Characterizing Child Growth Trajectories

Craig Anderson[*]   Ryan Hafen[†]   Louise M. Ryan[‡]

Abstract

There is a wide and growing literature on growth curve modeling based on longitudinal data, including parametric and semi-parametric (spline-based) random effects models, functional data analysis methods and latent growth curve models. We compare and contrast these various methods in terms of how well they do in terms of predicting individual child growth trajectories, based on data from the HBDGki project. Prediction accuracy is assessed using a "leave one out" strategy for fitting and then comparing predicted values with the observed values of those left out. Methods are also presented for extracting key growth features such as faltering and catch-up. We conclude with recommendations about how growth data can be most effectively modeled in epidemiological cohort studies such as the ones encountered in the HBGDki project.

Key Words: splines, FDA, growth curve, longitudinal, goodness of fit, faltering

1. Introduction

The Healthy Birth, Growth and Development knowledge integration (HBGDki) project is a collaboration funded by the Bill and Melinda Gates Foundation with the aim of integrating information from an extensive database of studies of child growth and development from across the world. The majority of these studies are from low or middle income countries, and the ultimate goal of this project is to create a knowledge platform to inform decisions about interventions in these lower and middle income countries to help improve overall health and wellbeing.

It is therefore crucial to understand the factors and conditions which impact on the development of children. It is critical that we have reliable methods that allow us to characterise of different growth patterns. For example, we need to distinguish between children who are growing successfully and those whose growth is faltering [1]. When children falter, we wish to quantify the timing and the extent of their recovery [2] [3] [4]. Once we have identified methodology for characterising growth patterns, we can start to investigate the factors which predict faltering and recovery. In turn, this will allow exploration of the relationship between faltering and other outcomes such as cognitive development [5] [6] [7].

Scientists and health professionals identified the link between socio-economic status as long ago as the early 19th century [8]. A historical overview of growth modelling is provided by [9], but this paper will focus primarily on modern statistical approaches. One of the first such methods proposed was the LMS method [10], a precursor to the SITAR method which is in common use today [11]. There has been a recent surge in the use of statistical growth modelling approaches for child development as a result of the increasing availability of computer software [12] [13]. The bulk of the existing literature has been based on raw growth measurements, rather

[*]School of Mathematical and Physical Sciences, University of Technology Sydney
[†]Department of Statistics, Purdue University
[‡]School of Mathematical and Physical Sciences, University of Technology Sydney

than modelling growth relative to a global or local standard. Most people will be familiar with the use of standardized growth charts to assess how an individual child compares to the population distribution for children of the same age and gender. These charts allow us to quantify a child's relative height or weight at a particular age using what is known as a Z-score. If we continue to monitor a child's progress as they get older, we will obtain multiple Z-scores at different ages. These Z-scores can be used to identify whether a child's relative growth is improving or declining over time. Standard Deviation Scores (SDS, [14]) can be computed to quantify these changes in Z-scores over time, but assessing the significance of such changes requires sophisticated consideration of the expected variation in centile crossing as well as the potential impact of regression to the mean [15] [16] [17].

SDS-based measurements identify growth changes between two timepoints for an individual child, and do not provide us with a mechanism for characterising more complex growth patterns. Many epidemiological studies, including those in the HBGDki database, involve children observed at multiple timepoints, and these timepoints are not necessarily the same for each child. In this paper we suggest that faltering and recovery can be investigated more extensively by estimating a growth trajectory for each child using longitudinal modelling techniques. The idea of extracting indicators and measures of child growth rates from fitted longitudinal models is not new. Grajeda et al. (2016) [18] consider linear mixed effects models based on regression splines, and derive the associated derivatives of these models to characterize child-specific growth rates. However, the application discussed in their paper is based on modelling of raw growth data rather than Z-scores. One important contribution of our paper will be to explore Z-score modelling in more detail, and to provide a quantitative comparison to raw data modelling.

Section 2 provides a detailed description of the data available in the HBGDki database, including a discussion of the types of outcome data which will be modelled. In Section 3, we provide an overview of the growth modelling literature and discuss their advantages and disadvantages in our context. In Section 4, a selection of these methods are applied to a variety of datasets from the HBDGki database, and we use a novel validation approach to test their efficacy. We conclude with a discussion in Section 5.

2. Data

The HBDGki project is an ongoing initiative which has so far amassed data from well over 100 studies. At the time of analysis, 23 of these studies contained data with sufficient longitudinal measures of the two main child growth outcomes, height (or length) and weight. In total, these longitudinal datasets contain around 800,000 observations made on over 100,000 children.

Table 1 provides a summary of the studies which were analysed in this paper. For data confidentiality reasons, we have labelled these datasets using letters rather than references to their location or source. The second column outlines the number of children in each study, while the third column displays the total number of height (and/or weight) measurements within the study. The next three columns provide the minimum, maximum and median number of observations per child, while the final three columns show the minimum, maximum and median age range over which

children were measured.

The studies vary in terms of the number of growth observations per child, with a median of 23 observations per child for Dataset C compared to just two per child for Dataset A. The studies cover a wide range of ages; for example Dataset K covers children from birth to roughly 18 months, while Datasets F and H measure subjects all the way to adulthood. The variety of these data makes it difficult to propose a 'one size fits all' modelling approach, but in this paper we seek to make recommendations which allow for a degree of consistency in the analysis of the datasets. This consistency will be crucial for the final aims of the HBGDki project, which include characterising growth patterns across multiple studies, and combining the results from these studies to identify global trends in growth.

Table 1: Summary of relevant studies within the HBGDki project

Dataset	# Children	# Num Obs	Obs Per Child			Child Age (in days)		
			Min	Max	Median	Min	Max	Median
A	7637	18983	1	4	2	168	927	541
B	197	2352	1	15	14	95	1903	804
C	373	12478	23	37	34	1	1111	558
D	3125	35506	1	37	9	1	1846	446
E	197	4405	10	41	21	1	702	116
F	20510	158892	1	19	6	1	6954	718
G	380	8436	2	26	23	1	1175	343
H	1544	28823	1	77	16	1	6954	2746
J	315	2548	1	13	10	119	493	269
K	203	1427	1	17	7	1	521	136
L	27363	122139	1	6	5	1	960	92
M	2954	41587	1	69	13	0	900	309
N	2144	46499	1	25	25	1	732	336
P	629	11828	1	43	21	1	2199	644
Q	529	8656	1	30	16	731	6696	3257
R	153	1839	1	16	13	1	679	185
S	412	2279	1	8	7	193	1282	628
T	16898	174233	1	14	11	1	3287	275
U	700	9741	1	16	16	1	756	175
V	302	1140	2	4	4	153	457	265
W	278	3177	1	33	13	1	525	211
X	2027	15637	1	10	8	18	1095	280
Y	14086	64867	1	10	5	1	1132	115

3. Methods

There is a large literature on growth modelling and it is not the purpose of this paper to provide an exhaustive review. However in this section we provide a broad overview of the different approaches that can be used for modelling growth data and provide some key references. Our focus lies in the characterization of individual growth trajectories, and therefore we consider only methods of relevance for longitudinal studies and do not discuss the extensive literature on the analysis of

cross-sectional growth data. While the range of growth trajectory methods are quite varied, they have an underlying commonality in that they hypothesise that individual children vary stochastically about a population curve [18]. In other words, it is assumed that there exists an overall mean curve for a particular population, and the differences between children can be explained as deviations from this mean curve.

Consider a study which observes the growth of N children over time in terms of a particular growth measurement. This measurement might reflect a physical characteristic such as height or weight, or may represent a mental characteristic such as a cognitive score. Suppose that the ith child has this growth measurement taken at a series of timepoints $t_{i1}, t_{i2}, \ldots t_{in_i}$ and let Y_{ij} represent the growth measurement taken at time t_{ij}. Note that there may be different numbers of growth measurements for each child, and that the measurements are not necessarily taken at regular intervals.

Clearly the growth of a child will depend on both their age and their gender. There are two broad ways to account for this; either the age and gender can be built into the modelling process, or the model can be based on age and gender standardised versions of the growth measurements. The majority of the papers on growth modelling work with raw growth data [10] [11] [13] [18] [19]. Our contribution will be to explore statistical approaches which can be used to model the standardised Z scores. Explicitly modelling age and gender effects might be interesting from a biological perspective [12], but the trade-off is that some of our degrees of freedom are used to capture the actual growth patterns rather than focusing on the trends. In this paper, we will compare modelling techniques based on raw and standardised data. Note that trajectories modelled under one approach can easily be converted to the other for illustrative purposes, so the purpose of our comparison is to see which form of data should be modelled on. Our standardised data is based on height and weight for age Z-scores (HAZ or WAZ), calculated with respect to the World Health Organisation standard population [20].

This section will discuss the existing growth methodologies, with a particular focus on the six proposed growth models which will be compared in Section 4.

3.1 Laird and Ware Linear Model

In an early paper that laid the groundwork for much of the last several decades of work on longitudinal growth curve modelling, Laird and Ware (1982) [21] proposed the use of random effects as a means of characterising child-specific departures from a global mean. Their approach allows each child to have a random intercept and slope via the following model

$$Y_{ij} = \beta_0 + \beta_1 t_{ij} + \gamma_{0i} + \gamma_{1i} t_{ij} + \epsilon_{ij}. \tag{1}$$

Here, γ_{0i} represents the ith child's deviation from the global intercept β_0 and γ_{1i} is their deviation from the global slope β_1. Here $(\gamma_{0i}, \gamma_{1i})$ are assumed to follow a joint normal distribution, independent of the error term ϵ, which also follows a normal distribution. This model is applicable on both the raw and standardised scale; in either case we would fit an individual straight line through our data for each child.

3.2 Laird and Ware Quadaratic Model

It is clear that the random intercept and slope model is very simplistic and is unable to capture nuances such as growth faltering and catchup. However, it is straightforward to extend this formulation to capture more complex non-linear trends. For example, one can add a quadratic time effect as follows:

$$Y_{ij} = \beta_0 + \beta_1 t_{ij} + \beta_2 t_{ij}^2 + \gamma_{0i} + \gamma_1 i t_{ij} + \gamma 2 i_i t_{ij}^2 + \epsilon_{ij}, \tag{2}$$

where γ_{2i} is an addtional random effect, representing the ith child's departure from the global quadratic term. Note that to avoid confusion we will hereafter refer to this method as *lwquad*, and use *lwlinear* to refer to the linear version outlined in the previous subsection. Higher degree polynomials can also be accounted for by adding further parameters to this formulation in a similar manner. However, such fully parametric approaches may struggle to capture the true growth trajectory, and we may be able to capture subtle aspects of the data more accurately using more flexible models. Spline-based approaches provide a more flexible framework for modelling individual growth trajectories, and have therefore been used extensively in this field.

3.3 SITAR

Cole et al. (2010) [11] proposed a method known as Superimposition by Translation and Rotation (SITAR), which involves each individual having a curve which is a shifted and transformed version of the mean growth curve. Shifting the curve up or down corresponds to mean changes, shifting it left or right corresponds to different growth times and the transformation of the curves.

The SITAR model is defined as follows

$$Y_{ij} = \omega_i + h\left(\frac{t_{ij} - \lambda_i}{\exp(-\kappa_i)}\right) \tag{3}$$

where ω_i, λ_i and κ_i are subject-specific random effects, and h is a natural cubic spline curve with $h(t)$ representing the mean curve.

A key advantage of the SITAR approach is that it describes each trajectory in terms of three biologically interpretable parameters. The parameter ω_i adjusts for child-specific differences in height, λ_i accounts for different timing of growth spurts, and κ_i accounts for different durations of growth spurts. The actual growth curve $h()$ forms an explicit part of the model. A consequence of this is that it is more natural to fit SITAR to the raw data, unlike the other methods outlined in this section which can be applied to either type of data..

3.4 Brokenstick

Van Buuren (2014) [22] proposed a piecewise linear model known as the 'brokenstick' model. The author proposes modelling growth via a combination of linear segments with different slopes. This approach is essentially a linear spline model, where the knots are used to represent changepoints in the growth trajectory.

This model is based on a partition with two knots at the endpoints of our dataset, and an additional M internal knots which represent changepoints. Linear segments can then be fitted between each pair of knots, giving a global trajectory with a total of $M+1$ segments. A set of subject specific random effects are used to control each individual child's deviation from each segment of the global trajectory. The brokenstick model is outlined as follows:

$$Y_{ij} = \sum_{m=0}^{M+1} \beta_m \tilde{t}_{im} + \sum_{m=0}^{M+1} \gamma_{im} \tilde{t}_{im} + \epsilon_{ij} \qquad (4)$$

where β_m is a fixed effect population coefficient and γ_{im} is a subject specific random effect for child i. Here, $\tilde{t}_{im}$ represents a B-spline transform of t_{ij}. The sum $\psi_{im} = \beta_m + \gamma_{im}$ can be interpreted as the conditional mean for child i at the mth knot, and the set of ψ_{im} values can be connected by linear segments in order to model the trajectory of child i.

It is important to give consideration to both the number and location of the internal knots when fitting this model. We must choose a sufficient number of knots to capture the changes in growth pattern over time, but we must also avoid overfitting. The author gives some general advice that one should not select more knots than the average number of growth observations per child. The locations of the knots are also important to the overall accuracy of the growth trajectory estimates. The author recommends that the locations are selected to represent specific stages in a child's development, but it should be noted that this is in the context of fitting on the raw scale. This choice may be less crucial when fitting on the Z-scale, because many developmental changes may already be accounted for by the transformation, and evenly spaced knots may provide a more straightforward representation of the growth trajectory.

3.5 Multilevel Spline Model

The brokenstick approach is based on linear splines, but higher degree polynomials can also be used to model growth trajectories. Durban et al. (2005) [23] proposed the use of cubic splines, thus allowing for more flexible global and individual growth trajectories. Additionally, they used penalisation as a means of reducing the impact of overfitting. A consequence of this is that they did not have to worry about knot choices when fitting the model. The penalised splines are represented as a mixed model, thus allowing for fast and computationally efficient fitting using existing mixed model software.

This model is defined as follows:

$$Y_{ij} = f(t_{ij}) + g_i(t_{ij}) + \epsilon_{ij} \qquad (5)$$

where f is a smooth function which represents the population trend and g_i is a smooth function which represents child i's deviation from the population trajectory. The individual trajectory g_i consists of a combination of a linear component and a non-linear component. The choice of penalised splines for both f and g_i is more robust to the user's choice of the number of knots, because of its inbuilt

penalty for overspecification of knots. Note that this method will hereafter be referred to using the shortened form *PS*.

3.6 Functional Principal Components Analysis

As was outlined for the previous model, the longitudinal growth data can be considered to be a form of functional data, and therefore techniques from the field of functional data analysis have been proposed. Xiao et al. (2016) [24] outlined the FACE approach, which was designed specifically for sparse longitudinal data of the form outlined in this paper. This approach assumes that the data take the form

$$Y_{ij} = f(t_{ij}) + h_i(t_{ij}) + \epsilon_{ij} \tag{6}$$

where f is a smooth function which represents the population curve and $h_i()$ represents individual departures from the population curve. Here, $h_i()$ is considered to be a stochastic process with mean 0 and covariance function $C()$. The covariance function $C()$ is estimated via a two-stage approach by first constructing a raw matrix and then applying a bivariate smoother. This covariance function is then used to specify $h_i()$ and thus identify the child-specific deviation from the mean curve.

3.7 Software

As part of the HBDGki initiative, we have developed the *hbgd* software package which allows the user to fit Models (1)-(6) under consistent conditions. This package is available at `https://github.com/HBGDki/hbgd`.

4. Comparison of selected methods

We wish to compare the six proposed models to identify the best performers in terms of the estimation of growth trajectories. In addition, we wish to determine whether modelling on raw or standardised data is more likely to yield accurate trajectories. Therefore, in total we will compare eleven different modelling approaches: SITAR on the raw data only, and each of the other five methods on both raw and standardised data.

We propose a novel leave-one-out validation approach to compare the performances of these models across the various datasets. Leave-one-out validation is commonly used in regression modelling, but has rarely been applied to longitudinal data. A single observation is removed from each child, and then the model is fitted to the remaining dataset. Note that children who were only observed at a single timepoint were removed from the validation process. The models obtained from the reduced dataset were then used to obtain a prediction of the removed observation, and the mean squared error (MSE) was used to quantify the difference between the true and predicted values. Let $\boldsymbol{y} = (y_1, \ldots, y_m)$ be a vector containing the observed values of our held out data from m children, and let $\hat{\boldsymbol{y}} = (\hat{y_1}, \ldots, \hat{y_m})$ be the vector of predictions for those values. Then the mean squared error is given by:

$$\text{MSE} = \frac{1}{m} \sum_{i=1}^{m} (y_i - \hat{y}_i)^2. \tag{7}$$

A lower MSE suggests that a model did a good job of accurately predicting the value of the removed observation. Note that to ensure consistency, the MSEs were always calculated on the Z-scale. In the cases where the models were fitted on the raw data, we transformed the resulting trajectory to the Z-scale in order to calculate the MSE. The Z-transformation is monotonic, and therefore our results are not affected by our decision to calculate the MSEs on the Z-scale rather than the raw scale.

Two different validation approaches were used; a 'random value' approach and a 'last value' approach. For the random value approach, we selected the validation observation at random from the set of all observations for the child. This approach tests the accuracy of the overall model fit, by focusing on how well it can interpolate at unobserved time points. A similar approach was outlined by Grajeda et al. (2016) [18], who randomly removed 20% of observations per child. However, we also include a novel 'last value' approach, which involves removing the final observation for each child, that is the observation at which the child is oldest. This approach tests the predictive ability of the models, with a particular focus on the type of short-term extrapolation for which such models could realistically be used. Prediction is particularly important in the context of the HBGDki project, where we may wish to use a child's observed trajectory to make inference about the effects of an intervention on future growth.

4.1 Illustrative Example - Dataset U

Section 4.2 will summarise the results obtained from each of our 23 datasets. However, for the purposes of illustration, we will present detailed results from Dataset U. This dataset contains height and weight measurements from 700 children in a low income country in Asia. A total of 9741 height observations were taken over the study period, with a median of 16 observations per child.

Figures 1 and 2 display the fitted growth trajectories of a single randomly selected child under each of our eleven proposed modelling strategies. In each case, we used the random holdout approach and fitted the model to the remaining data. For each panel of the plots, the points represent observed HAZ scores for the child, while the line is the fitted trajectory under the selected model. The filled point is the one which was held out for validation. Figure 1 displays the cases where the models were fitted on the raw data, while Figure 2 is based on the Z-scores. Note that the third panel of Figure 2 does not have a fitted line because we cannot fit the SITAR model on the Z-scores.

Figure 2 shows that the Laird and Ware approaches are not flexible enough to model a sensible growth curve based on these data. Each of the other three models appear to do a reasonable job of estimating the growth trajectories of the children. It appears that the brokenstick and face approaches perform best in terms of how well they predicted the holdout value. These models have more flexibility to account for this child's fluctuation in HAZ score between the ages of 200 and 300 days. The penalised spline model provides a reasonable fit, but appears to be slightly too smooth to capture this fluctuation.

4.2 Results

The comparison process outlined in Section 4.1 was repeated on each of the other 22 datasets. Each of these datasets have different features, and we are interested in comparing the methods in terms of MSE to test how robust each method is to different data structures.

The results of the random value validation approach are shown in Table 2, with the lowest MSE for each dataset displayed in bold. Across the 23 datasets, we can see that the brokenstick and FACE approaches provide lower MSE values than any of the other approaches whether they are fit on the Z-score or raw scale. In the majority of cases, the penalised spline method provides the third best MSE, while the lwlinear and lwquad approaches perform more poorly. The SITAR approach is unable to provide a fit on many of the datasets, but performs well in the cases where it managed to fit successfully. We note that for the random holdout approach, fitting on the Z-score scale tends to provide lower MSE values. This is unsurprising, since fitting on the raw scale will typically use up some of our degrees of freedom on the overall curve fit, rather than just focusing on an accurate fit relative to a standard growth curve.

If we focus on the two best approaches, we can see FACE provides better estimates than brokenstick for most datasets, though there are five datasets (Datasets A, B, S, V and Y) where the brokenstick approach is more effective. The main thing which these five datasets have in common is that they have a low number of observations per child, which makes it more difficult for FACE to accurately estimate the necessary principal components. There were also three datasets (Datasets F, N and T) where FACE was unable to provide a successful fit due to the large size of the datasets. FACE is a more computationally complex approach than brokenstick, and is thus more likely to run into such issues.

The relative performance of the modelling approaches was similar under the last value validation approach, as shown in Table 3. However, in this scenario, there was less of a difference in performance between the raw scale and the Z-score scale. This is likely to be because we are extrapolating from our data, which is likely to lead to larger deviations from the truth. If we are fitting on the Z-scale, these larger errors will be inflated when transforming to the true curve.

Again, FACE provides the lowest MSE for most of the datasets, but was unable to provide a fit for the very large datasets. The FACE approach also performs very poorly in Dataset A, which is likely to be due to the small number of observations per child. This dataset has a median of just two observations per child, which means that FACE is often trying to predict a child's trajectory with just a single data point. Overall, however, it appears that the FACE approach represents the most accurate of the modelling approaches discussed here, both in terms of internal and external prediction. However, the brokenstick approach provides a credible alternative, and may prove to be particularly useful in cases where we have larger datasets, or where the number of observations per child is very low. The brokenstick approach is likely to work successfully on a wider range of datasets, and as a result we recommend it as the optimal modelling approach.

We are also interested in comparing the models fitted on the Z-scale to those fitted on the raw data. Figure 3 presents a comparison of the MSEs obtained from modelling raw and Z-scale data using the brokenstick model with random holdout. Each point represents a single dataset, and the x-axis displays the MSE from fitting on the Z-scale while the y-axis displays the MSE from fitting on the raw data. We can see that almost all of the points lie above the line of equality, which means that the MSEs were lower when we fitted the data on the Z-scale. Figure 4 shows similar results for the FACE model. We can see that fitting on the Z-scale leads to more accurate estimation, and as a result, we recommend fitting our models on the Z-scale in all cases.

Table 2: MSE results for random validation. The lowest MSE for each dataset is displayed in bold.

	BS (Z)	BS (raw)	face (Z)	face (raw)	lwlin (Z)	lwlin (raw)	lwquad (Z)	lwquad (raw)	PS(Z)	PS (raw)	sitar (raw)
A	**0.02**		0.03	0.04					0.19	1.75	
B	**0.03**	0.03	0.03	0.03	0.05	0.06	0.05	0.05	0.06	0.60	
C	0.19	0.21	**0.10**	0.11	0.48	1.09	0.37	0.62	0.24	0.24	
D	0.12	0.27	**0.07**	0.09	0.31	1.14	0.28	0.72	0.25	0.27	
E	0.04	0.07	**0.03**	0.04	0.07	0.31	0.06	0.08			0.06
F	**0.16**	1.73			0.53	5.77	0.41	2.92	0.29	0.72	
G	0.12	0.12	**0.07**	0.07	0.33	2.63	0.27	0.44	0.14	0.13	0.21
H	0.06	0.73	**0.03**	0.04	0.21	5.10	0.17	1.57	0.14	0.17	
J	0.02	0.03	**0.02**	0.03	0.08	0.10	0.07	0.08			
K	0.14	0.16	**0.12**	0.13	0.49	1.00	0.45	0.54			0.55
L	0.16	0.50	**0.07**	0.08	0.46	3.76	0.43	1.28	0.41	4.00	
M	0.31	0.36	**0.27**	0.32	0.56	1.51	0.54	0.76	0.54	0.78	0.58
N	**0.09**	0.11			0.27	1.73	0.23	0.54	0.20		0.27
P	0.17	0.68	**0.13**	0.15	0.41	2.83	0.33	1.46	0.31	0.33	
Q	0.05	0.05	0.04	**0.04**	0.11	0.18	0.09	0.10			
R	0.06	0.11	**0.05**	0.07	0.15	0.89	0.13	0.23			0.16
S	0.02	**0.01**	0.03	0.03	0.05	0.06	0.06	0.05	0.07	2.15	
T	**0.61**	1.17			1.01	5.95	0.94	2.14	0.81	0.95	
U	0.05	0.11	0.03	**0.02**	0.17	1.47	0.13	0.52	0.11	0.11	0.16
V	0.02	**0.02**	0.03	0.02	0.08	0.09					
W	0.17	0.18	**0.11**	0.15	0.64	1.14	0.55	0.64			1.31
X	**0.09**	0.14	0.10	0.11	0.30	1.50	0.26	0.49	0.88	2.03	
Y	0.55	0.87	**0.19**	0.36	1.22	2.98	1.21	1.65	1.19	5.01	

Table 3: MSE results for last value validation. The lowest MSE for each dataset is displayed in bold.

	BS (Z)	BS (raw)	face (Z)	face (raw)	lwlin (Z)	lwlin (raw)	lwquad (Z)	lwquad (raw)	PS(Z)	PS (raw)	sitar (raw)
A	0.03	**0.01**	1.25	0.02					0.19	1.15	
B	0.05	0.04	0.04	**0.04**	0.13	0.08	0.11	0.10	0.28	0.37	
C	0.06	**0.04**	0.06	0.06	0.56	0.49	0.27	0.41	0.09	0.21	
D	0.10	0.14	0.10	**0.09**	0.50	0.88	0.47	2.04	0.54	5.15	
E	0.04	0.04	**0.03**	0.04	0.14	0.87	0.10	0.21			0.11
F	**0.10**	0.29			0.40	1.83	0.36	0.58	0.58	0.14	
G	0.03	**0.02**	0.02	0.02	0.22	1.00	0.35	0.80	0.06	1.37	0.22
H	0.03	0.15		**0.01**	0.33	4.57	0.23	0.55	0.20	5.43	
J	0.01	0.02	**0.01**	0.02	0.08	0.11	0.07	0.08			
K	0.13	0.12	0.10	**0.09**	0.45	0.99	0.40	0.53			0.49
L	0.10	0.28	**0.05**	0.06	0.78	22.53	7.73	77.90	0.53	0.56	
M	0.29	**0.24**	0.25	0.25	0.57	1.68	0.55	0.81	0.52	1.36	
N	0.05	**0.04**			0.25	1.21	0.19	0.50	0.14		0.19
P	0.11	0.21	0.07	**0.07**	0.49	1.23	0.36	0.67	0.53	0.39	
Q	0.02	0.02	**0.02**	0.02	0.16	0.46	0.14	0.18			
R	0.06	0.10	0.08	**0.05**	0.26	2.71	0.21	1.29			0.23
S	0.02	**0.02**	0.06	0.05	0.12	0.13	0.11	0.12	0.11	1.98	
T	0.42	**0.24**			0.86	12.72	0.81	3.06	0.80	2.21	
U	0.07	0.07	**0.07**	0.09	0.33	2.67	0.27	1.62	0.27	0.58	0.17
V			**0.02**	0.03	0.11	0.18					
W	0.19	0.18	**0.17**	0.18	0.68	1.35	0.58	0.77			0.71
X			0.26	**0.24**	0.88	4.29	1.04	2.59	0.73	2.03	
Y			0.32	**0.31**	1.06	5.37	1.08	4.12	1.18	11.51	

5. Discussion

This paper provides a thorough comparison of commonly used methods for characterising child growth trajectories. Six different models were tested across 23 datasets and compared in terms of the accuracy of their model fit and their ability to predict future growth patterns. Our results showed that two models, brokenstick and

FACE were consistently the best performing approaches. The FACE model provided slightly better estimation overall, but had some difficulties with larger datasets and also those with a very small number of observations per child. The brokenstick approach was more robust in these circumstances because it is a less computationally complex model. We identified that the Z-score models were superior to raw data models in terms of accurate fitting, and that there was little difference between the approaches in terms of predicting future growth. One of our overall goals is to provide an integrated modelling framework for all of these datasets, and therefore it is important to have consistency in our modelling approaches. As a result, we recommend the use of the brokenstick model with standardised Z-score data. A key advantage of the brokenstick model is that it is easier to fit and provides easily interpretable estimates of child growth trajectories.

The work presented in this paper may motivate future work in the area of growth modelling. For example, one may wish to use indicators extracted from our growth trajectories to investigate the effects of growth faltering on other outcomes, or to consider the factors which may lead to growth faltering in the first place. We have identified sensible techniques for accurately modelling growth trajectories, and have shown that they still perform well on sparse datasets. It may therefore be possible to design more efficient studies with a smaller number of measurements per child, while still retaining the ability to accurately model growth.

The goal of the HBGDki project is to pool together information from a wide range of studies from across the world in order to improve health and wellbeing in children. It is crucial that we identify accurate and reliable models for characterising growth trajectories in order to distinguish between children who have healthy growth and those whose growth is faltering. This allows us to explore factors which predict faltering, and also the effect of poor growth on future health, thus providing a framework for influencing decision making both in the field and at the governmental level.

References

[1] Shields B, Wacogne I, Wright C. Weight faltering and failure to thrive in infancy and early childhood. *BMJ* 2012; **345**.

[2] Martorell R, Khan L, Schroeder D. Reversibility of stunting: epidemiologic findings from children in developing countries. *European Journal of Clinical Nutrition* 1994; **48 (Suppl 1)**:S45–S57.

[3] Checkley W, Epstein L, Gilman R, Black R, Cabrera L, Sterling C. Effects of Cryptosporidium parvum infection in Peruvian children: growth faltering and subsequent catch-up growth. *American Journal of Epidemiology* 1998; **148**:497–506.

[4] Lundeen E, Behrman J, Crookston B, Dearden K, Engle P, Georgiadis A, Penny M, Stein A. Growth faltering and recovery in children aged 18 years in four low- and middle-income countries: Young Lives. *Public Health Nutrition* 2014; **17**:2131–2137.

[5] Norgan N. Long-term physiological and economic consequences of growth retardation in children and adolescents. *The Proceedings of the Nutrition Society* 2000; **59**:245–256.

[6] Crookston B, Penny M, Alder S, Dickerson T, Merrill R, Stanford J, Porucznik C, Dearden K. Children who recover from early stunting and children who are not stunted demonstrate similar levels of cognition. *The Journal of Nutrition* 2010; **140**:1996–2001.

[7] Fink G, Rockers P. Childhood growth, schooling, and cognitive development: further evidence from the Young Lives study. *American Journal of Clinical Nutrition* 2014; **182-188**.

[8] Tanner J. *A History of the Study of Human Growth*. Cambridge University Press, 1981.

[9] Hermanussen M, Bogin B. Auxology - an editorial. *Italian Journal of Pediatrics* 2014; **40**:8.

[10] Cole T, Green P. Smoothing reference centile curves: The LMS method and penalized likelihood. *Statistics in Medicine* 1992; **11**:1305–1319.

[11] Cole T, Donaldson M, Ben-Shlomo Y. Sitar, a useful instrument for growth curve analysis. *International Journal of Epidemiology* 2010; **39**:1558–1566.

[12] Howe L, Tilling K, Matijasevic A, Petherick E, Santos A, Fairley L, Wright J, Santos I, Barros A, Martin R, *et al.*. Linear spline multilevel models for summarising childhood growth trajectories: A guide to their application using examples from five birth cohorts. *Statistical Methods in Medical Research* 2013; .

[13] Fairley L, Petherick E, Howe L, Tilling K, Cameron N, Lawlor D, West J, Wright J. Describing differences in weight and length growth trajectories between white and Pakistani infants in the UK: analysis of the Born in Bradford birth cohort study using multilevel linear spline models. *Archives of Disease in Childhood* 2013; **98**:274–279.

[14] Cole T. Presenting information on growth distance and conditional velocity in one chart: practical issues of chart design. *Statistics in Medicine* 1998; **17**:2697–2707.

[15] Cole T. The development of growth references and growth charts. *Annals of Human Biology* 2012; **39**:382–394.

[16] Haymond M, Kappelgaard A, Czernichow P, Biller B, Takano K, Kiess W. Early recognition of growth abnormalities permitting early intervention. *Acta Paediatrica* 2013; **102**:787–796.

[17] Leroy J, Ruel M, Habicht J, Frongillo E. Using height-for-age differences (HAD) instead of height-for-age z-scores (HAZ) for the meaningful measurement of population-level catch-up in linear growth in children less than 5 years of age. *BMC Pediatrics* 2015; **15**:145.

[18] Grajeda L, Ivanescu A, Saito M, Crainiceanu C, Jaganath D, Gilman R, Crabtree J, Kelleher D, Cabrera L, Cama V, *et al.*. Modelling subject-specific childhood growth using linear mixed-effect models with cubic regression splines. *Emerging Themes in Epidemiology* 2016; **13**.

[19] Checkley W, Epstein L, Gilman R, Cabrera L, Black R. Effects of acute diarrhea on linear growth in Peruvian children. *American Journal of Epidemiology* 2003; **157**:166–175.

[20] Group WMGRS. Length/height-for-age, weight-for-age, weight-for-length, weight-for-height and body mass index-for-age: Methods and development. *Technical Report*, World Health Organisation 2006.

[21] Laird N, Ware J. Random-effects models for longitudinal data. *Biometrics* 1982; **38**:963–974.

[22] van Buuren S. Curve matching: a data-driven technique to improve individual prediction of childhood growth. *Annals of Nutrition and Metabolism* 2014; **65**:225–231.

[23] Durban M, Harezlak J, Wand M, Carroll R. Simple fitting of subject-specific curves for longitudinal data. *Statistics in Medicine* 2005; **24**:1153–1167.

[24] Xiao L, Zipunnikov V, Ruppert D, Crainiceanu C. Fast covariance estimation for high-dimensional functional data. *Statistics and Computing* 2016; **26**:409–421.
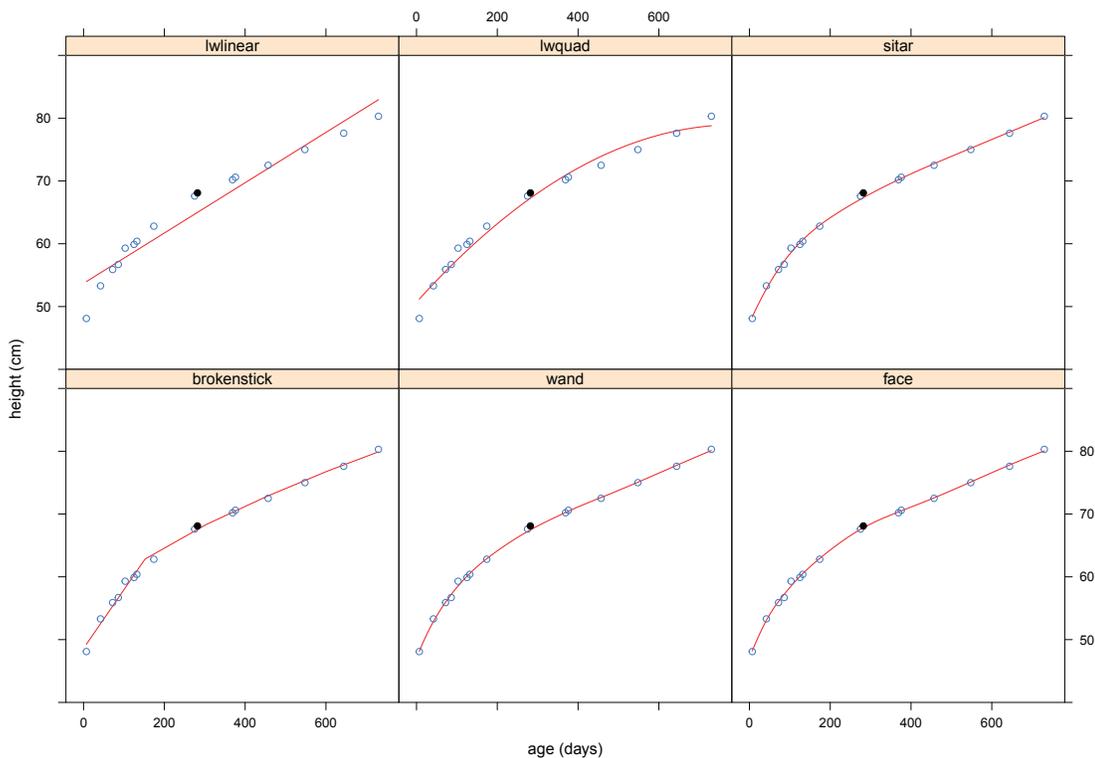
Figure 1: Fitted growth trajectory of a single child based on fitting each of our six models on the raw scale.
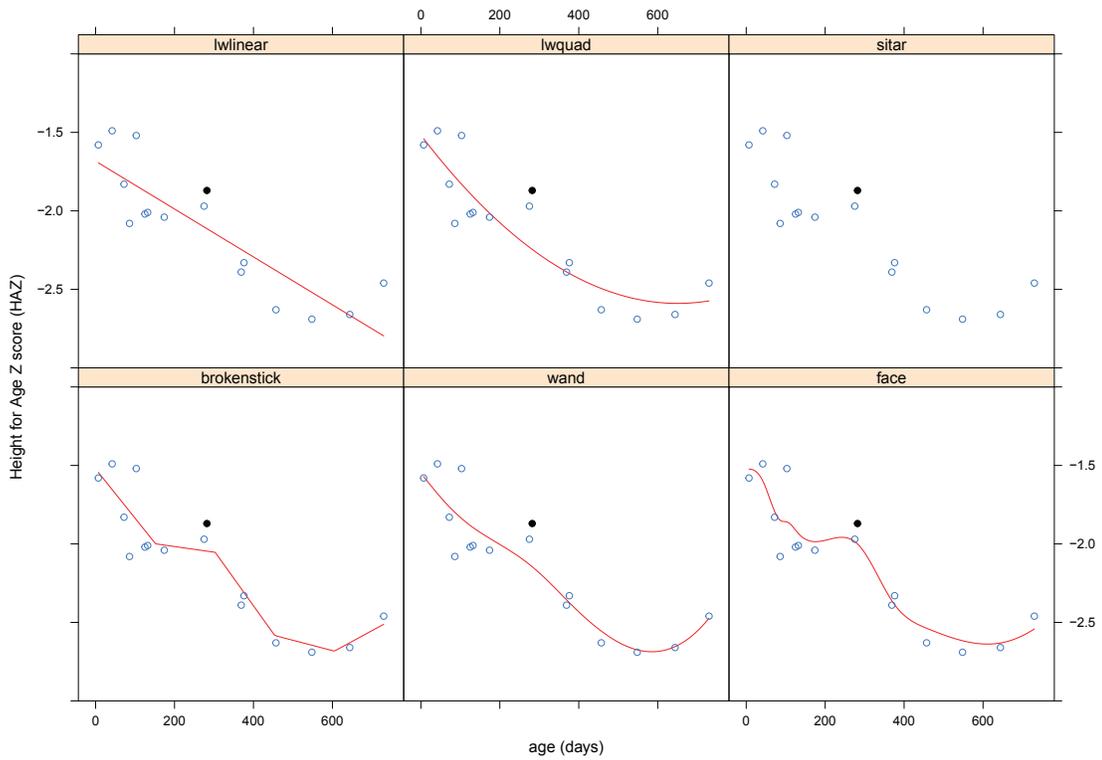
Figure 2: Fitted growth trajectory of a single child based on fitting five of our models on the Z-scale (SITAR was not fitted on this scale).
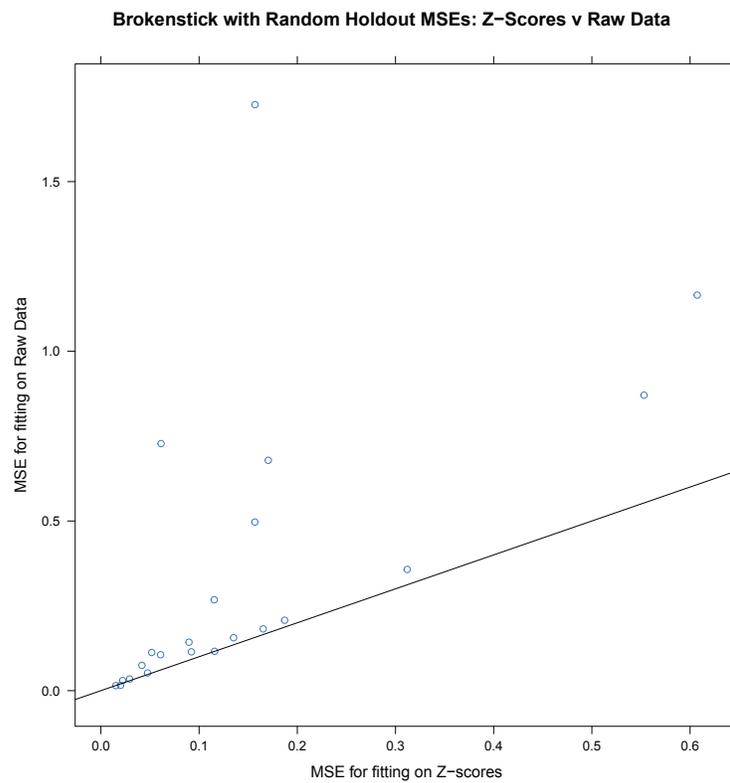
Figure 3: Comparison of results from fitting the brokenstick model on raw data and on Z-transformed data. Each point represents the MSE values obtained from random holdout on one dataset.
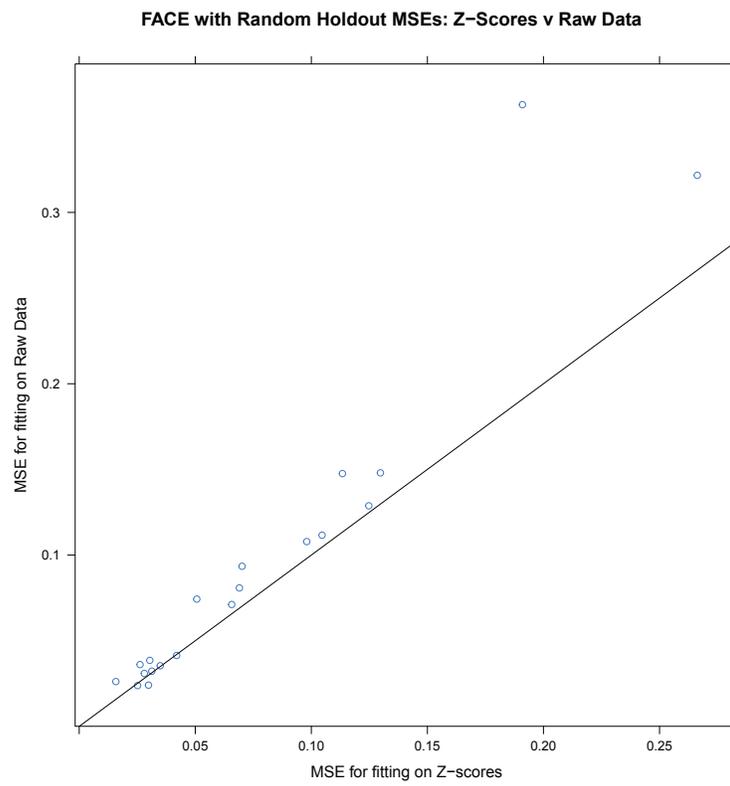
Figure 4: Comparison of results from fitting the FACE model on raw data and on Z-transformed data. Each point represents the MSE values obtained from random holdout on one dataset.