stunted children than no exposure.

- Exposure to alcohol during pregnancy has a greater proportion of stunted children than no exposure.
- The proportion of children stunted decreases with increasing socio-economic status.
- A greater proportion of children with an early gestational age were stunted, compared to those with a normal gestation age.

Table 8.1: Proportion stunted (0-23 months)

Variable	Level	Number	Percent
Sex	Female	77	19.11
	Male	149	34.02
Clinic	TC Newman	128	33.42
	Mbekweni	98	21.4
Maternal HIV status	Yes	50	27.17
	No	176	26.79
Tobacco during pregnancy	Yes	89	38.2
	No	137	22.53
Alcohol during pregnancy	Yes	60	41.96
	No	166	23.78
Socio economic quartile	Low	79	36.74
	Low to moderate	63	29.17
	Moderate to high	47	22.6
	High	37	18.32
Maternal marital status	Single (never married)	144	28.35
	Not married, but in a marriage like relationship	45	30
	Married	35	19.55
	Divorced	1	50
	Widowed	1	50
Initiated breastfeeding	Yes	204	26.32
	No	22	33.33
Early gestation (<38 weeks)	Yes	97	48.02
	No	129	20.19

Table 8.2 shows the summary statistics by stunted status (0-23 months) for the numeric covariates. The following observations were made:

- Maternal age does not differ significantly for stunted versus non-stunted children.
- Mothers of stunted children are significantly shorter than those of non-stunted children.
- The gestational age of stunted children is significantly lower than non-stunted children.

Table 8.2: Numeric covariates proportion stunted (0-23 months)

Variable	Stunted	N	Mean	SD	Median	t-test p.value
Maternal age	Yes	226	26.41	5.85	25.28	0.246
	No	615	26.94	5.75	26.31	
Maternal height	Yes	226	157.88	6.68	158	<0.001
	No	615	159.9	6.78	160	
Gestational age	Yes	226	37.39	3.13	38	<0.001
	No	615	38.93	2.09	39	

Table 8.3 shows the number and proportion stunted (24-48) months for each level of the categorical covariates. We observe a similar trend in stunting proportions as stunting within 0-23 months regarding sex, clinic, tobacco exposure, alcohol exposure, socio-economic status and early gestational age.

Table 8.3: Categorical covariates proportion stunted (24-48 months)

Variable	Level	Number	Percent
Sex	Female	49	12.16
	Male	69	15.75
Clinic	TC Newman	71	18.54
	Mbekweni	47	10.26
Maternal HIV status	Yes	22	11.96
	No	96	14.61
Tobacco during pregnancy	Yes	50	21.46
	No	68	11.18
Alcohol during pregnancy	Yes	29	20.28
	No	89	12.75
Socio economic quartile	Low	52	24.19
	Low to moderate	30	13.89
	Moderate to high	20	9.62
	High	16	7.92
Maternal marital status	Single (never married)	75	14.76
	Not married, but in a marriage like relationship	25	16.67
	Married	17	9.5
	Divorced	0	0
	Widowed	1	50
Initiated breastfeeding	Yes	112	14.45
	No	6	9.09
Early gestation (<38 weeks)	Yes	45	22.28
	No	73	11.42

Table 8.4 shows the summary statistics by stunted status (24-48 months) for the numeric covariates. Again we see that maternal height and gestational age differ by stunting status.

Table 8.4: Numeric covariates proportion stunted (24-48 months)

Variable	Stunted	N	Mean	SD	Median	t-test p.value
Maternal age	Yes	118	26.52	6.19	25.24	0.599
	No	723	26.84	5.71	26.31	
Maternal height	Yes	118	155.94	6.55	155.75	<0.001
	No	723	159.91	6.69	160	
Gestational age	Yes	118	37.68	3.04	38	0.001
	No	723	38.66	2.38	39	

8.2 Classification models

Two methods were explored to model stunting as a function of the risk factors: Logistic regression and random forests. The purpose of the modelling process is two-fold:

- Identify which risk factors have an influence on the probability of stunting.
- Compare which model is better suited to predict the probability of stunting based on the risk factors available at birth.

8.2.1 Logistic regression

According to Hilbe (2011), logistic regression is the most commonly used method to model binary response data. It has often been used to evaluate childhood nutrition and stunting status in various studies. In Thibault et al. (2013), logistic regression was used to evaluate obesity in children aged 5-11 years in France. In I. O. Senbanjo et al. (2013), logistic regression was used to evaluate underweight and wasting in rural and urban areas from Nigeria. Akombi et al. (2017) uses logistic regression to model childhood stunting, underweight and overweight in Indonesia.

Here logistic regression was used to determine the risk factors associated with stunting.

The model equates the log-odds of the probability of stunting with a linear combination of the predictor variables, i.e.

$$ln\left(\frac{p(y_i=1)}{1-p(y_i=1)}\right)=\beta_0+\sum_j^p x_{ij}\beta_j$$

for $i=1,...,n$

where

the x_{ij}'s represent the covariate values, p is the number of covariates

and

$$y=\begin{cases}1 & \text{for} \quad stunted\\ 0 & \text{for} \quad non-stunted\end{cases}$$

Lasso regression was used to select the most appropriate variables. Lasso was used beacause it sets the coefficients of some variables to zero, thereby avoiding overfitting. It does this by regularisation, which adds a penalty equal to the absolute value of the magnitude of the coefficients Hastie, Tibsharani, and Friedman (2009). The algorithm minimises:

$$\sum_i^n(y_i-\sum_j^p x_{ij}\beta_j)^2+\lambda\sum_i^p|\beta_j|$$

The value of the penalty parameter, λ, used for the lasso was determined using cross validation based on mean-squared error.

Figure 8.1 shows the results of the cross validation. The value of $\log(\lambda)$ that resulted in the lowest mean-square error was -6.6725519, corresponding to a lambda value of 0.0012652. In order to balance accuracy and simplicity, we want the model with the smallest number of coefficients that also gives a good accuracy.To do this, we find the value of lambda that gives the simplest model but also lies within one standard error of the optimal value of lambda. This ensures the model is the most regularised while still minimising the error. The resulting value of lambda is 0.0299147. The selected covariates are then used to fit the logistic regression model.

Table 8.5 shows the results of a logistic regression model of stunting (0-23 months).

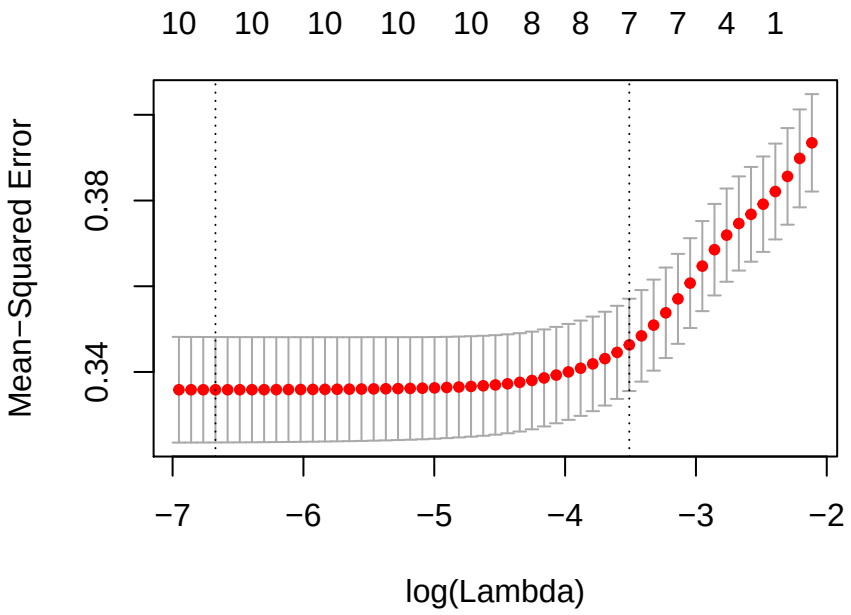


Figure 8.1: Cross validation results

The following observations were made:

- Being male increases the chance of being stunted.
- Children born in TC Newman have a greater chance of being stunted.
- Children with taller mothers have a lower chance of being stunted.
- Tobacco during pregnancy increases the chance of being stunted. The p-value of this coefficient is 0.18 however the Lasso has indicated that this variable has predictive power.
- Alcohol during pregnancy increases the chance of being stunted.
- Having a socio-economic class of low increases the chance of being stunted.
- The greater the gestation age, the lower the chance of being stunted.

Table 8.5: Logistic regression model (0-23 months)

	Estimate	Std. Error	z value	Pr(>|z|)
(Intercept)	15.6421	2.51276	6.22507	<0.001
SexMale	0.82326	0.17539	4.69391	<0.001

	Estimate	Std. Error	z value	Pr(>|z|)
ClinicMbekweni	-0.3454	0.20738	-1.66555	0.0958
Maternal height	-0.03917	0.01291	-3.03479	0.00241
Tobacco during pregnancyNo	-0.29276	0.21987	-1.33153	0.18301
Alcohol during pregnancyNo	-0.61498	0.22207	-2.76929	0.00562
Socio-economic quartile: LowNo	-0.63516	0.18877	-3.36465	<0.001
Gestation Age	-0.24976	0.03609	-6.92022	<0.001

Table 8.6 shows the results of the logistic regression model of stunting (24-48 months). The following observations were made:

- Children born in TC Newman have a greater chance of being stunted in years 2 to 4.
- Children with taller mothers have a lower chance of being stunted in years 2 to 4.
- Tobacco during pregnancy increases the chance of being stunted in years 2 to 4.
- An early gestational age (<38 weeks) increases the chance of being stunted in years 2 to 4.

Table 8.6: Logistic regression model (24-48 months)

	Estimate	Std. Error	z value	Pr(>|z|)
(Intercept)	17.35842	2.95007	5.88407	<0.001
Maternal height	-0.0867	0.01626	-5.33058	<0.001
Tobacco during pregnancyNo	-0.68241	0.21885	-3.11812	0.00182
Socio-economic quartile: LowNo	-0.99812	0.21742	-4.59081	<0.001
Gestation age	-0.11383	0.03803	-2.99281	0.00276

8.2.2 Decision Trees

Tree based learning algorithms are one of the most widely used supervised learning methods. Unlike linear models, as in logistic regression, they can map non-linear relationships. Decision trees can be a useful tool for data exploration as they can easily identify the most significant variables as well as relationships between variables.

Decision trees can be used for classification problems by splitting the population into homogeneous sub-populations based on the most significant splitting input variable. In order for the tree to decide at which variable to split, a measure of node impurity (variability within nodes), called the gini index is used,

$$G = \sum_j G_j \quad and \quad G_i = \sum_k \hat{p}_{jk}(1 - \hat{p}_{jk})$$

where $\hat{p}_{jk}$ is the proportion of observations in response category $k = 1, ..., K$ within node $j = 1, ..., J$. We would like the terminal nodes to be as homogenous as possible, ideally only including observations of only one response category. The smaller the gini index, the purer the node. Therefore at each step of the tree growth, the split that produces the greatest reduction in the Gini index is chosen.

Figure 8.2 shows the decision tree of stunting within the first 24 months of life.

The figures in the plot can be interpreted as follows:

- The top number in the node refers to the way that node is voting (1 for stunted and 0 for non-stunted)
- The numbers in the second line of the node represent the proportion split of that population, based on the way that node is voting.
- The bottom number in the node indicates the percentage of the population that resides in that node.
- The last layer of nodes are called terminal nodes. The predicted value of an observation will be the most commonly occuring class in this node.

In the first node, which is 100% of the population, we can see that 27% are stunted. The first split is based on early gestation, indicating that this is the most significant contributor to the probability of being stunted. For those with an early gestational age, which is 24% of the population, 48% are stunted. For this group, tobacco during

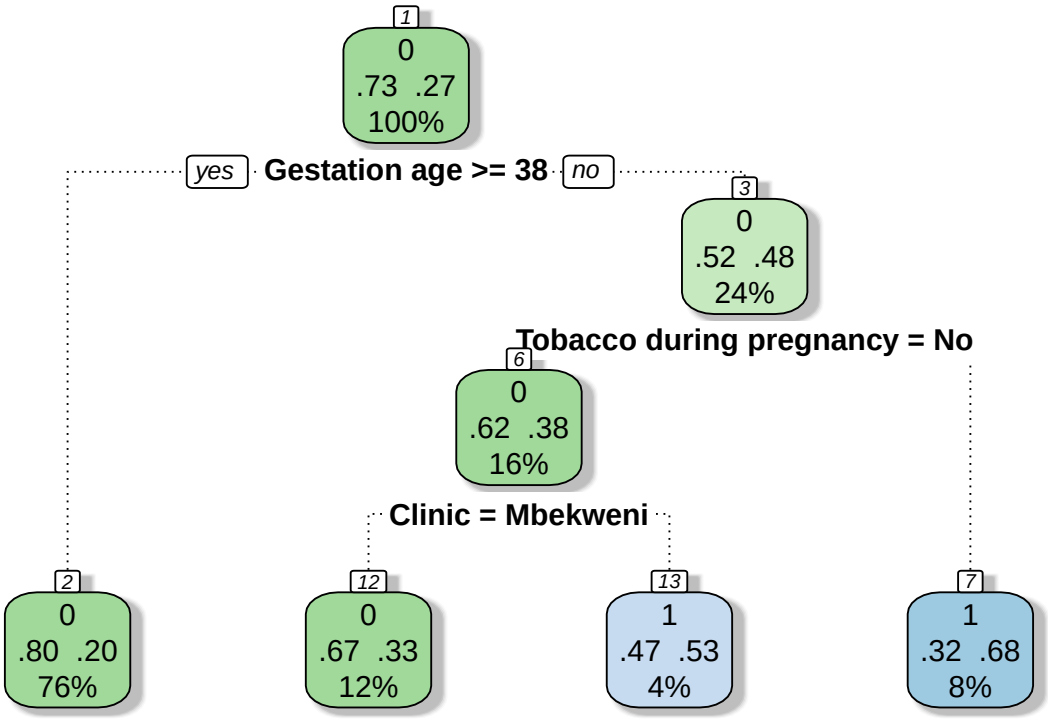


Figure 8.2: Decision tree of stunting within the first 24 months

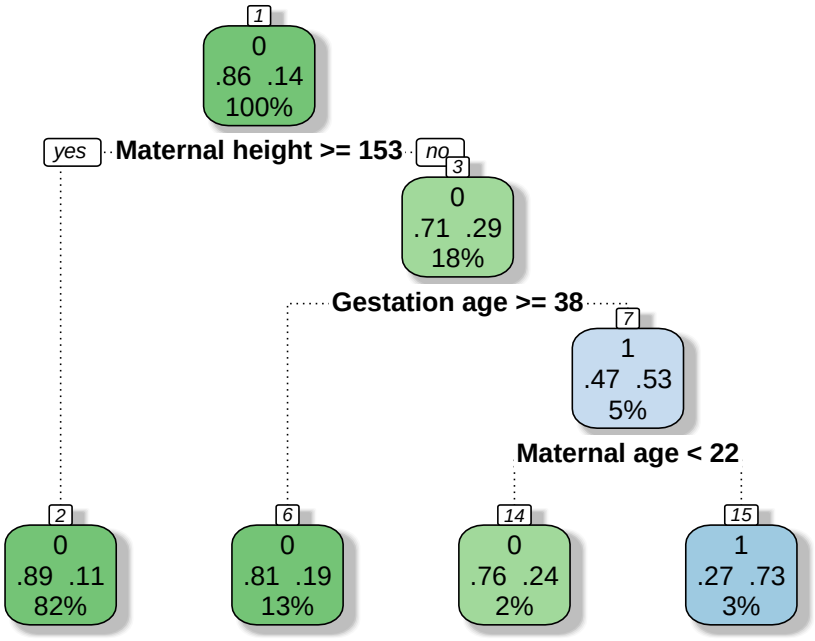


Figure 8.3: Decision tree of stunting within the second 24 months

pregnancy is the next most significant contributor. Of the early gestation babies who's mothers did have tobacco during pregnancy, 32% are stunted. Of the early gestation babies who's mothers did not have tobacco during pregnancy, 62% are not stunted. As we continue down the tree, we see which combination of variables are most likely to result in a child being stunted. For this tree, we used a minimum splitting criteria of 50. Therefore there must exist at least 50 cases of stunting in a node in order for a split to be attempted.

Figure 8.3 shows the decision tree of stunting within the second 24 months of life. We can see that maternal height, early gestation and maternal age are the most important predictors of stunting. Since there are fewer cases of stunting in the second 24 months of life, the minimum splitting criteria were reduced to 30 cases.

8.3 Classification

In order to predict whether a child will be stunted based on the covariates available at birth, we compare the logistic regression model to a random forest model. The data is split into a training set (60%) and testing set (40%). Table 8.7 shows the split of stunted to non-stunted children in the original dataset compared with the training and testing datasets. We can see that the ratio is similar in each dataset.

Table 8.7: Data split

	Non stunted	Stunted
Original	0.7313	0.2687
Train	0.7411	0.2589
Test	0.7164	0.2836

First we predict stunting within the first 24 months by using the logistic regression model. The model parameters were derived using the training dataset and prediction done on the testing dataset. In order to determine the classification threshold to use, i.e. the probability at which the model classifies stunting, we evaluate the Receiver Operating Characteristic (ROC) curve for each model. At each threshold value, the true positive rate (sensitivity) and false positive rate (1-specificity) are calculated and plotted. This shows the trade-off between the rate at which the model correctly predicts stunting with the rate of incorrectly predicting stunting.

Sensitivity = number of true positives / (number of true positives + number of false negatives)

Sensitivity and specificity are commonly used to evaluate the performance of a predictive model. Sensitivity indicates how often the prediction is correct when the actual outcome is stunted, i.e. the proportion of stunted children that were correctly identified by the model as being stunted. A highly sensitive model is reliable when it does predict stunting. However it is not useful for ruling out stunting since it does not take into account false positives, i.e. those who are incorrectly predicted to be stunted.

Figure 8.4 shows the ROC graph for each model. The Area Under the Curve (AUC) summarises the overal performance of the model over all possible thresholds. The

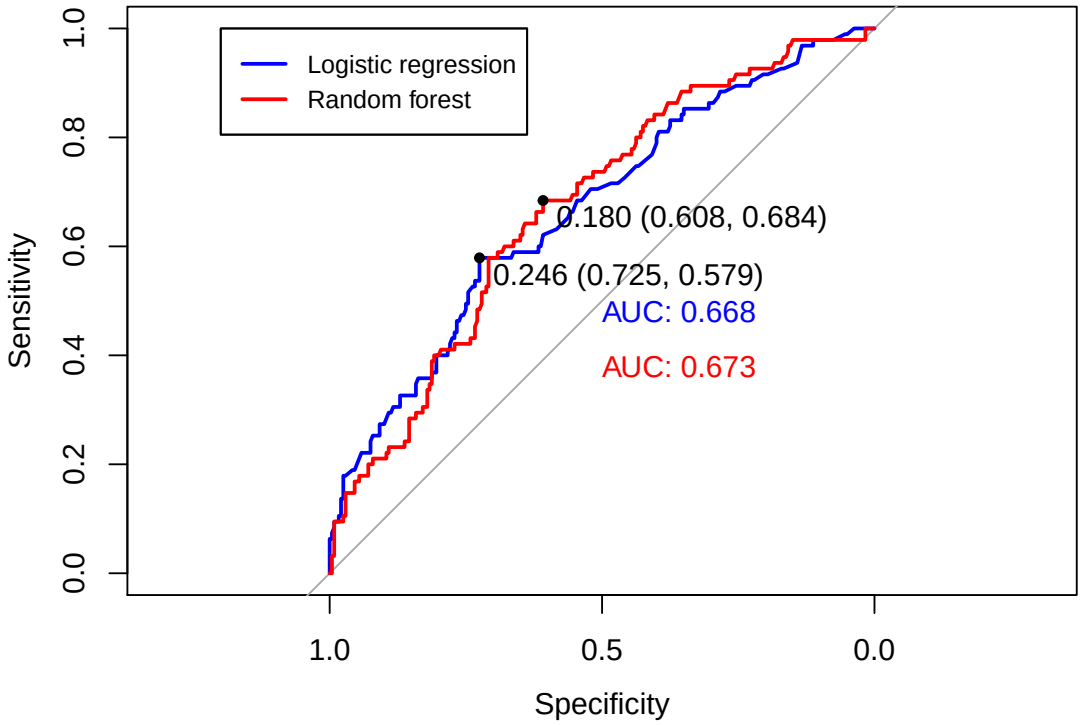


Figure 8.4: ROC comparison

higher the AUC, the better the classifier. The random forest model has a greater AUC than the logistic regression model. The points plotted on each curve shows the best threshold for the model with the highest sum of sensitivity + specificity. For the logistic regression model this is 0.246, with a resulting sensitivity of of 0.579, and for random forest model this is 0.18, with a resulting sensitivity of 0.684.

Table 8.8 shows the confusion matrix of the logistic regression prediction.

Table 8.8: Logistic Regression 24 month stunting confusion matrix

	Actual: Not stunted	Actual: Stunted	Sum
Predicted: Not stunted	174	40	214
Predicted: Stunted	66	55	121
Sum	240	95	335

The accuracy (proportion of correctly classified observations) of the logistic regression model predictions is 0.6836.

Next we predict stunting using a random forest model. The random forest algorithm builds trees similar to the way a decision tree does. However, the bagging (bootstrap aggregating) algorithm is used to create random samples of the data with which to train the model. Given a dataset, say D1, it creates dataset D2 by sampling n cases with replacement. About 1/3 of the rows from D1 are excluded, known as the Out Of Bag (OOB) samples. The model trains using D2 and calculates an unbiased estimate of the error using the OOB sample. Every time a split has to be made, it uses only a subset of the inputs to make the split, instead of the full set of inputs. It builds multiple trees using the same process, and then takes the average of all the trees to obtain the final model.

The parameters that need to be set in the random forest model are the number of variables randomly sampled as candidates at each split and the number of trees to grow in the forest. Figure 8.5 shows the accuracy (through repeated cross validation) for different combinations of these variables. We can see that the most accurate model occurs with 1500 trees and 3 candidate variables at each split.

The model was then trained using the training dataset and predicted values computed for the testing dataset. Table 8.9 shows the confusion matrix of the random forest model.

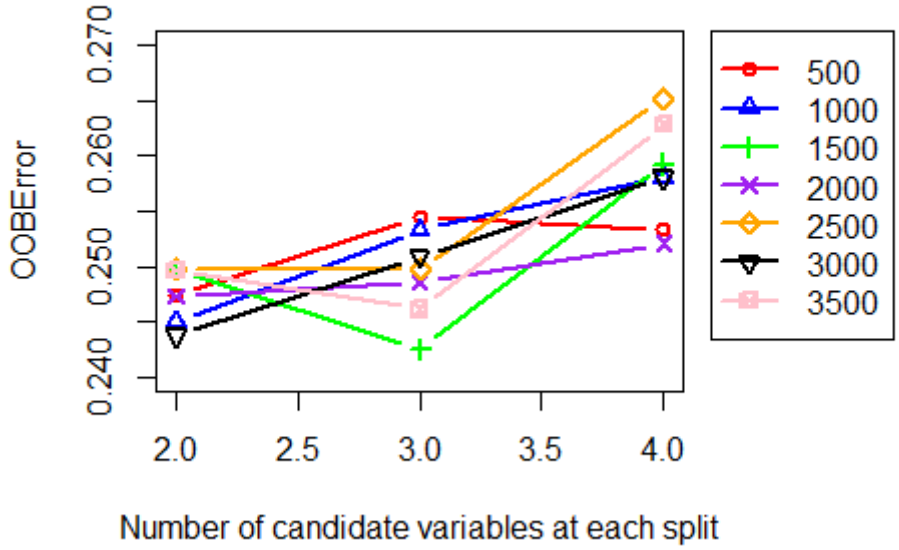


Figure 8.5: Random forest parameter selection

Table 8.9: Random Forest 24 month stunting confusion matrix

	Actual: Not stunted	Actual: Stunted	Sum
Predicted: Not stunted	146	30	176
Predicted: Stunted	94	65	159
Sum	240	95	335

The accuracy of the random forest model predictions is 0.6299. This is a smaller prediction accuracy than the logistic regression model accuracy of 0.6836. However other measures of classification performance may be more informative than the accuracy. Table 8.10 shows a comparison of the predictive performance of the logistic regression model and the random forest. The random forest model has a higher sensitivity/recall but lower specificity than the logistic regression model. The logistic regression model outperforms the random forest model in precision. This represents the accuracy of a stunting prediction, i.e. how often the prediction is correct when stunting is predicted. The balanced accuracy, ($\frac{sensitivity+specificity}{2}$), is higher for the logistic regression model.

Table 8.10: Comparison of logistic regression and random forest 24 month stunting model

	Logistic regression	Random forest
Sensitivity	0.5789	0.6842
Specificity	0.725	0.6083
Precision	0.4545	0.4088
Balanced Accuracy	0.652	0.6463

8.4 Interpretation of predictive models

We can assess which variables were most important in the random forest model using the variable importance plot. Figure 8.6 shows the variable importance plot of the random forest model, derived using the training dataset. The MeanDecreaseAccuracy on the x-axis shows how much removing each variable reduces the accuracy of the model.

Comparing the logistic regression model with the random forest, we need to take interpretability as well as predictive power into account. In order to identify the factors that are more likely to result in stunting, we interpret the contribution of each factor in the predictive model. For logistic regression, we do this by evaluating the parameter coefficients. Table 8.11 shows the parameter coefficients of the logistic regression model, derived using the training dataset. For a random forest, we evaluate the variable importance using variable importance plots.

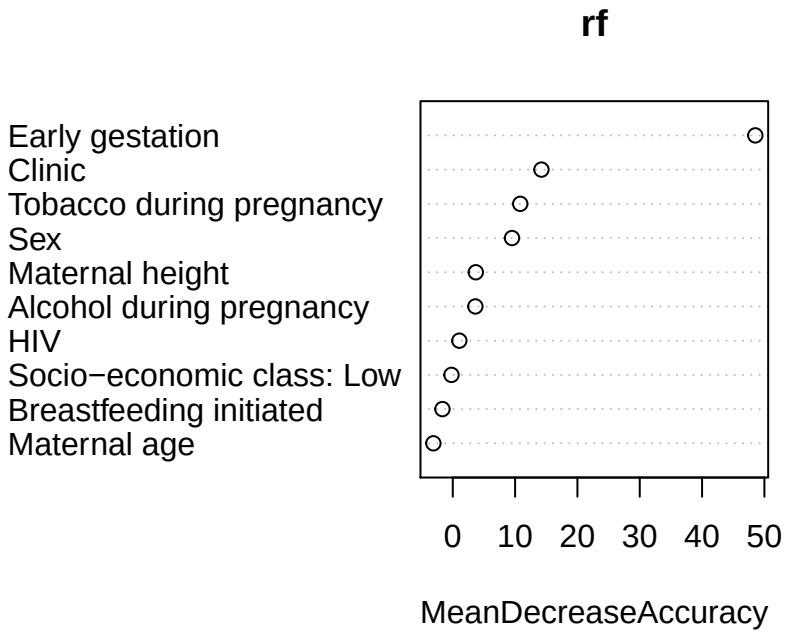


Figure 8.6: Random forest variable importance

Table 8.11: Logistic regression model (0-23 months)

	Estimate	Odds Ratio	Std. Error	z value	Pr(>|z|)
(Intercept)	7.91038	2725.427	2.70944	2.91956	0.00351
SexMale	0.90167	2.464	0.23202	3.88615	<0.001
Maternal height	-0.04664	0.954	0.01707	-2.73231	0.00629
Tobacco during pregnancyNo	-0.51156	0.6	0.25448	-2.01024	0.04441
Alcohol during pregnancyNo	-0.66167	0.516	0.27986	-2.36429	0.01806
Early gestationNo	-1.64823	0.192	0.24308	-6.78066	<0.001

Comparing the variable importance plot to the logistic regression coefficients, we see that in both cases, early gestation has the greatest magnitude. Tobacco and alcohol during pregnancy are also identified by both models as being important predictors. The models differ in their relative importance of sex and clinic. The disadvantage of the random forest interpretation is that we do not know in which direction the variable is affecting the outcome, whereas in the regression results the sign of the coefficient indicates whether the effect of the variable is to increase or decreases the log odds of

9 Conclusions

In this dissertation we evaluated various methods for assessing growth and stunting in a South African birth cohort. Our intent was to find the most appropriate growth curve to model height over time and identify predictors of the growth trajectory. The potential of using an Artificial Neural Network as an alternative modelling strategy to the more conventional mixed-effect model was explored. Various mixed-effect models were fit and compared to neural networks in terms of model fit, interpretability of parameters as well as predictive power.

The mixed-effect modelling followed a progressive approach in which unconditional models were first fit using non-structural forms (i.e. polynomials) and then structural forms, where the model parameters have a biological interpretation. Linear and non-linear mixed-effect models were used. Subsequently, conditional models were fit where the effects of covariates on the curve shape parameters were investigated.

We required a model function that captured the rapid growth following birth and then subsequent deceleration in growth. Therefore various common growth curves used for modelling height in children were fit. This included the linear, quadratric, cubic and fractional polynomial models, the Count model, the Berkey-Reeds model, the Logistic model, the Gompertz model, the Exponential model, the Karlberg model and the Jenss-Bayley model. The Berkey-Reeds 2nd order model was shown to have the best fit. The parameters of this model are birth height, linear growth rate, deceleration in growth velocity and two growth inflection points. The Berkey-Reeds 2nd order model was then used to fit the conditional mixed-effects model.

The covariates that were used were sex, clinic, maternal height, maternal age, gestational age, breastfeeding, alcohol during pregnancy, tobacco during pregnancy, low socio-economic quartile and HIV status. We assessed the effect of each covariate on the various shape parameters of the Berkey-Reeds model and then introduced them into the model in a stepwise approach.

This revealed that having an early gestational age lowers the birth height, male babies have a higher birth height, tobacco during pregnancy lowers the birth height, taller mothers have babies with higher birth heights, a positive HIV status lowers the birth height, TC Newman babies have a smaller linear growth rate than Mbekweni babies, male babies have a smaller linear growth rate, having an early gestational age increases

the deceleration in growth velocity, belonging to the lowest socio-economic class decreases the deceleration in growth velocity, taller mothers have babies with a higher deceleration in growth velocity, alcohol during pregnancy is associated with a smaller inflection point and male babies have a larger second inflection point.

The variable selection process used in deriving the conditional mixed-effect model was to examine the effect of all covariates on each of the model features successively, starting with the intercept and moving to the other parameters. The approach to this could have rather been to test a covariate on all model parameters before considering the next covariate, especially since the random effects are allowed to be correlated.

We then evaluated Artificial Neural Networks (ANNs) for the purpose of modelling child height. The standard approach to modelling longitudinal data is to use a mixed-effect model since they take account of the within-subject association through the addition of subject-specific effects. ANNs allow for a nonlinear model to be developed with little knowledge about the actual relationship that exists between variables. However, they do not take account of the repeated measures per subject. Three approaches were explored to introduce subject-level conncections in the network: an age-specific intercept network, age-specific networks, and a subject-dummy variable network.

We saw that neural networks can be a useful tool for providing insight into the variables affecting child growth. In particular, the age-specfic networks allowed us to determine the covariates affecting child growth at each age. The age-specific intercept has the advantage of allowing for one model to be fit across all ages, due to the nonlinear modelling capability of neural networks. The subject-dummy model explicity allowed for a subject-specific effect to be incorporated and enabled us to understand the common characteristics that apply to child growth by seperating the common effects and subject-specific effects.

Comparing the mixed-effect model with the neural network, the structural nature of the Berkey-Reeds model allows for a more in-depth understanding of how certain risk factors affect various aspects of child growth. The neural network has no associated mathematical formula and therefore interpretation can be difficult. However, with the use of the Garson method, variable importance can be determined from the neural network, allowing us to derive insights into the risk factors affecting child growth.

For the purpose of growth prediction using covariates available at birth, we compared

the conditional Berkey-Reeds model with the age-specific intercept neural network by splitting the data into training and testing sets. The Berkey-Reeds model had a smaller RMSE. However the predictive performance of the neural network showed promising results and did not greatly under-perform relative to the mixed model.

Although attempts were made to separate the subject specific effects from the common effects in the neural network, none of the approaches take into account the within-subject correlation. Possible future work could be to use the 'mixed-effect neural network', described by Tandon, Adak, and Kaye (2006), to add a random effect after the nonlinear relationship has been established by the neural network. Another approach would be to treat the subject input seperately from the other inputs by feeding the subject variable into a second hidden layer, as described by Maity and Pal (2013). The semiparametric panel data model using neural networks, by Crane-Droesch (2017), may be a step towards effectively modelling longitudinal data using neural networks.

Lastly, models for childhood stunting were fit and compared. Stunting is defined as a shortfall in height relative to age and is an indication that a child is failing to develop. Logistic regression was used to explain the relationship between various pre- and post-natal risk factors and stunting. The results were compared to a random forest model which revealed that both models show the same order of variable importance. The random forest model also appeared to perform similarly in terms of predictive power.

Both models had a relatively poor performance in predicting stunting in this birth cohort. As seen in figure 4.3, the heights for this cohort were generally lower than expected using the WHO criteria. This brings into question the appropriateness of using the WHO references and whether South Africa specific references should be developed. Using lower height-for-age z-score cut-off points for stunting such as $z<-2.5$ or $z<-3$ could have been more appropriate and hence resulted in improved predictive models.

A Appendix A

A.1 Mixed-effect model results

A.1.1 Linear model

Table A.1: Linear2 Random Effects

	StdDev	Corr
(Intercept)	1.49462858	(Intr)
age	0.08874764	0.578
Residual	4.57820956	

Table A.2: Linear2 Covariance Estimates

	(Intercept)	age
(Intercept)	2.234	0.07668
age	0.07668	0.007876

A.1.2 Quadratic model

Table A.3: Quadratic3 Random Effects

	StdDev	Corr
(Intercept)	2.156354760	(Intr)
age	0.170229645	-0.001
I(age^2)	0.003328908	0.014
Residual	2.937013404	

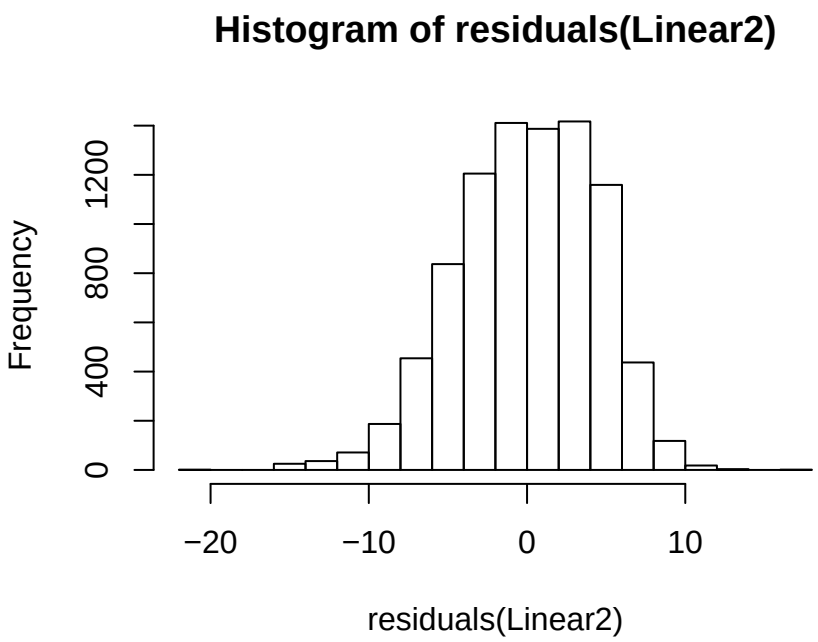


Figure A.1: Linear2 residuals

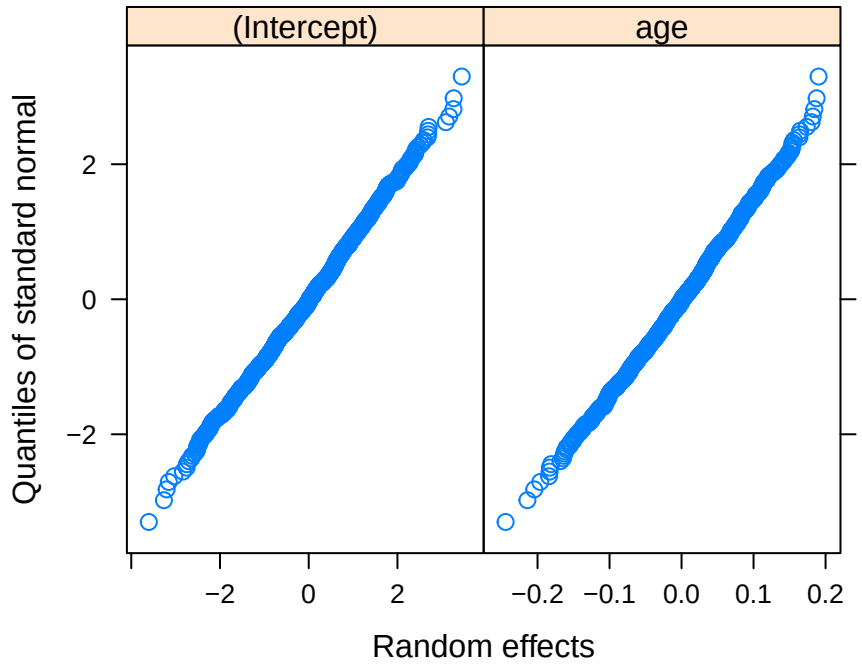


Figure A.2: Linear2 Random effects normal plot

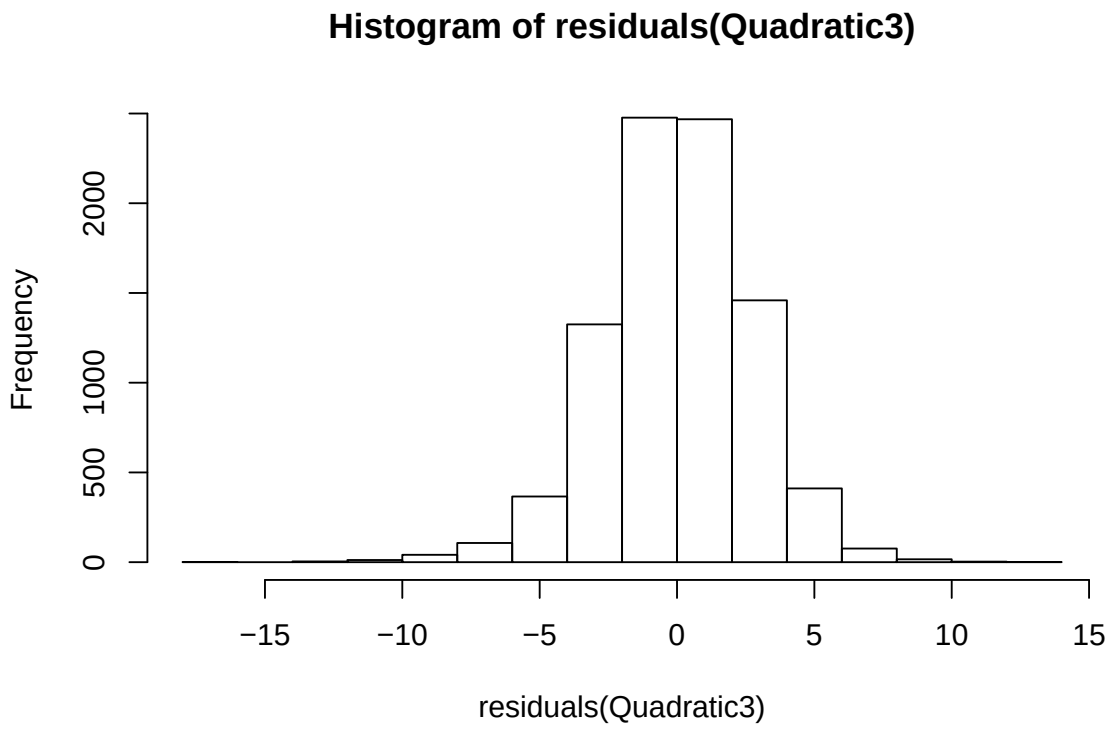


Figure A.3: Quadratic3 residuals

Table A.4: Quadratic3 Covariance Estimates

	(Intercept)	age	I(age^2)
(Intercept)	4.65	-0.0003432	0.0001022
age	-0.0003432	0.02898	-0.0004991
I(age^2)	0.0001022	-0.0004991	0.00001108

A.1.3 Cubic model

Table A.5: Cubic3 Random Effects

	StdDev	Corr
(Intercept)	2.510564301	(Intr)

	StdDev	Corr
age	0.186321306	-0.249
I(age^2)	0.002883006	0.259
Residual	2.268447616	

Table A.6: Cubic3 Covariance Estimates

	(Intercept)	age	I(age^2)
(Intercept)	6.303	-0.1165	0.001875
age	-0.1165	0.03472	-0.0004928
I(age^2)	0.001875	-0.0004928	0.000008312

A.1.4 Fractional model

Table A.7: Fractional3 Random Effects

	StdDev	Corr
(Intercept)	3.4308656	(Intr)
age	0.1187967	0.460
I(log(age + 2))	1.7609330	-0.705
Residual	2.1419414	

Table A.8: Fractional3 Covariance Estimates

	(Intercept)	age	I(log(age + 2))
(Intercept)	11.77	0.1875	-4.261
age	0.1875	0.01411	-0.1468
I(log(age + 2))	-4.261	-0.1468	3.101

A.1.5 Count model

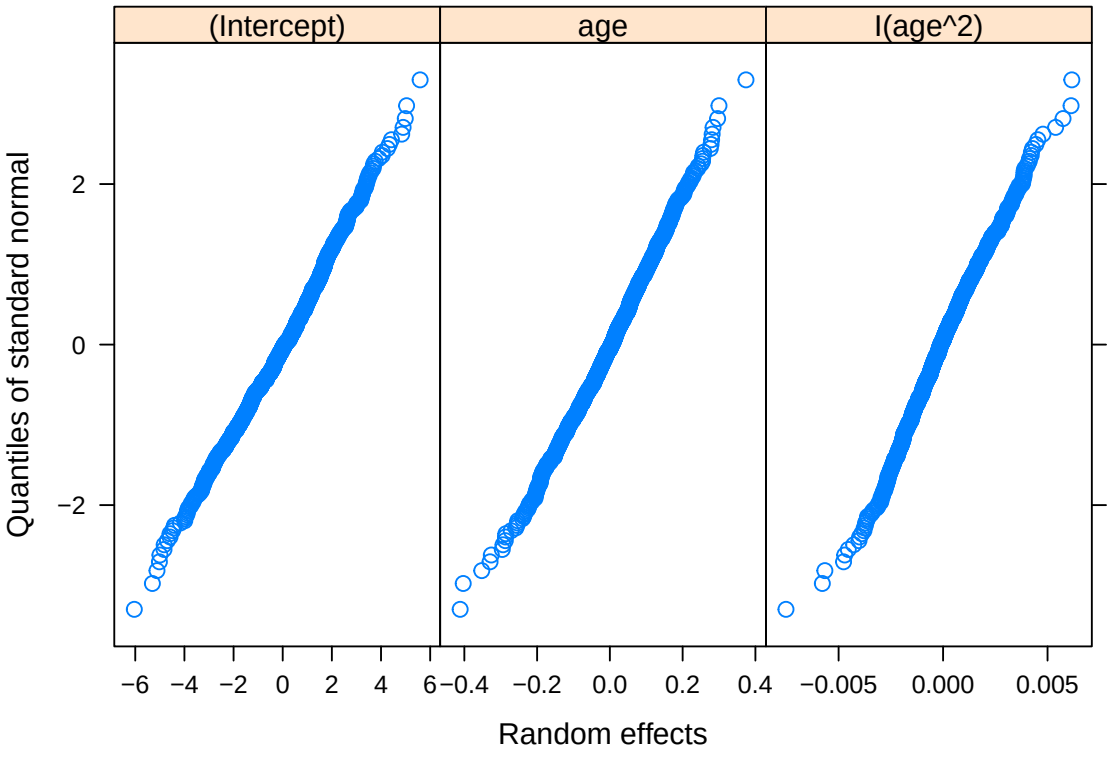


Figure A.4: Quadratic3 Random effects normal plot

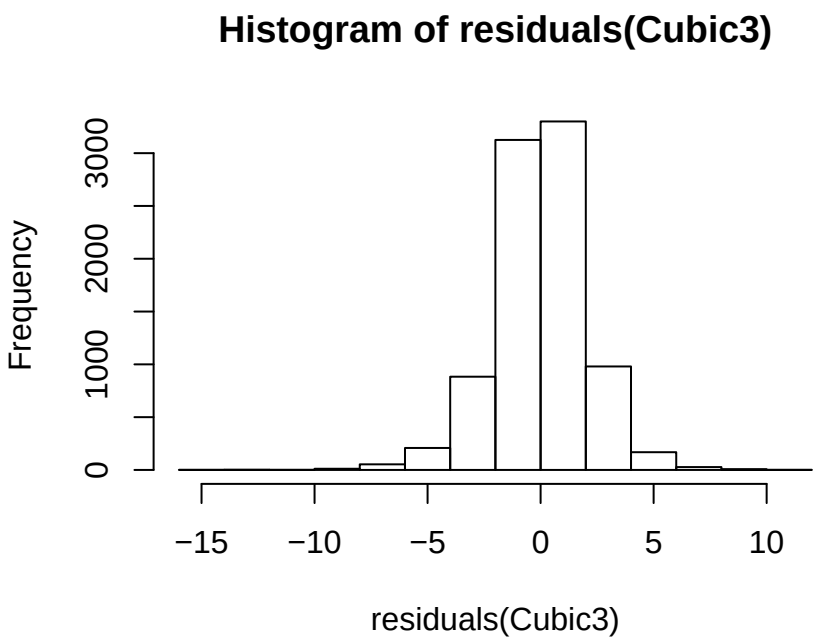


Figure A.5: Cubic3 residuals

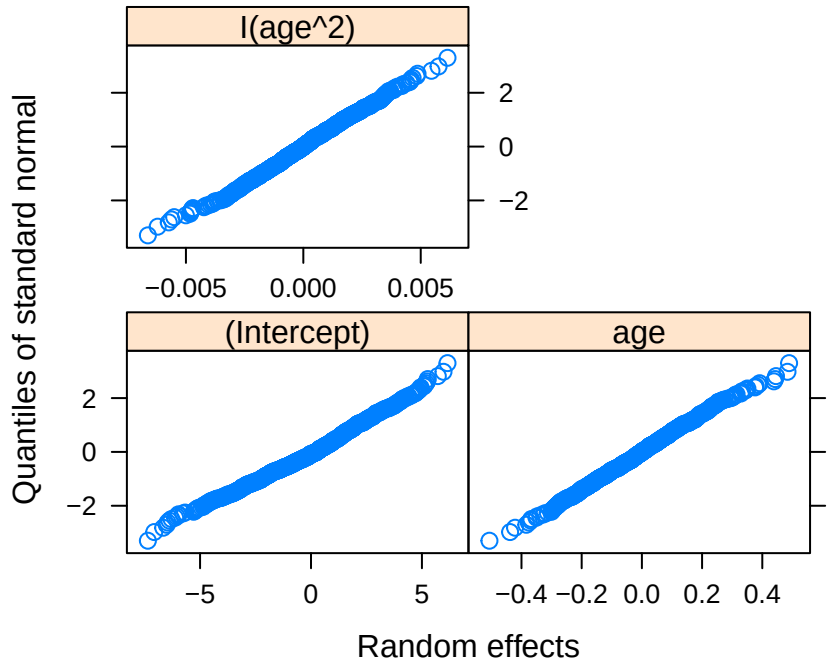


Figure A.6: Cubic3 Random effects normal plot

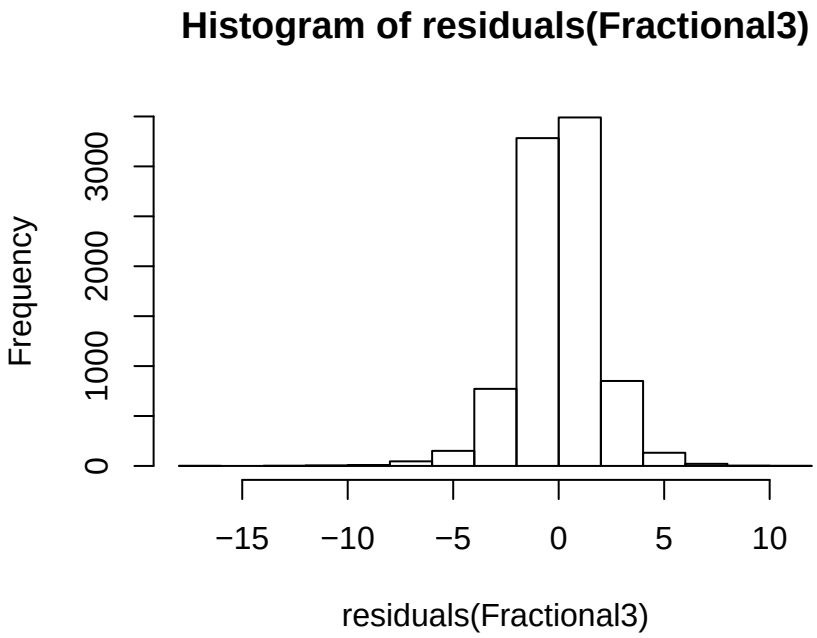


Figure A.7: Fractional3 residuals

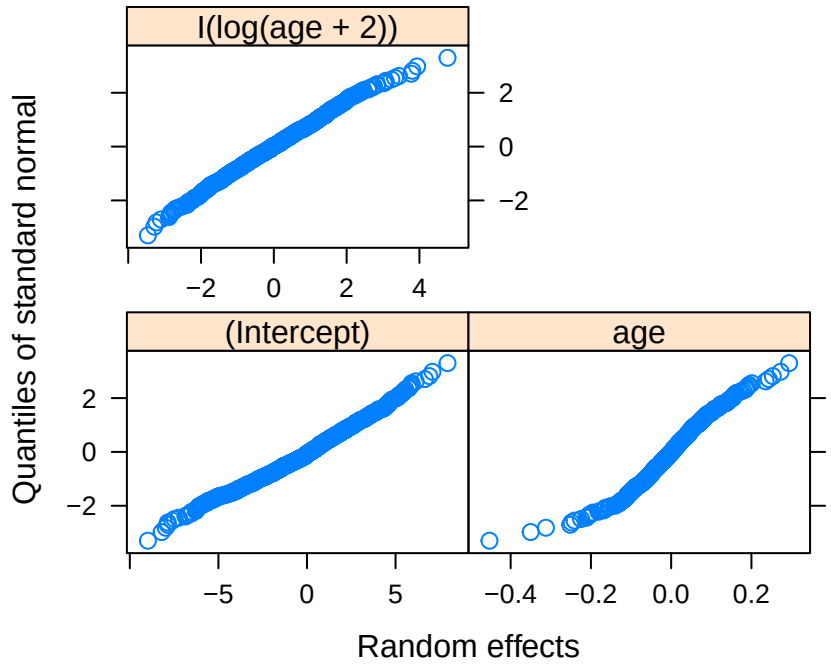


Figure A.8: Fractional3 Random effects normal plot

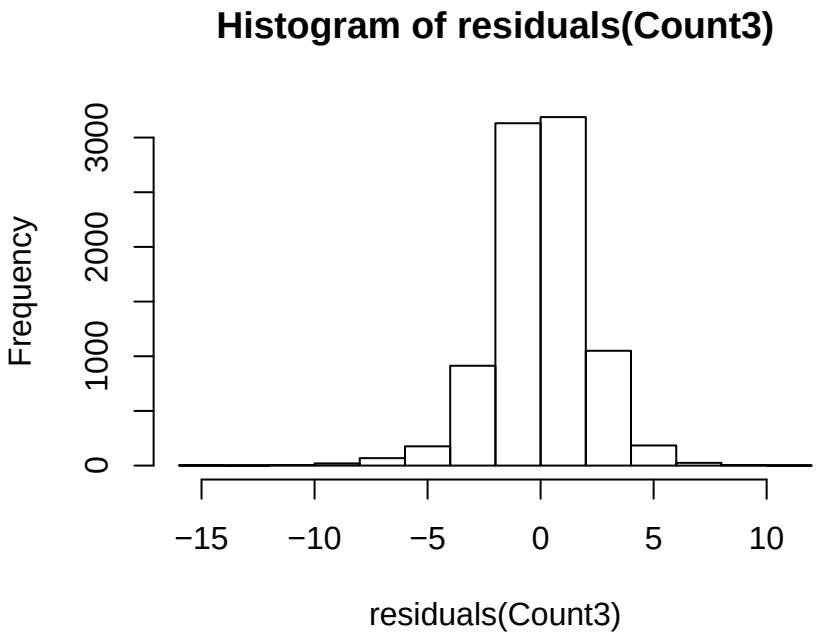


Figure A.9: Count3 residuals

Table A.9: Count3 Random Effects

	StdDev	Corr
(Intercept)	2.66250286	(Intr)
age	0.09657565	0.247
I(log(age + 1))	1.13028430	-0.463
Residual	2.32858805	

Table A.10: Count3 Covariance Estimates

	(Intercept)	age	I(log(age + 1))
(Intercept)	7.089	0.06363	-1.394
age	0.06363	0.009327	-0.05723
I(log(age + 1))	-1.394	-0.05723	1.278

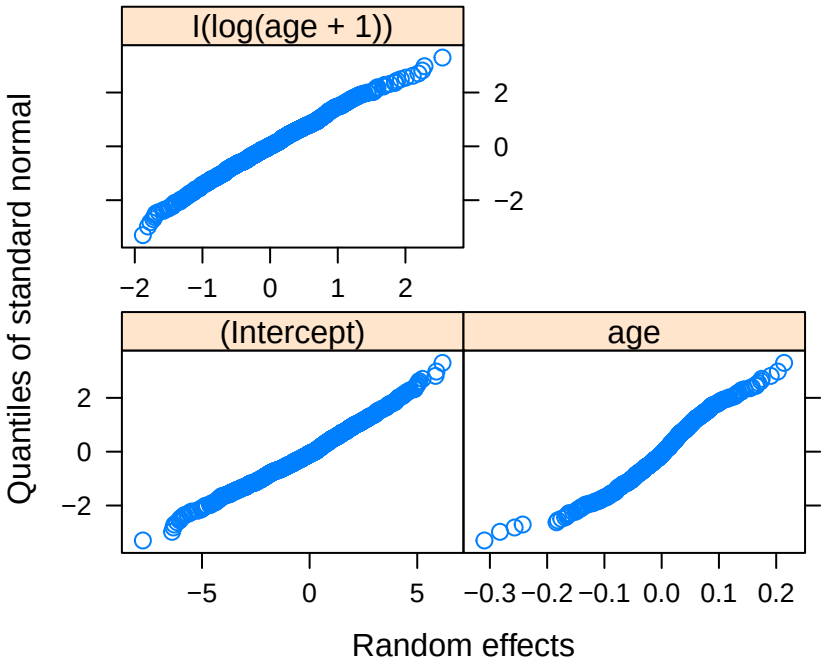


Figure A.10: Count3 Random effects normal plot

A.1.6 Berkey-Reeds 1st order model

Table A.11: BerkeyReed1.4 Random Effects

	StdDev	Corr
(Intercept)	6.2218255	(Intr)
age	0.1164029	0.566
I(log(age + 1))	2.7818495	-0.889
I(1/(age + 1))	5.7885423	-0.892
Residual	2.0209073	

Table A.12: BerkeyReed1.4 Covariance Estimates

	(Intercept)	age	I(log(age + 1))	I(1/(age + 1))
(Intercept)	38.71	0.4098	-15.39	-32.14
age	0.4098	0.01355	-0.2334	-0.3206
I(log(age + 1))	-15.39	-0.2334	7.739	14.09
I(1/(age + 1))	-32.14	-0.3206	14.09	33.51

A.1.7 Berkey-Reeds 2nd order model

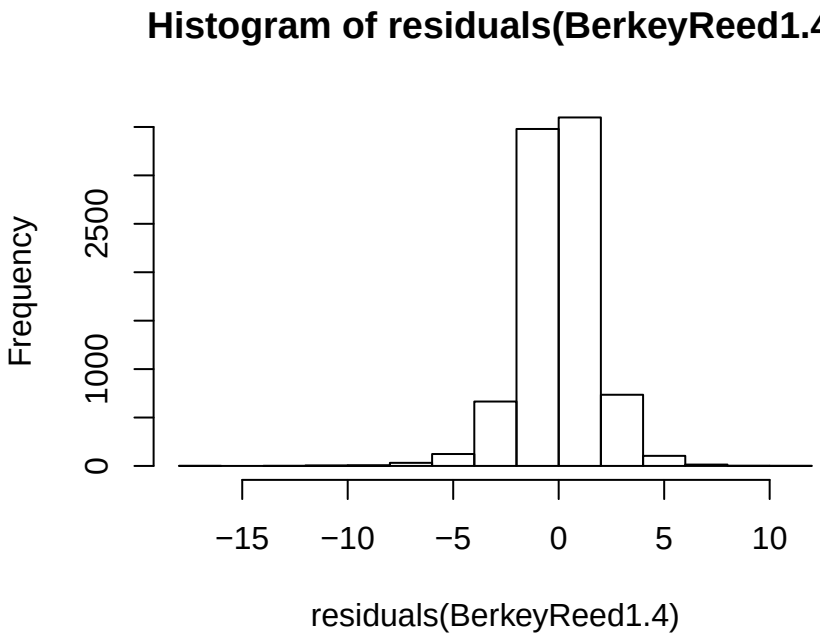


Figure A.11: BerkeyReed1.4 residuals

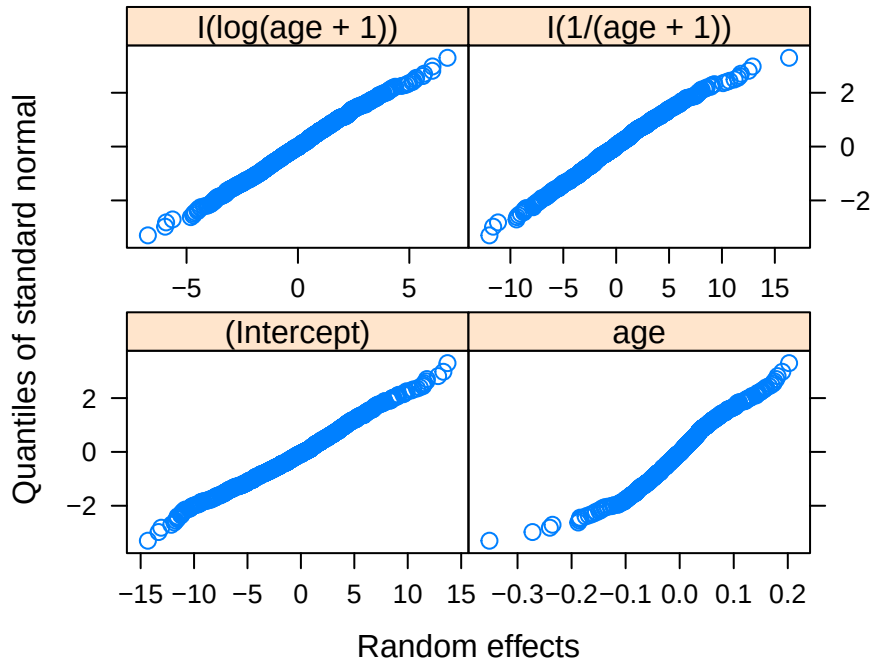


Figure A.12: BerkeyReed1.4 Random effects normal plot

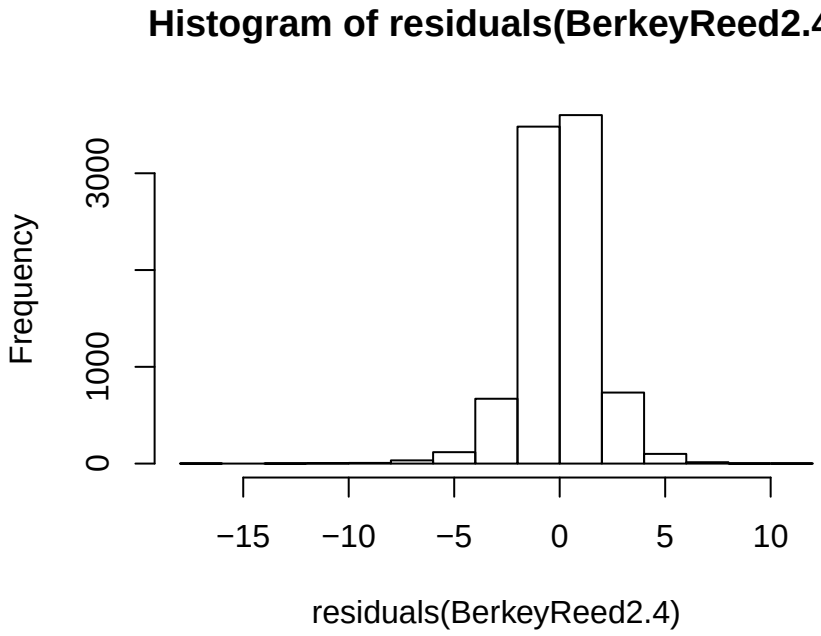


Figure A.13: BerkeyReed2.4 residuals

Table A.14: BerkeyReed2.4 Covariance Estimates

	(Intercept)	age	I(log(age + 1))	I(1/(age + 1))
(Intercept)	43.53	0.4772	-17.37	-36.52
age	0.4772	0.01411	-0.257	-0.3799
I(log(age + 1))	-17.37	-0.257	8.511	15.86
I(1/(age + 1))	-36.52	-0.3799	15.86	37.51

A.1.8 Logistic model

Table A.15: Three parameter logistic model Random Effects

	StdDev	Corr
phi1	4.265617	phi1
phi2	1.157305	0.557

	StdDev	Corr
Residual	2.889520	

A.1.9 Gompertz model

Table A.16: Gompertz Model Random Effects

	StdDev	Corr	NA
alpha	6.48486422	alpha	beta
beta	0.05667145	0.718	
mu	0.01329644	0.763	0.594
Residual	2.57972870		

A.1.10 Exponential model

Table A.17: Exponential Model Random Effects

	StdDev	Corr
beta0	1.2156887328	beta0
beta1	0.0007422079	0.913
Residual	5.5336596280	

A.1.11 Jenns Bayley model

Table A.18: Jenss Bayley Model Random Effects

	StdDev	Corr	NA
beta0	3.3709207	beta0	beta1
beta1	0.0920468	-0.317	
beta2	0.1751734	0.687	-0.395
Residual	2.0714552		

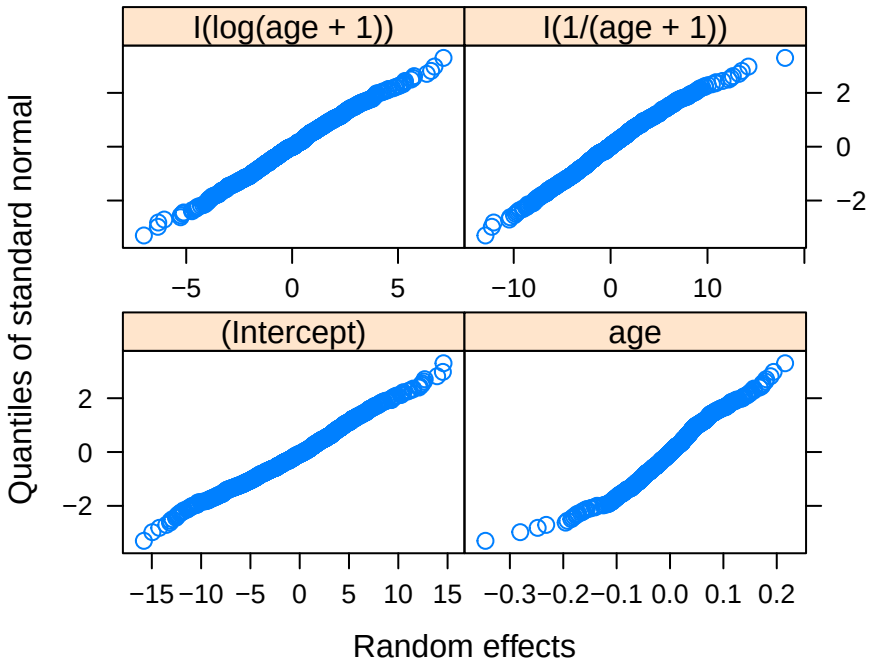


Figure A.14: BerkeyReed2.4 Random effects normal plot

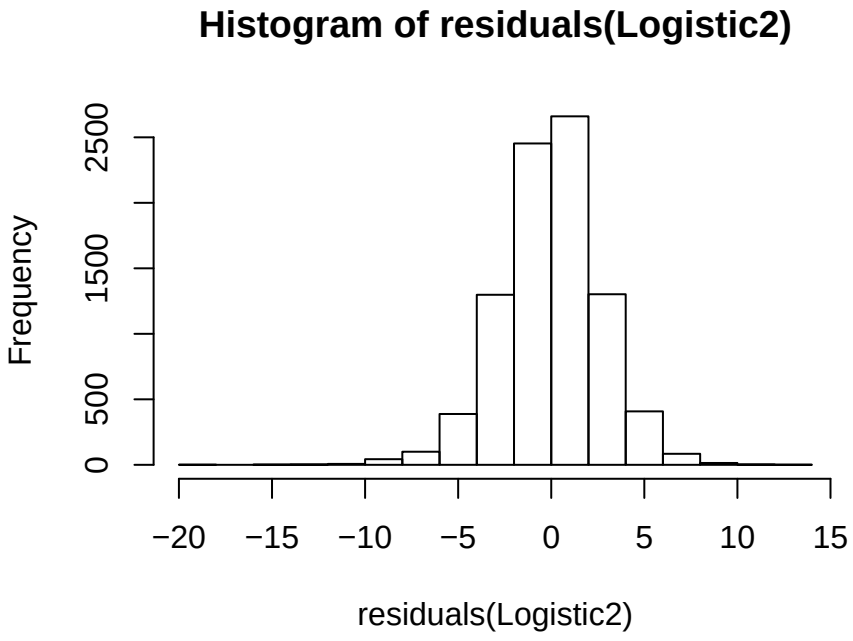


Figure A.15: Logistic2 residuals

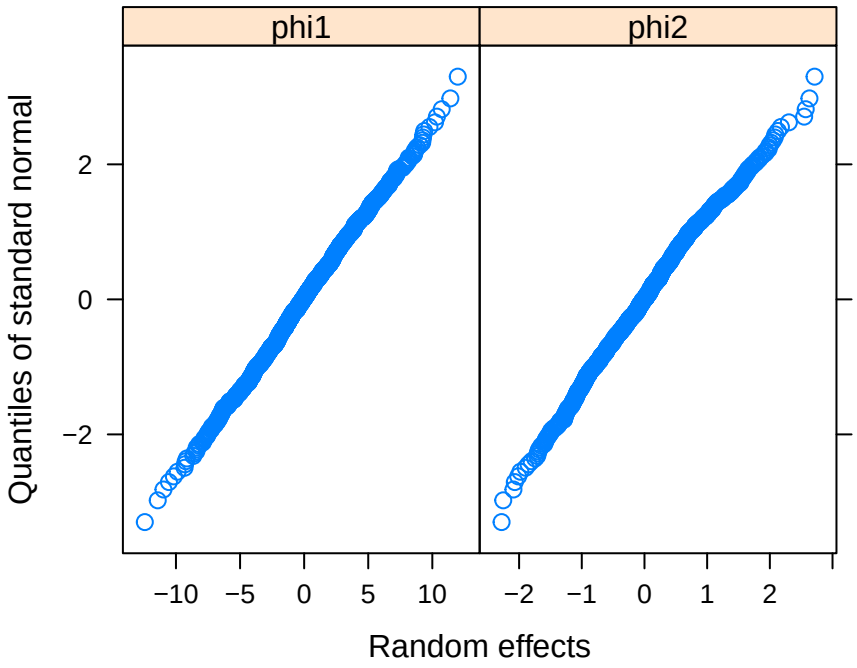


Figure A.16: Logistic2 Random effects normal plot

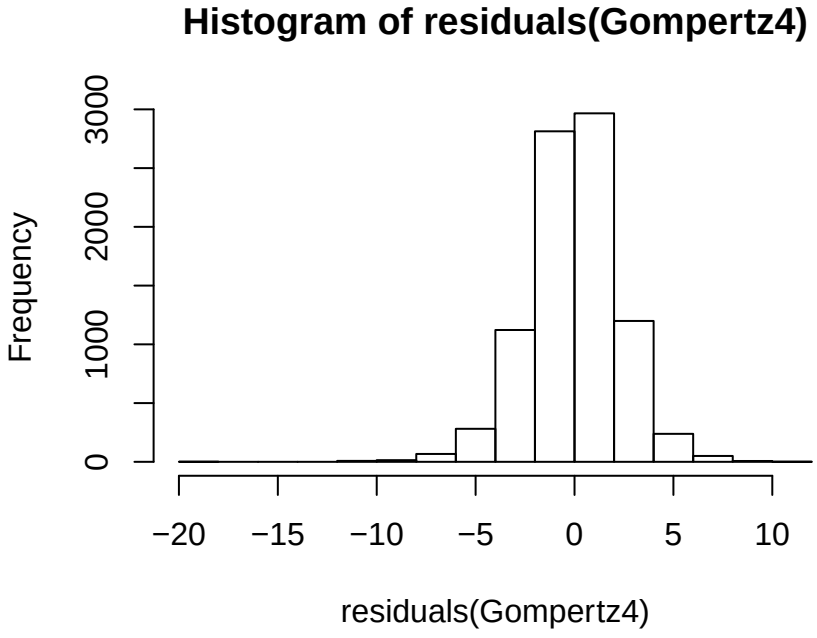


Figure A.17: Gompertz4 residuals

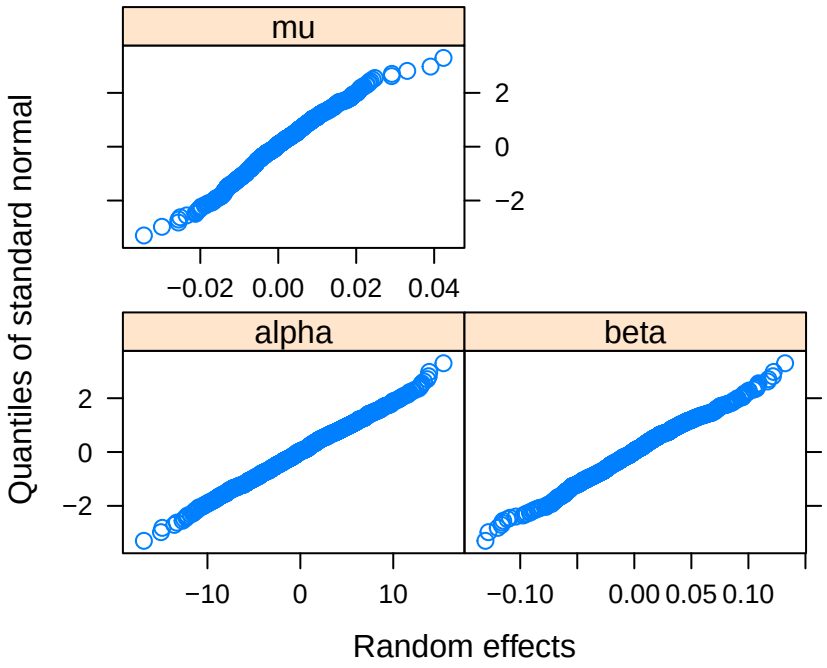


Figure A.18: Gompertz4 Random effects normal plot

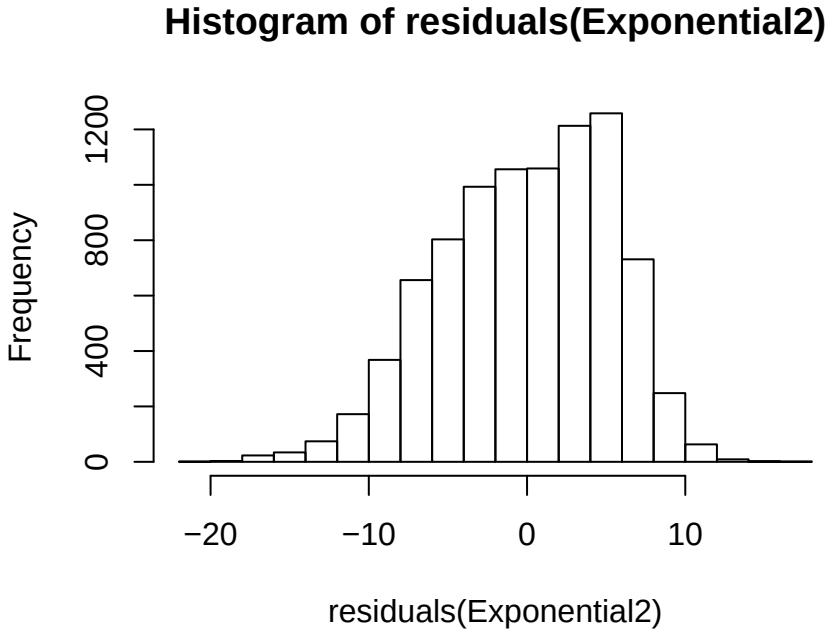


Figure A.19: Exponential2 residuals

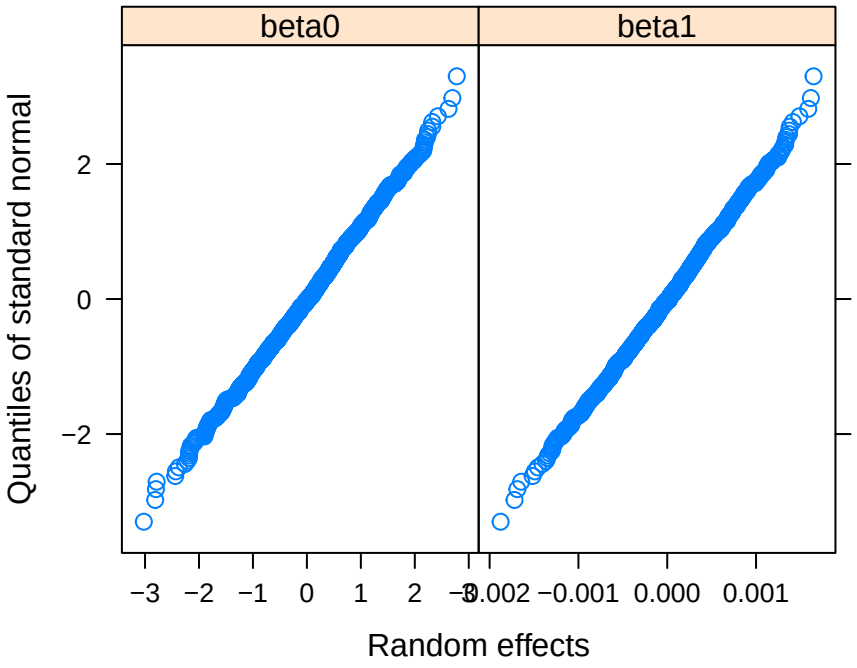


Figure A.20: Exponential2 Random effects normal plot

A.1.12 Karlberg model

Table A.19: Karlberg Model Random Effects

	StdDev	Corr	NA
beta0	2.5058412	beta0	beta1
beta1	6.6946221	0.118	
beta2	0.0124898	-0.290	-0.804
Residual	2.4239577		

Table A.20: Final conditional model random effects

	StdDev	Corr
Intercept	4.55786426293	(Intr)
age	0.05525458301	0.109
ln(age)	2.02145636754	-0.836
1/age	4.22106850255	-0.854
Residual	2.06644440814	

Table A.21: Final conditional model covariance estimates

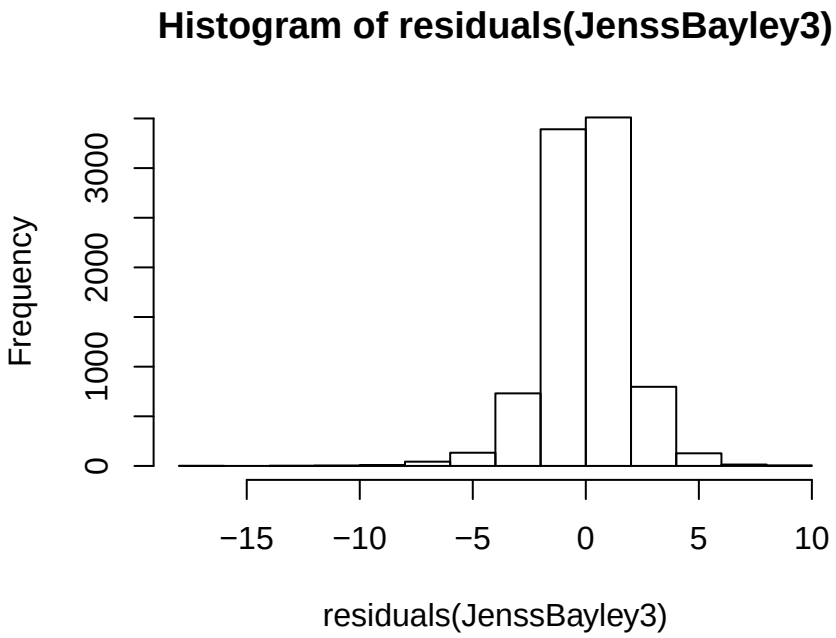


Figure A.21: JennsBayley3 residuals

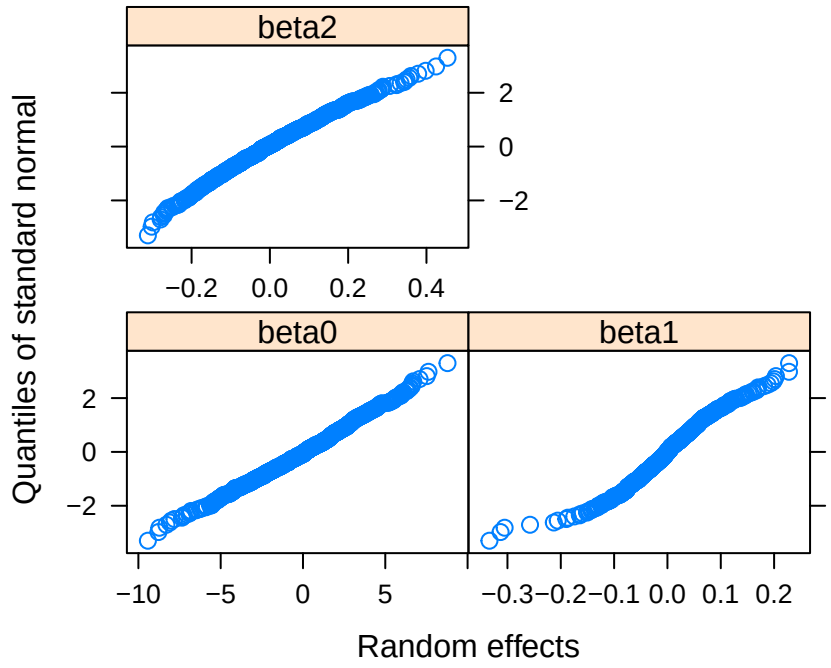


Figure A.22: JennsBayley3 Random effects normal plot

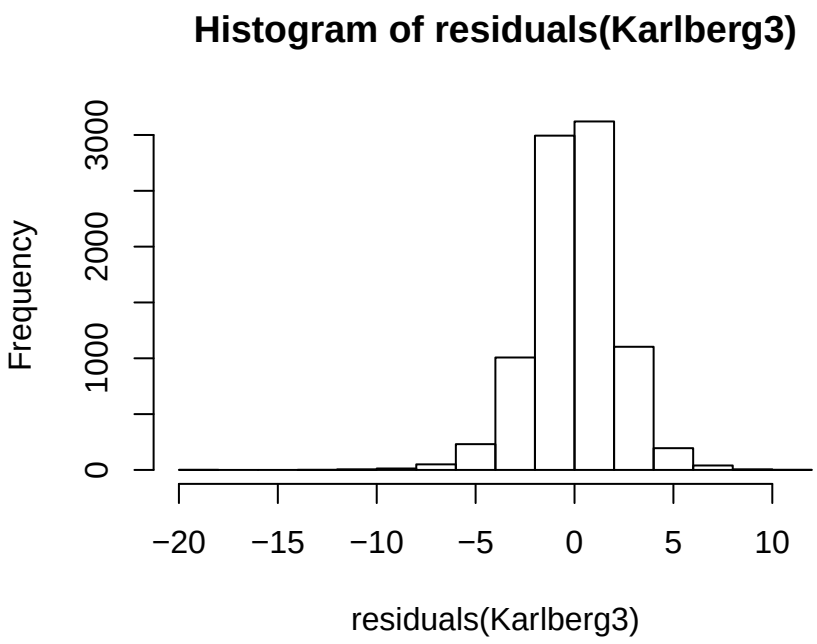


Figure A.23: Karlberg3 residuals

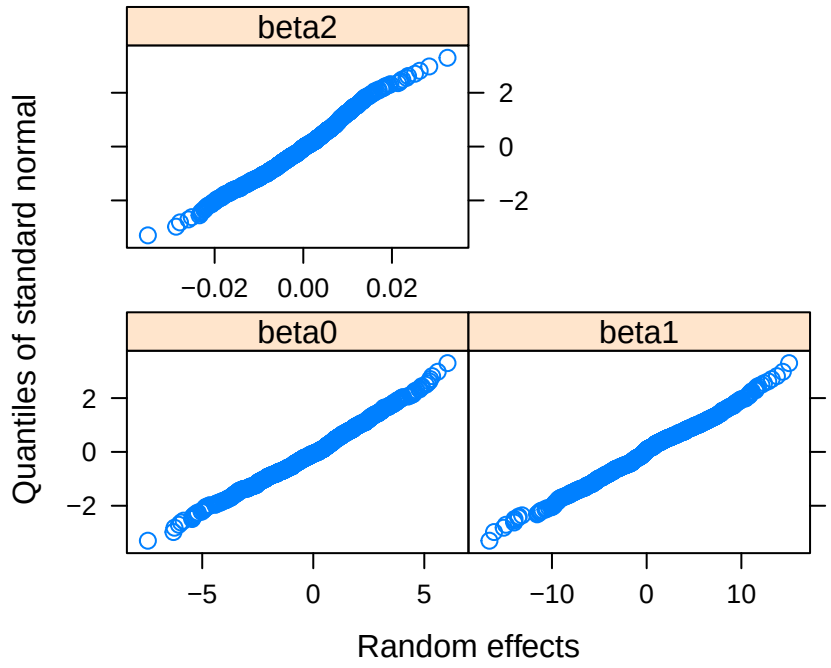


Figure A.24: Karlberg3 Random effects normal plot

References

Aggrey, S. E. 2009. "Logistic nonlinear mixed effects model for estimating growth parameters." *Poultry Science* 88 (2): 276–80. doi:10.3382/ps.2008-00317.

Akombi, Blessing Jaka, Kingsley Emwinyore Agho, John Joseph Hall, Dafna Merom, Thomas Astell-Burt, and Andre M.N. Renzaho. 2017. "Stunting and severe stunting among children under-5 years in Nigeria: A multilevel analysis." *BMC Pediatrics* 17 (1). BMC Pediatrics: 1–16. doi:10.1186/s12887-016-0770-z.

Berkey, C S. 1982. "Comparison of two longitudinal growth models for preschool children." *Biometrics* 38 (1): 221–34. http://www.ncbi.nlm.nih.gov/pubmed/7082759.

Black, M M, and A Krishnakumar. 1999. "Predicting longitudinal growth curves of height and weight using ecological factors for children with and without early growth deficiency." *Journal of Nutrition* 129 (2 SUPPL.): 539S–543S. doi:10.1093/jn/129.2.539S.

Black, Robert E, Cesar G Victora, Susan P Walker, Zulfiqar A Bhutta, Parul Christian, Mercedes de Onis, Majid Ezzati, et al. 2013. "Maternal and child undernutrition and overweight in low-income and middle-income countries." *Lancet (London, England)* 382 (9890): 427–51. doi:10.1016/S0140-6736(13)60937-X.

Budree, Shrish, Elizabeth Goddard, Kirsty Brittain, Shihaam Cader, Landon Myer, and Heather J. Zar. 2017. "Infant feeding practices in a South African birth cohort—A longitudinal study." *Maternal and Child Nutrition* 13 (3): 1–9. doi:10.1111/mcn.12371.

Chirwa, Esnat D., Paula L. Griffiths, Ken Maleta, Shane A. Norris, and Noel Cameron. 2014. "Multi-level modelling of longitudinal child growth data from the Birth-to-Twenty Cohort: a comparison of growth models." *Annals of Human Biology* 41 (2): 168–79. doi:10.3109/03014460.2013.839742.

Crane-Droesch, Andrew. 2017. "Semiparametric panel data models using neural networks," 1–15. http://arxiv.org/abs/1702.06512.

Dommelen, P. van, S. van Buuren, G. R J Zandwijken, and P. H. Verkerk. 2005. "Individual growth curve models for assessing evidence-based referral criteria in growth monitoring." *Statistics in Medicine* 24 (23): 3663–74. doi:10.1002/sim.2234.

Dwyer, J T, E M Andrew, C Berkey, I Valadian, and R B Reed. 1983. "Growth in 'new' vegetarian preschool children using the Jenss-Bayley curve fitting technique."

The American Journal of Clinical Nutrition 37 (5): 815–27. doi:10.1093/ajcn/37.5.815.

Ganesan, R., P. Dhanavanthan, C. Kiruthika, P. Kumarasamy, and D. Balasubramanyam. 2014. "Comparative study of linear mixed-effects and artificial neural network models for longitudinal unbalanced growth data of Madras Red sheep." *Veterinary World* 7 (2): 52–58. doi:10.14202/vetworld.2014.52-58.

Grimm, Kevin J., Nilam Ram, and Fumiaki Hamagami. 2010. "Nonlinear Growth Curves in Developmental Research" 8 (24): 4017–8. doi:10.1002/bmb.20244.DNA.

Hastie, Trevor, Robert Tibsharani, and Jerome Friedman. 2009. "Springer Series in Statistics The Elements of." *The Mathematical Intelligencer* 27 (2): 83–85. doi:10.1007/b94608.

Hilbe, Joseph M. 2011. "Logistic Regression." *Encyclopedia of Statistical Science*, 1–2. doi:10.1198/tech.2003.s30.

Johnson, William, Nagalla Balakrishna, and Paula L Griffiths. 2014. "Modelling physical growth using mixed effects models." *Am J Phys Anthropol* 150 (1): 612–25. doi:10.1002/ajpa.22128.Modeling.

Karkach, Arseniy S. 2006. "Trajectories and models of individual growth." *Demographic Research* 15. doi:10.4054/DemRes.2006.15.12.

Karlberg, J. 1987. "On the modelling of human growth." *Statistics in Medicine* 6 (March 1986): 185–92.

Maity, Tanmay Kumar, and Asim Kumar Pal. 2013. "Subject Specific Treatment to Neural Networks for Repeated Measures Analysis." *Proceedings of the International Multi Conference of Engineers and Computer Scientists* I: 60–65.

Pizzi, Costanza, Tim J. Cole, Camila Corvalan, Isabel dos Santos Silva, Lorenzo Richiardi, and Bianca L. De Stavola. 2014. "On modelling early life weight trajectories." *Journal of the Royal Statistical Society. Series A: Statistics in Society* 177 (2): 371–96. doi:10.1111/rssa.12020.

Riedmiller, M., and H. Braun. 1993. "A direct adaptive method for faster backpropagation learning: theRPROP algorithm." *IEEE International Conference on Neural Networks* 1 (7): 586–91. http://deeplearning.cs.cmu.edu/pdfs/Rprop.pdf.

Senbanjo, Idowu O., Ibiyemi O. Olayiwola, Wasiu A. Afolabi, and Olayinka C. Sen-

banjo. 2013. "Maternal and child under-nutrition in rural and urban communities of Lagos state, Nigeria: The relationship and risk factors." *BMC Research Notes*. doi:10.1186/1756-0500-6-286.

Simondon, Kirsten B., Francois Simondon, Francis Delpeuch, and Andre Cornu. 1992. "Comparative study of five growth models applied to weight data from congolese infants between birth and 13 months of age." *American Journal of Human Biology* 4 (3): 327–35. doi:10.1002/ajhb.1310040308.

Steele, Fiona. 2008. "Multilevel models for longitudinal data." *Journal of the Royal Statistical Society. Series A: Statistics in Society* 171 (1): 5–19. doi:10.1111/j.1467-985X.2007.00509.x.

Tandon, Reeti, Sudeshna Adak, and Jeffrey A. Kaye. 2006. "Neural networks for longitudinal studies in Alzheimer's disease." *Artificial Intelligence in Medicine* 36 (3): 245–55. doi:10.1016/j.artmed.2005.10.007.

Thibault, Hélène, Caroline Carriere, Coralie Langevin, Edouard Kossi Déti, Pascale Barberger-Gateau, and Sylvie Maurice. 2013. "Prevalence and factors associated with overweight and obesity in French primary-school children." *Public Health Nutrition* 16 (2): 193–201. doi:10.1017/S136898001200359X.

Tilling, Kate, Corrie MacDonald-Wallis, Debbie A. Lawlor, Rachael A. Hughes, and Laura D. Howe. 2014. "Modelling childhood growth using fractional polynomials and linear splines." *Annals of Nutrition and Metabolism* 65: 129–38. doi:10.1159/000362695.

World Health Organization. 2010. "Global Database on Child Growth and Malnutrition." *Chemistry & …* 2: 2–3. doi:10.1093/tandt/11.7.180.

Zar, H. J., W. Barnett, L. Myer, D. J. Stein, and M. P. Nicol. 2015. "Investigating the early-life determinants of illness in Africa: The Drakenstein Child Health Study." *Thorax* 70 (6): 592–94. doi:10.1136/thoraxjnl-2014-206242.