Journal of the Royal Statistical Society
SERIES A Statistics in Society

J. R. Statist. Soc. A (2014)
177, *Part* 2, *pp.* 371–396

On modelling early life weight trajectories

Costanza Pizzi,

Department of Medical Sciences of the University of Turin and Centro di Riferimento per l'Epidemiologia e la Prevenzione Oncologica in Piemonte, Turin, Italy, and London School of Hygiene and Tropical Medicine, UK

Tim J. Cole,

University College London Institute of Child Health, UK

Camila Corvalan,

University of Chile, Santiago, Chile

Isabel dos Santos Silva,

London School of Hygiene and Tropical Medicine, UK

Lorenzo Richiardi,

University of Turin and Centro di Riferimento per l'Epidemiologia e la Prevenzione Oncologica in Piemonte, Turin, Italy

and Bianca L. De Stavola

London School of Hygiene and Tropical Medicine, UK

[Received August 2012. Revised February 2013]

Summary. There is broad recognition that early life growth trajectories can contribute to the study of the onset and development of several health outcomes. We review the random-effects specifications of two models that have been purposely developed to describe anthropometric data and a shape invariant random-effects model recently proposed in the statistical literature. They are compared in terms of their ability to extract salient and biologically meaningful features of growth in infancy and also to represent the data validly. We discuss advantages and limitations in choosing and interpreting each of the models by using longitudinal weight data taken from 0 to 4 years from three contemporary birth cohorts.

Keywords: Growth curve; Jenss–Bayley model; Random effects; Reed model; Superimposition by translation and rotation model; Splines; Weight

1. Introduction

Much interest in modelling growth data comes from research in life course epidemiology (Barker, 1998; Huxley *et al.*, 2000; Kuh and Ben-Shlomo, 2004; Baird *et al.*, 2005), where summary growth parameters are used as explanatory variables for the onset of a later outcome. In such settings analyses consist of two stages: the first focused on modelling the growth data, and the

Address for correspondence: Costanza Pizzi, Cancer Epidemiology Unit, Department of Medical Sciences of the University of Turin, Via Santena 7, 10126 Turin, Italy.
E-mail: costanza.pizzi@lshtm.ac.uk

 0964–1998/14/177371

second aimed at relating parameters from the growth model, e.g. the age at peak weight velocity (APWV), to distal health outcomes, e.g. cardio-vascular disease later in life. This field of research has been enriched by several new European (CHICOS, 2010) and worldwide (Richter *et al.*, 2011; Brown *et al.*, 2007) birth cohorts established to study early life influences on the onset and development of a wide range of later diseases. As a consequence, a large body of longitudinal growth data, collected in different populations and periods and with different approaches and levels of detail, is available. There is therefore an increasing need for understanding and applying modelling strategies that properly exploit these data, also taking into account both the pattern and the quality of the growth measurements.

The traditional approach of analysing anthropometric data consists of finding a parametric curve which is suitable for the variable and the age range of interest, while fitting it separately to each child. For example the *Jenss–Bayley* (JB) model was suggested (Jenss and Bayley, 1937) to describe weight or height growth from birth to age 6 years. It is a fully parametric model that includes a linear and a non-linear (exponential) part. Another model that is widely used with early life height and weight is the four-, or five-, parameter *Reed* model (Berkey and Reed, 1987), where different parameterizations can be used to model postnatal growth deceleration. Typically these and other anthropometric models are either specific to a particular growth measurement or age range and therefore are not widely applicable, and, most importantly, their parameters do not have an intrinsic biological interpretation. Moreover their fitting of separate individual curves is inefficient. An alternative approach that partly overcomes the latter problem consists of fitting whatever model is selected not separately to each child but simultaneously, by using random-effects (i.e. mixed) models (Verbeke and Molenberghs, 2000). Within this framework child-specific parameters are assumed to be drawn from some distribution. Applications include a random-effects specification of the JB model used to study the height trajectories of children with and without Turner syndrome (van Dommelen *et al.*, 2005), and two specifications of the Reed model used to analyse growth in weight during the first years of life in an Ethiopian and a Finnish birth cohort (Asefa *et al.*, 1996; Tzoulaki *et al.*, 2010). General specifications of random-effects models have also been used to analyse growth data (Verbeke and Molenberghs, 2000). Linear and higher degree polynomial functions of time are specified within this framework (dos Santos Silva *et al.*, 2002; Yang *et al.*, 2011), with random effects generally included only for some of the parameters because of identification constraints (Goldstein, 2011). However, these models can produce only a limited range of possible shapes, which are not usually sufficient to describe infant growth. Furthermore growth patterns may not all cluster around a single population average polynomial curve, therefore leading to poor fit to the data.

For this reason the implementation of spline-based models within the mixed effects framework has raised considerable interest because of their flexibility in fitting different anthropometric variables over different age ranges and the possibility of achieving a satisfactory balance between goodness of fit and smoothness of the growth curve by choosing the number of knots (Ruppert *et al.*, 2003). Among these models, the shape invariant random-effects model that was proposed by Beath (2007) is particularly relevant. It was originally used to model infant growth in weight (Beath, 2007), but it has been used by Cole to analyse height growth during puberty (Cole *et al.*, 2010), and more recently by Hui *et al.* (2010) and Johnson *et al.* (2011) to study factors influencing infant growth. In this model a common spline function is modified by shifting the two axes and scaling the x-axis to fit the individual data. The scaling and shifting identify how each child departs from the common spline function and are captured by parameters which have a direct biological interpretation (Beath, 2007; Cole *et al.*, 2010), unlike any of the methods that were mentioned above. The application of the shape invariant random-effects model, however, requires fairly advanced computing skills. Its interpretation also requires careful understanding of the parameterization.

The aim of this paper is to compare the two models that were purposely developed for anthropometric data, the JB and Reed models, both specified within a random-effects framework, with the shape invariant random-effects model in terms of their ability to extract salient and biologically meaningful features of growth in infancy and also in their ability to represent the data validly. Comparisons will be carried out by using longitudinal weight data from three recent birth cohort studies. The information that is available from these cohorts differs in term of age range and number and timing of follow-up observations (regular or irregular), and therefore illustrates a variety of settings that are likely to be encountered in practice.

The programs that were used to analyse the data can be obtained from

`http://www.blackwellpublishing.com/rss`

2. Data

We have analysed data from three contemporary birth cohorts: the Southern European Web-based 'Nascita e infanzia gli effetti dell'ambiente' (NINFEA) and 'Geracão XXI' birth cohort studies, and the Chilean 'Growth and obesity cohort study' (GOCS).

2.1. Nascita e infanzia gli effetti dell' ambiente

The NINFEA study is an Italian Web-based cohort study which started in 2005 and aimed to recruit pregnant women via the Internet and to follow up their children. The study is still on going and targets pregnant women, who have access to the Internet and sufficient knowledge of the Italian language to understand the questionnaires. They are recruited via a wide range of advertising campaigns (for more details see Richiardi *et al.* (2007)). Registration is carried out at the study Web site (`www.progettoninfea.it`) where participants complete the first questionnaire Q1. The cohort is then actively followed up via other on-line questionnaires administered at around 6 (Q2), 18 (Q3) and 48 (Q4) months of age of the child.

Revisions of the questionnaires were undertaken after enrolment of the first 1500 participants; thus the available data vary with year of recruitment. In particular, women who enrolled until November 2008 were asked to report the child's weight at birth, 3 and 6 months at the time of the second questionnaire, whereas at the third they were asked to report the weight of the baby at 12 and 18 months. When the second and third questionnaires were updated, a new question was introduced regarding the child's weight at the actual time of completion of these questionnaires. In contrast the fourth questionnaire was set up from the outset to include questions on anthropometric measures both at 4 years of age and at the time of completion of the questionnaire.

Because of these variations in questionnaire design the data that were used for these analyses include all singleton children who, at the time of the data download (November 2011), were eligible for completion of the fourth questionnaire. These are the children for whom growth measurements are available at fixed time points only (0, 3, 6, 12, 18 and 48 months). This sample includes 845 children.

2.2. Geracão XXI study

The Gercão XXI study is the first Portuguese birth cohort study, which was established in 2005 in the Porto region. All live children born of women resident in one of the six regional districts, admitted to one of the five hospitals in Porto for delivery with a gestational age at birth that was greater than 24 weeks, were eligible to participate. The recruitment period lasted from the end of April 2005 until August 2006. Women were enrolled during an appointment a few days before their due date and, for the majority, baseline questionnaires were completed between 24 and 72 h after delivery. In total the baseline data consist of 8311 singleton children. Children were actively followed up through interviewer-administered questionnaires planned at 3, 6, 12 or 15, and 24 months. Because of logistic and financial limitations it was not possible to interview every participant at each follow-up visit. Therefore a restricted time period was allocated for each follow-up occasion. At the 2-years follow-up visit growth data from the child's health records card, with measures obtained prospectively by health professionals, were gathered from the parents for entry into the database.

The data that were used for this analysis consist of the information from the baseline and the 2-years follow-up, available for 783 singleton babies. These include anthropometric measures reported in the child's health record and referring to measures taken at about 1, 2, 4, 6, 9, 12, 15, 18 and 24 months of age, together with the actual dates of measurement. The values at 24 months were obtained directly by the interviewers. Up to six additional measurements and dates reported in the health records were also entered into the database. There are therefore up to 16 weight measurements per child, taken from birth to (about) 24 months.

2.3. Growth and obesity Chilean cohort study

The GOCS is an on-going Chilean cohort aiming to study the association of early growth with children's maturation, adiposity and associated metabolic complications. The study was initiated in 2006 when all children aged 2.6–4 years attending public nursery schools in six counties of Santiago were invited to participate if they

(a) were singleton births with a gestational age at birth between 37 and 42 weeks,
(b) had a birth weight between 2500 and 4500 g and
(c) had no physical or psychological conditions that could affect growth (six children were excluded because of these conditions).

Among the 1498 children who were eligible to participate, 1195 accepted (80%). For almost all subjects, weight and height measurements from birth up to 36 months of life were retrospectively gathered from health records, and they were prospectively measured every year after recruitment by a dietitian who visited the nursery school.

For these analyses we used growth data up to only around age 4 years for direct comparison with the NINFEA cohort. This leads to including a maximum of 11 measurements per child. After exclusion of subjects with missing growth data, the sample that was used in these analyses consists of 1149 children.

3. Methods

In this paper we compare two classes of anthropometric random-effects models: the JB model and the four-parameter Reed model, and the shape invariant random-effect model. Below we define these models and the specific issues arising from fitting them to the data.

3.1. Jenss–Bayley model

3.1.1. Original specification

The JB model (Jenss and Bayley, 1937), has been widely applied (see van Dommelen *et al.* (2005), Deming and Washburn (1963), Manwani and Agarwal (1973) and Berkey (1982)) to describe childhood growth during the first 6 years of life in term of both height and weight. According to this model, the observed growth variable $y_i(t)$ of child i at age t, for $i=1,\ldots,n$ and $t=1,\ldots,T_i$, can be expressed as

$$y_i(t)=c_i+d_i t-\exp(a_i+b_i t)+\varepsilon_{it} \tag{1}$$

where a_i, b_i, c_i and d_i are unknown parameters to be estimated separately for each child and ε_{it} is the error term at age t specific to child i, that is assumed to be normally distributed with mean 0 and variance σ_i^2. Equation (1) defines $y_i(t)$ as a negatively accelerated exponential curve in t, whose asymptote is a positive straight line. Hence the model accounts for the rapid decelerating growth rate that is usually observed during infancy via the exponential component $\exp(a_i+b_i t)$, whereas c_i and d_i represent the intercept and the slope of the asymptote respectively. From equation (1) it follows that the predicted size at birth for child i is $c_i-\exp(a_i)$.

3.1.2. Random-effects specification

Instead of modelling each child separately, we can add some distributional assumptions for the child-specific parameters, and to their relationship with the residual errors ε_{it}, to define an overall model for all children. Using the same notation as in equation (1) we now specify the child-specific parameters as

$$\begin{aligned} a_i &= a_0+a_{1i}, \\ b_i &= b_0+b_{1i}, \\ c_i &= c_0+c_{1i}, \\ d_i &= d_0+d_{1i} \end{aligned}$$

where a_0 represents the fixed effect and a_{1i} the child-specific random effect, with similar definitions for the components of b_i, c_i and d_i. We also assume that a_{1i}, b_{1i}, c_{1i} and d_{1i} are drawn from a multivariate normal distribution with mean $\mathbf{0}$ and covariance matrix Φ and that the errors ε_{it} are independent normally distributed random variables with mean 0 and constant variance σ^2, which are independent of the child-specific random effects. The curve that is implied by the model when all random effects are set to 0 (the 'population level curve'; Pinheiro and Bates (2000)) is therefore $c_0+d_0 t-\exp(a_0+b_0 t)$. This model does not allow for an inflection point (a maximum or minimum in the weight velocity curve), and therefore no APWV, because its second derivative with respect to t is an exponential function in t.

3.2. Reed model

3.2.1. Original specification

The Reed model was suggested by Berkey and Reed (1987), as an extension to an earlier model that was suggested by Count (1943). It has two versions, with four or five parameters. For comparison with the JB model we focus on the four-parameter version, which has been shown not to be inferior in term of goodness of fit to the five-parameter model (Simondon *et al.*, 1992). The four-parameter specification of the Reed model is

$$y_i(t)=a_i'+b_i't+c_i'\ln(t)+\frac{d_i'}{t}+\varepsilon_{it}' \tag{2}$$

where, again, $y_i(t)$ is the ith child's growth variable at age t, and a_i', b_i', c_i' and d_i' are the child-specific parameters and ε_{it}' is the child- and age-specific error term, which is assumed to be normal with mean 0 and variance $\sigma_i'^2$. This specification, unlike that for the JB model, is linear in its parameters and can accommodate one inflection point. From equation (2) it follows that c'_i and d'_i are deceleration terms moderating the original increase that is captured by b'_i.

3.2.2. Random-effects specification

As for the JB model, we can specify the Reed model by using a random-effects approach. Again we define the child-specific parameters as

$$\begin{aligned} a'_i &= a'_0 + a'_{1i}, \\ b'_i &= b'_0 + b'_{1i}, \\ c'_i &= c'_0 + c'_{1i}, \\ d'_i &= d'_0 + d'_{1i} \end{aligned}$$

with a'_0 representing the fixed effect and a'_{1i} the child-specific random effects, and similarly for the other parameters. The parameters a'_{1i}, b'_{1i}, c'_{1i} and d'_{1i} are assumed to be drawn from a multivariate normal distribution with mean $\mathbf{0}$ and covariance matrix Φ'. As before the error terms ε'_{it} are assumed to be independent normally distributed random variables with mean 0 and constant variance σ'^2, and to be independent of the random effects. This model can have an inflection point and therefore the APWV can be derived by differentiating the weight velocity curve (the first derivative of the fitted curve) and setting it to 0.

3.3. Shape invariant random-effects model

The shape invariant random-effects model was introduced by Beath (2007) to describe infant weight growth and by Cole *et al.* (2010) to analyse pubertal growth in height. Cole used a slightly modified parameterization, which is the parameterization that is adopted here, and named the model 'Super-imposition by translation and rotation' (SITAR). It is a model where a common spline function for all subjects is modified by shifting the two axes and scaling the x-axis to adapt it to the individual trajectories. This model is not specific to an age range or to an anthropometric dimension. The use of a spline function allows this model to deal naturally with non-linearity and therefore to identify inflection points. In this application a natural cubic spline with a B-spline basis matrix is used to fit the data by using a non-linear random-effects model, where the coefficients of the spline function are treated as fixed effects whereas the parameters of the shape invariant model are treated as random. Formally, let $y_i(t)$ be again the growth dimension of the ith child at age t; then the SITAR model is specified as

$$y_i(t) = \alpha_i + h\left\{\frac{t-\beta_i}{\exp(-\gamma_i)}\right\} + \eta_{it} \tag{3}$$

where $h(z)$ is the natural cubic spline curve of the growth variable regressed on z (the transformed age) and $\alpha_i = \alpha_0 + \alpha_{1i}$ is a subject-specific coefficient with α_0 representing the fixed effect and α_{1i} the random effect, with similar definitions for the other two parameters. The three random effects α_{1i}, β_{1i} and γ_{1i} are assumed to be drawn from a multivariate normal distribution with mean $\mathbf{0}$ and covariance matrix Λ. As before the error terms η_{it} are assumed to be independent normally distributed random variables with mean 0 and constant variance τ^2, and to be independent of the random effects. Weight velocity curves can easily be derived by differentiating the spline curve, and, from these, the APWV can be derived.

The specific random effects α_{1i}, β_{1i} and γ_{1i} for subject i correspond to the shift parameter for the y-axis (measure), and the shift and the scale parameter for the x-axis (age) respectively. Cole *et al.* (2010) referred to these three parameters as *size*, *tempo* and *velocity*, where the effects of the first two lead to a translation (in measure and age), whereas that of γ to a rotation. Size is expressed in the units of the y-variable and tempo in the units of the t-variable, whereas velocity is a multiplier and therefore is scale free. Therefore, from a biological perspective, when analysing weight in infancy α_{1i} captures differences in size (which are greater for heavier children), β_{1i} the timing of growth (which is negative for babies with an earlier APWV) and γ_{1i} the growth rate (which is positive for children with steeper growth). When these subject-specific random effects are set equal to 0, again the 'population level curve' is obtained.

3.4. Growth models for weight in early life

We fitted the random-effects specifications of the JB and Reed models and of the SITAR model to the weight data from the three cohorts that were described above, initially separately by sex. We used all the available data (for ages 0–4 years in the NINFEA study and GOCS, and for ages 0–2 years in the Geracão XXI study) as well as the NINFEA and GOCS data restricted to ages 0–2 years for comparison with Geracão XXI. In each set of analyses all subjects with at least one observation were included under the assumption that missing visits occurred at random (Rubin, 1976).

The JB and Reed models were both originally conceived for an untransformed growth measure. Despite this, natural logarithmic transformations of the observed weight measures were considered when fitting all models and results compared with those obtained when weight was not transformed.

A transformation of the timescale was needed for the Reed model because our data include measures at birth, i.e. at time 0. One option, which was suggested by Berkey and Reed themselves, is to replace age since birth (in months) t, with t^* where $t^* = (t+9)/9$, so that $t^* = 0$ at around conception and $t^* = 1$ at birth. With this transformation the size at birth for child i that is predicted by the Reed model is $a_i' + b_i' + d_i'$. The same transformation was also considered when fitting the SITAR model when investigating its best specification for the data.

The SITAR model was fitted by using a natural cubic spline with B-spline basis matrix, placing the internal knots of the spline $h(t)$ at the quantiles of the age distribution, as in Cole *et al.* (2010), according to the number of degrees of freedom specified. Degrees of freedom were chosen according to the richness of available weight measurements over the age timescale. Because of identification issues, alternative constraints on the values of the fixed effects β_0 and γ_0 were considered. In particular we considered either fixing both β_0 and γ_0, or just β_0 or just γ_0 to be equal to 0. To aid interpretation, models were also refitted with the males and females combined.

Estimation of all models was carried out by maximum likelihood estimation implemented in the `nlme` function in R (Lindstrom and Bates, 1990). The cubic spline function with B-spline basis matrix was fitted by using the `ns` function also in R.

3.5. Comparison of alternative models

Specifications of the SITAR model with different weight and age scales first, and then with alternative constraints on the fixed and random effects, were fitted on the data from all three cohorts, and then compared in terms of the Akaike information criterion AIC and Bayesian information criterion BIC. Plots of predicted population level curves, and of individual growth and growth velocity curves, for a selection of children were also examined to assess their closeness to the observed data. Specifications of the random-effect JB model obtained with or without

Table 1. Baseline characteristics and weight and weight changes at selected ages, by cohort

	Results for Geracão XXI (n = 783), age range 0–2 years					*Results for GOCS (n = 1149), age range 0–4 years*					*Results for NINFEA (n = 845), age range 0–4 years*				
	N	*Range*	*Median*	*Mean*	*SD*	*N*	*Range*	*Median*	*Mean*	*SD*	*N*	*Range*	*Median*	*Mean*	*SD*
Number of measures per child	783	1–16	10	9.4	1.9	1149	2–11	8	7.6	1.9	845	1–6	5	5.0	1.2
Gestational age (weeks)	748	28–43	39.5	39.2	1.6	959	37–42	39.5	39.6	1.3	837	26–43	39.9	39.7	1.7
Preterm births (<37 weeks)	61	8.2%	—	—	—	0	0.0%	—	—	—	40	4.8%	—	—	—
Female	373	49.3%	—	—	—	579	50.4%	—	—	—	399	47.3%	—	—	—
Birth weight (kg)	756	0.6–4.7	3.2	3.2	0.5	1148	2.5–4.8	3.4	3.4	0.4	793	0.8–4.8	3.2	3.2	0.5
Weight at 2 or 3 months† (kg)	573	2.5–8.1	5.0	5.0	0.7	800	3.4–7.9	5.4	5.4	0.6	662	2.3–8.7	5.7	5.7	0.8
Weight at 6 months (kg)	522	4.2–12.8	7.6	7.6	0.9	747	5.5–11.2	7.8	7.9	0.9	664	4.6–11.5	7.6	7.6	1.0
Growth rate 0–2 or 3 months† (kg month^{-1})	553	0.3–1.8	0.9	0.9	0.2	799	0.1–2.0	1.0	1.0	0.3	658	0.1–1.8	0.8	0.8	0.2
Growth rate 0–6 months (kg month^{-1})	506	0.4–1.4	0.7	0.7	0.1	746	0.3–1.3	0.7	0.7	0.1	661	0.3–1.4	0.7	0.7	0.1
Weight at 18 months‡ (kg)	291	7.9–15.7	11.3	11.2	1.3	729	7.5–16.2	11.1	11.3	1.3	752	8.0–16.0	11.2	11.4	1.3
Weight at 48 months§ (kg)	—	—	—	—	—	136	12.7–31.3	17.2	17.5	2.6	620	11.0–30.0	16.0	16.5	2.3

†Weight data around 2 months are available for the GOCS and Geracão XXI studies, whereas weight data at 3 months are available for the NINFEA study.
‡For the GOCS and Geracão XXI studies, children with weight measures collected between 17.5 and 18.5 months are included.
§For the GOCS, children with weight measures collected between 47 and 49 months are included.

log-transformation of the weight scale were compared by using the same criteria and likewise for the random-effects Reed model. When the weight scale was transformed, the adjusted deviance was calculated (Box and Cox, 1964), and corrected AIC and BIC derived.

For each data set the goodness of fit of the best specifications of each model was then compared again using the same criteria. Moreover, child-specific APWVs were estimated, when possible, and their averages compared with published data (Tzoulaki *et al.*, 2010; Johnson *et al.*, 2011).

For every model and every data set, the distribution of the estimated error terms was examined graphically by using normal quantile plots.

Only a selection of the comparisons and of the plots described above are reported below.

4. Results

4.1. Data description

Table 1 summarizes the baseline characteristics of the children who were included in the analyses, by cohort. It also includes descriptors of weight and weight change at selected ages. Geracão XXI is the richest cohort in terms of weight measurements gathered, with a median of 10 observations per child within the first 2 years of life, compared with medians of 8 in the GOCS and of 5 in the NINFEA study where growth measures were reported at exactly 0, 3, 6, 12, 18 and 48 months. Mean or median birth weight, as well as weight at 2 and 6 months, are slightly higher in the GOCS compared with the other two cohorts possibly because of its exclusion of preterm births (gestational age less than 37 weeks). This is also reflected by the average weight change in the first months of life, which is again slightly higher in the GOCS (1.0 kg month^{-1}) than in the Geracão XXI (0.9 kg month^{-1}) and NINFEA studies (0.8 kg month^{-1}, in the first 3 months). The lower rates in the European cohorts are probably because preterm children have not yet completed their catch-up period by 2 or 3 months, as opposed to the term children of the GOCS. Indeed the rates in weight change are identical across the three cohorts when evaluated from birth to 6 months. Mean weight and corresponding standard deviation (SD) at 18 months are similar across the three cohorts, whereas GOCS children at 4 years are on average slightly heavier than those in the NINFEA study (we have no Geracão XXI measures at 4 years). All these comparisons are, however, based on different subsets of children at each time point because not all children in each cohort have measures at all times considered.

Plots of weight trajectories for a random selection of children (dark lines for males and lighter lines for females), superimposed over the scatter of data for their full cohort, are shown in Fig. 1 (where the cohort differences in terms of frequency and timing of the observations are also evidenced). Each child selected from each cohort represents one of six strata, defined by gender and birth weight category (low, middle or high, defined with cut-off points at 2.5 and 3.8 kg). These plots clearly show how weight growth is faster during the first 3–6 months of life and then starts to decelerate, especially after the first year of birth. As expected males are generally slightly heavier than females in each birth weight category. Fig. 1 highlights specific growth profiles:

(a) children with a low weight at birth who remain small (relatively to the other same sex children);
(b) children with a low weight at birth who experience a high postnatal rate of growth (e.g. the low birth weight male in the Geracão XXI study);
(c) those with a high birth weight who remain constantly heavier compared with the others (e.g. the high birth weight male in the GOCS);

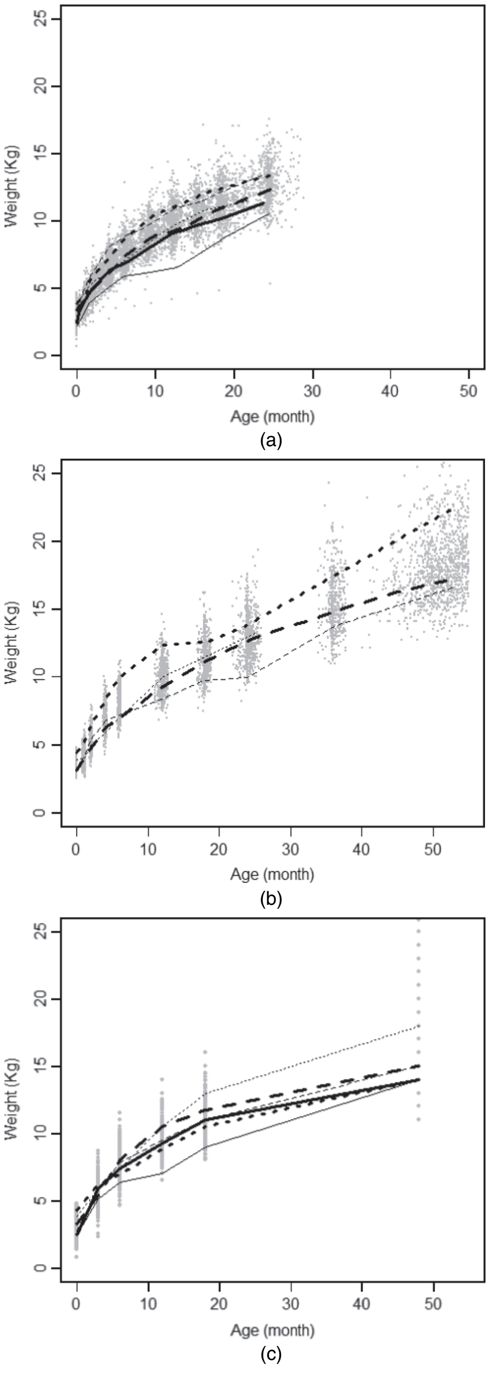


Fig. 1. Observed weight growth curves of a stratified random selection of children by cohort, superimposed on the observed data (by gender and birth weight category) (———, — —, ------, males; ———, — —, - - - -, females; ———, low birth weight; — —, medium birth weight; ------, high birth weight; ●, observed measurements); (a) Geracão XXI; (b) GOCS; (c) NINFEA

(d) those with an average weight at each observational time point (e.g the medium birth weight male in the Geracão XXI study);
(e) finally children with a high birth weight who experience a greater deceleration after the first months of life and thus return to an average weight (e.g. the high birth weight male in the NINFEA study).

4.2. *Growth modelling*

4.2.1. *Jenss–Bayley and Reed models*

Log-transforming the weight scale did not generally improve the fit of the two anthropometric models according to the fit criteria. Therefore only results obtained on the original scale are reported in Table 2, separately by age, gender and cohort. The JB model failed when fitted on the NINFEA 0–2-years data; this is probably due to the small number of measurement times that were available (see Fig. 1), the restricted age range analysed (since the JB model was suggested to describe growth between birth and 6 years) and the measurement error that may be affecting the NINFEA recalled measurements more than those from the other cohorts.

The estimates of the JB fixed effects were quite stable across cohorts, age and gender, with $\hat{a}_0$ and $\hat{c}_0$ slightly lower in females than in males. These estimated parameters were also slightly lower in both GOCS males and females than in the other cohorts, as were the estimates of b_0. The corresponding SDs of the random effects were similar across cohorts and gender, and lower for the 0–4-years data compared with the 0–2-years data.

In agreement with the JB model, the Reed fixed effect estimates differed in the GOCS, with all $\hat{a}_0'$, $\hat{b}_0'$, $\hat{c}_0'$ and $\hat{d}_0'$ considerably larger in absolute terms compared with the NINFEA and Geracão XXI results. The random effects SDs for this model were extremely large because of their high correlations (see below). However, they were lower when estimated on the 0–4-years data, as were the random-effects correlations.

4.2.2. *Super-imposition by translation and rotation models*

As described in Section 3 we compared alternative specifications of the SITAR model, by allowing different scales for the dependent variable, weight, and the age timescale, and different numbers of knots for the spline function, and by including constraints on the values of the fixed effects.

Models with log-transformed weight always performed better in terms of AIC and BIC than those with untransformed weight; therefore results from only the former models are reported. Instead models with log-transformed age performed better only in some settings and hence both sets of results (models with age and log-transformed age) are reported. Four internal knots were used when analysing the Geracão XXI study and the GOCS and three when analysing the NINFEA cohort, because of its fewer time occasions. Results were not affected when the number of internal knots for Geracão XXI and the GOCS was increased to 5 or 6. Alternative constraints on the fixed effects were examined to address identification issues. Among the various combinations fixing either β_0 and γ_0 (model 1), or just β_0 (model 2), to be equal to 0 led to the best-fitting models.

The results are shown in Table 3. According to AIC and BIC, model 2 performed slightly better than model 1 in most of the combinations; moreover the estimates of the APWV were more stable when obtained from model 2. As regards the choice of the age scale, the models fitted best on the untransformed age scale in most cases.

The estimated SDs of the three random effects α_{1i}, β_{1i} and γ_{1i} did not vary substantially across genders, nor across cohorts and age ranges with the exception of the estimated SD of

Table 2. Estimated fixed effects and SDs of the random effects of the JB and Reed models fitted on the original weight scale and stratified by cohort, gender and age range

Study	*Gender*	*Results for JB model*								*Results for Reed model*							
		Fixed effects				*Random effects (SD)*				*Fixed effects*				*Random effects (SD)*			
		a_0	b_0	c_0	d_0	a_{1i}	b_{1i}	c_{1i}	d_{1i}	a'_0	b'_0	c'_0	d'_0	a'_{1i}	b'_{1i}	c'_{1i}	d'_{1i}
0–2 years																	
Geracão XXI	Male	1.70	−0.19	8.88	0.17	0.37	0.06	2.08	0.07	12.52	0.85	−0.08†	−10.25	9.78	3.65	14.81	13.08
	Female	1.59	−0.17	8.19	0.18	0.42	0.07	1.84	0.08	9.72	0.79	1.33†	−7.49	9.48	3.84	14.74	12.89
GOCS	Male	1.52	−0.27	8.11	0.20	0.36	0.11	1.81	0.07	19.80	3.29	−10.47	−19.74	11.19	4.10	16.57	14.92
	Female	1.45	−0.22	7.70	0.19	0.36	0.08	1.61	0.07	16.87	3.20	−8.96	−16.77	10.66	3.86	15.58	14.12
NINFEA	Male	—	—	—	—	—	—	—	—	10.95	0.66†	1.48†	−8.28	13.85	5.79	22.28	19.22
	Female	—	—	—	—	—	—	—	—	6.76	−0.41†	5.87	−3.22	11.80	5.55	20.57	16.90
0–4 years																	
GOCS	Male	1.55	−0.25	8.18	0.20	0.26	0.07	1.35	0.05	18.38	2.69	−8.15	−17.71	6.65	1.52	7.54	7.86
	Female	1.40	−0.24	7.49	0.21	0.29	0.08	1.27	0.05	15.40	2.63	−6.66	−14.71	6.63	1.55	7.56	7.88
NINFEA	Male	1.76	−0.17	9.33	0.16	0.32	0.06	1.94	0.07	11.39	0.91	0.61†	−8.97	8.34	2.02	9.98	10.04
	Female	1.50	−0.18	7.98	0.17	0.33	0.05	1.58	0.06	9.24	0.96	1.08†	−7.08	6.21	1.63	7.66	7.46

†Not statistically significant at the 5% level.

Table 3. Results from alternative specifications of the SITAR model stratified by cohort, gender and age range†

Study	*Model*	*Results for log(kg) and month*						*Results for log(kg) and log(t‡)*					
		Random effects (SD)			*APWV (months)*	*AIC*	*BIC*	*Random effects (SD)*			*APWV (months)*	*AIC*	*BIC*
		α	β	γ				α	β	γ			
Males, 0–2 years													
Geracão XXI	1	0.18	0.72	0.32	1.53	*5332*	*5407*	0.23	0.09	0.29	1.16	5503	5577
	2	0.18	0.71	0.32	1.65	5333	5414	0.22	0.09	0.28	1.50	5479	5560
GOCS	1	0.17	0.63	0.35	0.78	6120	6193	0.19	0.08	0.22	0.73	6277	6350
	2	0.16	0.60	0.30	1.29	6102	6182	0.12	0.06	0.15	1.40	*6054*	*6134*
NINFEA	1	0.14	0.70	0.26	2.58	4117	4178	0.22	0.10	0.31	1.51	4129	4190
	2	0.15	0.73	0.30	1.98	*4044*	*4110*	0.16	0.08	0.21	1.97	4203	4270
Males, 0–4 years													
GOCS	1	0.18	0.70	0.34	1.98	9306	9383	0.13	0.06	0.10	1.01	9051	9127
	2	0.15	0.64	0.26	1.10	*9031*	*9114*	0.14	0.06	0.10	0.95	9039	9122
NINFEA	1	0.15	0.72	0.31	2.64	5398	5460	0.22	0.11	0.25	1.47	5394	5456
	2	0.17	0.77	0.36	1.83	*5304*	*5372*	0.12	0.07	0.11	1.75	5458	5526
Females, 0–2 years													
Geracão XXI	1	0.18	0.72	0.35	1.62	*4716*	*4790*	0.24	0.10	0.29	1.29	4833	4907
	2	0.18	0.71	0.34	1.90	4717	4797	0.23	0.10	0.31	1.62	4794	4874
GOCS	1	0.19	0.74	0.37	0.72	6029	6103	0.24	0.10	0.25	0.55	6065	6139
	2	0.17	0.70	0.32	1.29	*5813*	*5893*	0.13	0.07	0.15	1.28	5936	6017
NINFEA	1	0.14	0.74	0.24	2.61	3372	3432	0.17	0.08	0.25	1.66	*3322*	*3382*
	2	0.15	0.77	0.28	2.04	3360	3425	0.15	0.07	0.23	1.96	3375	3441
Females, 0–4 years													
GOCS	1	0.18	0.74	0.30	1.87	9252	9329	0.12	0.06	0.09	0.93	8963	*9040*
	2	0.13	0.60	0.20	0.94	9075	9159	0.13	0.06	0.09	0.89	*8957*	*9040*
NINFEA	1	0.17	0.80	0.31	2.64	4627	4689	0.22	0.11	0.22	1.41	4660	4722
	2	0.19	0.89	0.34	2.11	*4536*	*4603*	0.23	0.11	0.21	0.98	4684	4751

†Model 1, $\beta_0 = 0$ and $\gamma_0 = 0$; model 2, $\beta_0 = 0$. Within each stratum defined by study, gender and age range the lowest AIC and BIC are reported in italics.
‡$t = (\text{month} + 9)/9$.

β_{1i}, that was generally lower in the GOCS than in the other cohorts, probably because of its inclusion criteria (Table 3).

4.3. Models comparison

4.3.1. Goodness of fit

Results from the specification of the SITAR model which best fitted each data set were compared with those obtained from the JB and Reed models in term of AIC and BIC (Table 4). The two information criteria were always in agreement and show that the Reed model fits the data best, especially among females. The JB model gave the worse fit and failed when fitted on 0–2 years NINFEA data.

Examinations of the normal quantile plots of the residual errors estimated for each model showed that the assumption of normality appeared to be better satisfied by the SITAR model than the two anthropometric models (the data are not shown). This was seen for each data set, although even the SITAR estimated residuals showed some kurtosis (which was also reported by Beath (2007)).

4.3.2. Population level predicted curves

The population level predicted curves from the best specification of the three models fitted on the Geracão XXI data (0–2 years) are shown in Fig. 2, stratified by gender. In all the figures curves predicted by the SITAR models fitted on the log-weight scale have been back transformed so that weights are shown on the same scale for all the three models. As expected the

Table 4. Goodness of fit of the best specification of the three models by cohort, gender and age range

Study	*Results for JB model*		*Results for Reed model*		*Results for SITAR model*	
	AIC	*BIC*	*AIC*	*BIC*	*AIC*	*BIC*
Males, 0–2 years						
Geracão XXI†	5406	5499	5217	5310	5332	5407
GOCS‡	6047	6139	5889	5981	6054	6134
NINFEA§	—	—	3869	3952	4044	4110
Males, 0–4 years						
GOCS§	8949	9045	8813	8908	9031	9114
NINFEA§	5484	5570	5143	5228	5304	5372
Females, 0–2 years						
Geracão XXI†	4929	5022	4592	4685	4716	4790
GOCS§	5914	6007	5681	5774	5813	5893
NINFEA§§	—	—	3253	3335	3322	3382
Females, 0–4 years						
GOCS‡	9124	9220	8846	8942	8957	9040
NINFEA§	4569	4654	4486	4571	4536	4603

†SITAR: model 1, log(kg) and month scales.
‡SITAR: model 2, log(kg) and log(t) scales.
§SITAR: model 2, log(kg) and month scales.
§§SITAR: model 1, log(kg) and log(t) scales.

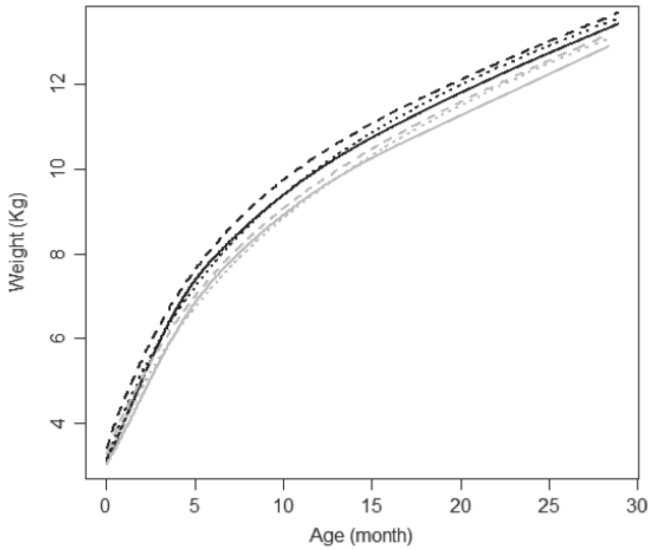


Fig. 2. Population predicted weight curves from the best specification of the three models fitted on the Geracão XXI data by gender (———, — — —, · · · · · ·, males; ———, — — —, · · · · · · , females): ———, SITAR model; — — —, JB model; · · · · · ·, Reed model

predicted curves for males lie above those for the females, regardless of the models that were used, with the difference increasing from birth till about 1 year and becoming approximately constant after that. The curve that was predicted by the JB model differs from those of the other two models, especially during the first months. In contrast the curves that were predicted by the Reed and the SITAR models overlap up to about 12–15 months before diverging slightly. These differences do not necessarily reflect differences in goodness of fit, but rather in the shape of their respective predicted curves when all random effects are set at 0. For this reason the curves that were predicted by the three models for the randomly selected Geracão XXI females of Fig. 1 are compared in Fig. 3. They show how the Reed and SITAR model most closely interpolate the original data in line with the results from the information criteria.

Further comparisons were made across the three cohorts. Fig. 4 shows the predicted population curves and predicted weight velocities that were obtained by the Reed and SITAR models on the 0–2-years males data of the three cohorts (with the SITAR model fitted by using model 2 specification on the log-weight and age scales; we do not show the results for the JB model because it failed on the 0–2-years NINFEA data). Even if the four-parameter Reed model in general allows for an inflection point, our transformation of the timescale, which we used to include measures taken at birth, did not lead to an estimation of a maximum for the growth velocities. According to both models, the typical trajectory of GOCS males (i.e. the trajectory of a child with zero random effects) is predicted to be slightly heavier than those from the other cohorts especially during the first year of life, possibly reflecting again the inclusion criteria that were used in this study. This is also seen in the velocity curves that were obtained from the SITAR model, with that for GOCS boys having their peak at a slightly earlier age compared with the rest (see also the predicted APWV that is reported in Table 3), and reaching a much higher size at that peak. The predicted APWVs for the NINFEA study are larger compared with those of the other cohorts but may be biased upwards because its growth data are collected only at birth, 3 and 6 months. To validate this we refitted the SITAR model on Geracão XXI

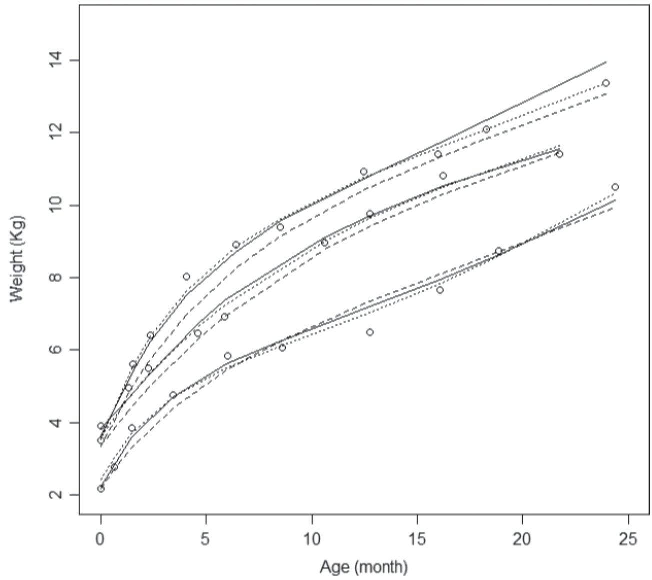
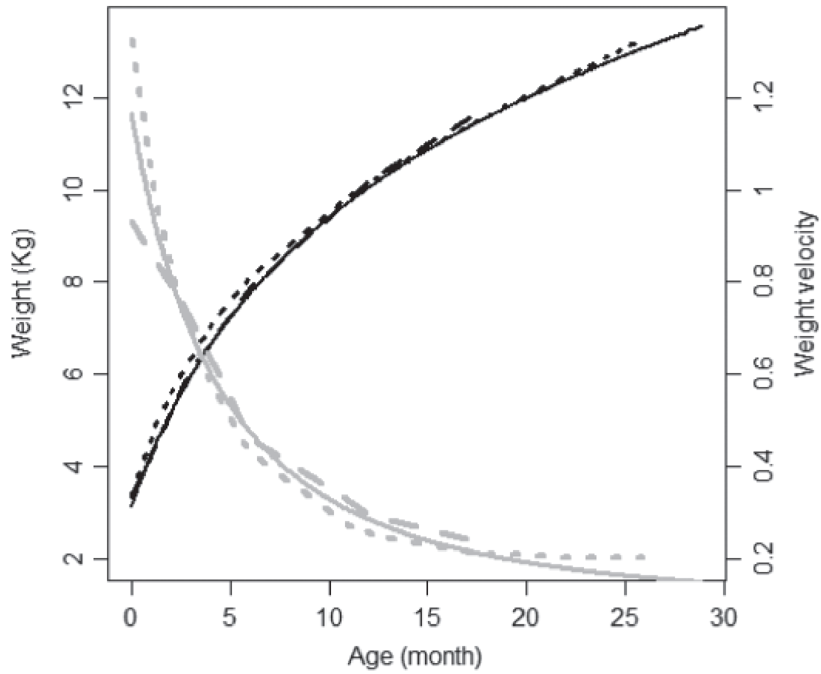


Fig. 3. Predicted weight growth curves of a stratified random selection of Geracão XXI females, superimposed on the observed data (○): ———, SITAR model; – – –, JB model; · · · · · ·, Reed model

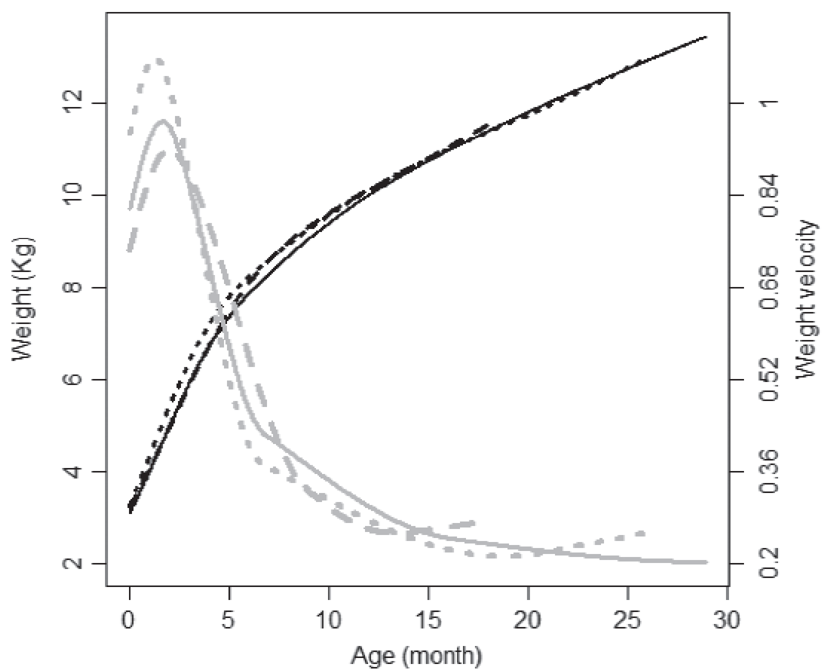
males after excluding all observations in the period 0.1–2.9 months to resemble the data in the NINFEA study: the new predictions of APWV were 2.17 months in males (instead of 1.65) and 2.13 months in females (instead of 1.90), becoming similar to those obtained in the NINFEA study (1.98 in males and 2.04 in females; see Table 3).

4.3.3. Random effects
Table 5 reports the SDs and correlations of the random effects from the three models when fitted on GOCS males (SITAR model: model 2 on log-weight and age scales). The results that were obtained from the other cohorts and for females are similar and are not reported for simplicity. Table 5 shows that the correlations of the JB and Reed random effects are extremely high (especially for the Reed model) whereas those of the SITAR model are substantially lower.

4.4. Model interpretation
4.4.1. Random effects
To aid the interpretation of the parameters of the three models, we refitted them on the combined males and females data, separately by cohort, and then compared the gender-specific distributions of their predicted random effects (best linear unbiased predictors) (Pinheiro and Bates, 2000). In Table 6 we report only the descriptive results that were obtained on the 0–2-years data separately by cohort. In each cohort the mean of the JB predicted random effects c_{1i} and a_{1i}, corresponding to the intercept of the asymptote and the intercept for the exponential term respectively, were higher in males than in females. However, b_{1i} and d_{1i} were not different. The Reed model random effects that most differed in terms of gender were a'_{1i}, which was on average higher among males, and c'_{1i} and d'_{1i}, the two deceleration terms, which were on average higher among females.



(a)



(b)

Fig. 4. Population predicted weight and velocity growth curves from (a) the Reed and (b) the SITAR model 2 log-weight and age models fitted on 0–2 years males from the three cohorts (——, – – –, ------, weight; ——, – – –, - - - -, weight velocity): ——, Geracão XXI; – – –, NINFEA; ------, GOCS

Table 5. SDs and correlations of the random effects estimated by the three models fitted on the GOCS males by age range

Parameter	*Results for the JB model*				*Results for the Reed model*				*Results for the SITAR model*		
	SD	*Correlations*			*SD*	*Correlations*			*SD*	*Correlations*	
0–2 years											
a (a') (α)	0.36	a	b	c	11.18	a'	b'	c'	0.16	α	β
b (b') (β)	0.11	0.91			4.1	0.91			0.60	0.76	
c (c') (γ)	1.81	0.99	0.90		16.57	−0.96	−0.98		0.30	−0.77	−0.70
d (d')	0.07	−0.73	−0.74	−0.66	14.92	−0.99	−0.95	0.99	—	—	—
0–4 years											
a (a') (α)	0.26	a	b	c	6.65	a'	b'	c'	0.15	α	β
b (b') (β)	0.07	0.82			1.52	0.85			0.64	0.74	
c (c') (γ)	1.35	0.98	0.70		7.54	−0.95	−0.95		0.26	−0.73	−0.66
d (d')	0.05	−0.49	−0.52	−0.40	7.86	−0.99	−0.90	0.98	—	—	—

Table 6. Mean of the predicted random effects by gender, model and cohort (0–2 years data)

Study	*Model*	*Means for a, a' or α*		*Means for b, b' or β*		*Means for c, c' or γ*		*Means for d or d'*	
		Males	*Females*	*Males*	*Females*	*Males*	*Females*	*Males*	*Females*
Geracão XXI	JB	0.05	−0.05	0.00	0.00	0.30	−0.27	0.00	0.00
	Reed	0.95	−1.19	−0.05	−0.03	−0.25	0.57	−0.85	1.16
	SITAR	0.00	0.00	−0.05	0.06	0.06	−0.07	—	—
GOCS	JB	0.05	−0.05	0.00	0.00	0.24	−0.23	0.00	0.00
	Reed	1.17	−1.15	0.08	−0.07	−0.66	0.65	−1.20	1.19
	SITAR	0.01	−0.02	−0.04	−0.04	0.04	−0.02	—	—
NINFEA	JB	—	—	—	—	—	—	—	—
	Reed	1.24	−1.38	0.23	−0.25	−0.98	1.06	−1.39	1.53
	SITAR	0.02	−0.03	−0.04	0.02	0.04	−0.04	—	—

Results from the SITAR model varied across cohorts: whereas on average the predicted growth velocities γ_{1i} were higher among males in each cohort, size α_{1i} was higher in males in the NINFEA and the GOCS but not in the Geracão XXI study, and tempo β_{1i} was higher in females only in the Geracão XXI study (and to a lesser extent in the NINFEA study), in agreement with the corresponding differences in predicted APWV (see Table 2).

4.4.2. *Super-imposition translation and rotation model parameters*

Both Beath (2007) and Cole *et al.* (2010) stressed that a biological interpretation could be attached to the coefficients that are derived from the SITAR model. To clarify this we examined the model-predicted trajectories and random effects for a selection of children and compared them with their respective population-predicted curve. Here we report the findings for a GOCS male who had a very high birth weight (4.55 kg) and who was still considerably overweight at 24 months. His predicted random coefficients are $\hat{\alpha}_{1j} = 0.44$ (the maximum of the distribution

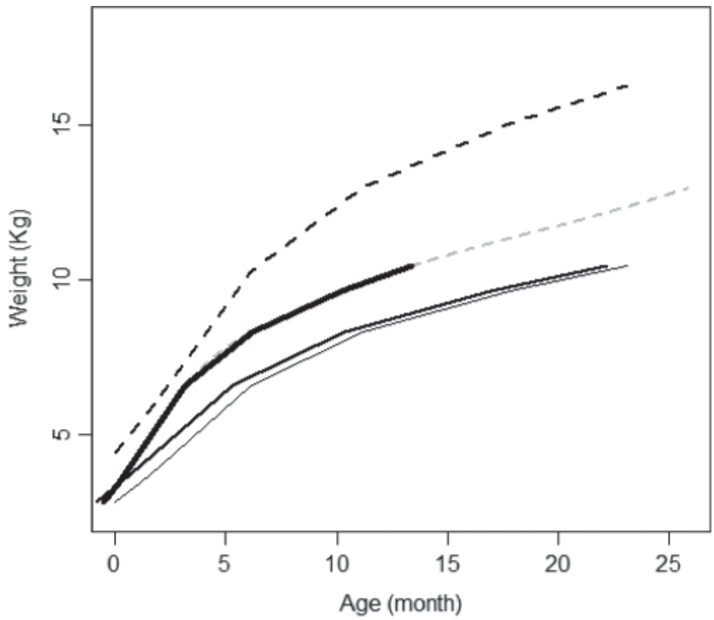


Fig. 5. Predicted individual growth curve of a selected GOCS male and new curves created after progressively removing the three random effects from the SITAR model (fitted on the GOCS males aged 0–2 years on the log-weight and age scale ($\hat{\alpha}_i = 0.44$; $\hat{\beta}_i = 0.88$ months; $\hat{\gamma}_i = -0.51$): – – –, population predicted curve; – – –, individual predicted curve; ——, removing effect of $\hat{\alpha}_i$; ——, removing effect $\hat{\alpha}_i + \hat{\beta}_i$; ——, removing effect of $\hat{\alpha}_i + \hat{\beta}_i + \hat{\gamma}_i$

of $\hat{\alpha}_{1i}$), $\hat{\beta}_{1j} = 0.88$ (approximate 95th percentile of the distribution of $\hat{\beta}_{1i}$) and $\hat{\gamma}_{1j} = -0.51$ (approximate fifth percentile of the $\hat{\gamma}_{1i}$-distribution). Whereas we expected the values of the first two predicted random effects to be high, the negative value for $\hat{\gamma}_{1j}$ was somewhat surprising, owing to the large size of this child, observed at all ages. However, Fig. 5 clarifies how this happens as it shows that the child-specific predicted curve (the dark dotted curve) is first shifted downwards when his departure from the population average size, due to his larger weight, $\hat{\alpha}_{1j}$, is removed (the thin full curve) and then shifted to the left when his departure from the population tempo, due to his late APWV (which was estimated to be at around 3 months), $\hat{\beta}_{1j}$, is removed (the medium thick full curve). Because of these realignments the curve is less steep than the predicted population curve, explaining the negative value of $\hat{\gamma}_{1j}$. When the contribution of $\hat{\gamma}_{1j}$ is also removed the individual predicted curve and the population predicted curve, as expected, overlap (the thick full curve).

This example illustrates how to attribute biological meanings to random effects predicted from a SITAR model. They also highlight their close interrelationships and the care needed for their interpretation: the value of each parameter is strongly influenced by the values of the other two. This was also evidenced by the correlations shown in Table 5. Estimated correlations between α_{i1} and β_{i1} are positive, suggesting that children with lower weight are also those with an earlier tempo of growth (and earlier APWV). Estimated correlations between α_{i1} and γ_{i1} and between β_{i1} and γ_{i1} are instead of similar strength but negative, suggesting that smaller children, or children with an earlier, i.e. negative, tempo experience greater growth velocities.

These observations lead to the conclusion that examination of the SITAR random-effects parameters, even if biologically interpretable, requires a consideration of their conditional distributions. In this light we examined the marginal associations between the gender of the child

Table 7. Marginal and mutually adjusted effects of gender on estimated size, tempo and velocity, by cohort (fitted on the 0–2 years data)†

Dependent variable	*Effect of gender (male versus female)*					
	Geracão XXI		*GOCS*		*NINFEA*	
	Marginal coefficient	*Adjusted‡ coefficient*	*Marginal coefficient*	*Adjusted‡ coefficient*	*Marginal coefficient*	*Adjusted‡ coefficient*
Size ($\hat{\alpha}_{1i}$)	−0.001 (0.012)	0.049 (0.008)	0.030 (0.008)	0.039 (0.005)	0.045 (0.008)	0.068 (0.006)
Tempo ($\hat{\beta}_{1i}$)	−0.112 (0.048)	−0.095 (0.041)	−0.0001 (0.034)	−0.003 (0.019)	−0.056 (0.042)	−0.129 (0.038)
Velocity ($\hat{\gamma}_{1i}$)	0.132 (0.021)	0.129 (0.015)	0.063 (0.015)	0.078 (0.009)	0.074 (0.016)	0.116 (0.009)

†Standard errors are given in parentheses.
‡Adjusted for the other SITAR random effects.

and each of size, tempo and velocity and also the same associations but conditional on the other random effects (Table 7). In the Geracão XXI and NINFEA studies the conditional effects differ from the marginal effects, with males having larger size when differences in the other two dimensions are accounted for. The latter means that, comparing the growth curve of a female with that of a male with similar velocity and tempo, the males curve will lie above the females curve or, in other words, that the mean of the size coefficient, stratified by—for example—quartiles of the velocity and tempo distributions is expected to be greater in males than in females. In the NINFEA study the conditional effects of gender on tempo and velocity are also stronger than the corresponding marginal effects. In contrast in the GOCS there is no evidence of a difference in tempo by gender, with the differences in size and velocity only marginally increased by adjustment. Given that GOCS children are selected to be more similar in terms of gestational age and birth weight than those from the other cohorts, this indicates that the degree of variation in gestational age and birth weight both contribute to the random-effects correlations.

4.4.3. Influence of gestational age

Several results, as observed above, identified differences between estimates that were obtained when analysing children in the GOCS cohort as opposed to the other cohorts. The most likely explanation is that participation in the GOCS was restricted to the term babies. Indeed it is well known that gestational age has a very strong influence not just on size at birth but also on the growth trajectory after birth (Sullivan *et al.*, 2008). To examine whether such heterogeneity had an influence on the estimated parameters and therefore on the individual growth curves predicted by the SITAR model, we compared the original results that were obtained when analysing the Geracão XXI data with results obtained when either

(a) preterm babies were excluded or
(b) the age scale was redefined as time since conception (i.e. age plus gestational age, both measured in weeks).

Excluding preterm babies (61 children) did not lead to any changes in the predicted population curves, whereas some changes were found in the estimated measures of variation: the SD of $\hat{\beta}_{1i}$

and its correlation with the other two random effects were smaller, becoming more similar to the values that were found in the GOCS. In contrast the average APWVs were virtually unchanged: 1.66 months in males and 1.95 months in females (instead of 1.65 and 1.90 months respectively). When we re-examined the conditional associations between the gender of the child and each of size, tempo and velocity, we observed that the effect on $\hat{\beta}$ decreased from -0.095 (standard error SE $= 0.041$) to -0.066 (SE $= 0.033$), whereas the effects on the other two parameters were substantially unchanged.

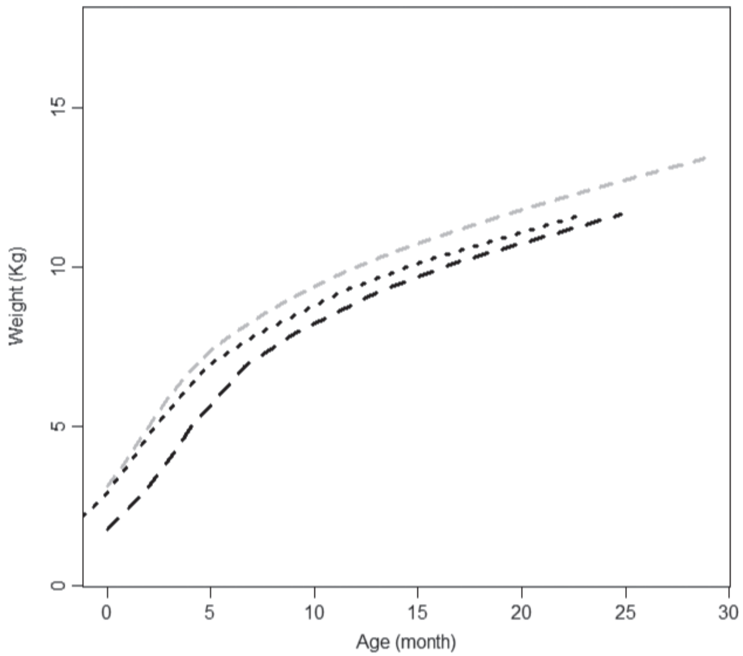
When analyses were rerun on the new age scale of time since conception (in weeks), part of the information held in the tempo-parameter β_{1i} appeared to have been removed by the new timescale. Fig. 6 illustrates this. Fig. 6(a) shows the predicted growth curve for a male from the Geracão XXI cohort whose gestational age was 32 weeks and whose weight at birth was 1.86 kg. In Fig. 6(b) the growth curve of the same child is shown when predicted by the model fitted on time since conception. The respective predicted population curves are also shown. The estimated child-specific parameters from the two models are $\hat{\alpha}_{1i} = -0.07$, $\hat{\beta}_{1i} = 1.77$ and $\hat{\gamma}_{1i} = 0.02$, when fitted on the original age scale, and $\hat{\alpha}_{1i} = -0.04$, $\hat{\beta}_{1i} = 1.29$ (in weeks, which corresponds to about 0.3 months) and $\hat{\gamma}_{1i} = -0.05$, for the time-since-conception scale (centred at about 39 weeks). Hence tempo is reduced to about a sixth when gestational age is accounted for. Indeed when the effect of $\hat{\beta}_{1i}$ is removed from the predicted individual curve in Fig. 6(a), the shape of the derived curve (the dotted curve) resembles the predicted individual curve in Fig. 6(b) (the broken curve). Removing only the effect of $\hat{\alpha}_{1i}$ from the latter curve (on the weeks-from-conception scale) was indeed sufficient to reach a near superimposition with the predicted population curve (the full curve in Fig. 6(b)).

5. Discussion

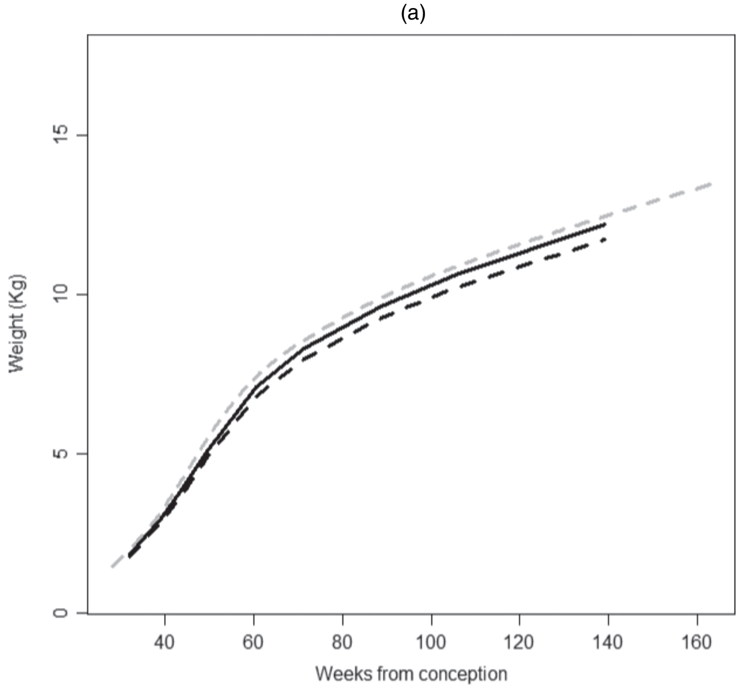
In this exploration of models for infant weight data we have used three recently established birth cohorts to compare three models where individual variations in growth are accounted for by specific random-effects parameters. Our main finding is that the choice of which model to adopt varies with the aim of the study and, less crucially, on the richness of the available data. If interest focuses on describing individual early weight growth trajectories and/or typical profiles then all the three models that were considered in this paper are suitable. Among them the four-parameter Reed model performed best in terms of standard fit criteria and, unlike the other two models, is linear in its parameters and is therefore easier to estimate. However, if the focus is on extracting salient features of the growth trajectories to include them in further analyses where growth is either the exposure or the outcome, then the SITAR model may be the preferred option because of the biological interpretability of its parameters. However, several other factors should be considered before a choice is made.

5.1. Range and frequency of observations over time

We have shown that including measures taken at birth requires the transformation of the timescale when fitting the Reed model because it contains a logarithmic and an inverse function of age. We have adopted one particular transformation that gives a value of 1 to time of birth (and a value of 0 to 9 months before birth, i.e. roughly at 'conception'). An alternative transformation had been suggested in the literature that adds a value of 1 to time only to the terms that involve the logarithmic and the inverse function of time. This hybrid solution was, in our view, unsatisfactory and indeed performed slightly worse in terms of goodness of fit than the solution that we have adopted. A drawback of the latter solution, however, is that, when applied to our data, the model's derivative (i.e. the weight velocity curve) did not present a maximum



(a)



(b)

Fig. 6. Growth curve for a selected child, Geracão XXI cohort, predicted when fitted on (a) the age timescale ($\hat{\alpha}_i = -0.07$; $\hat{\beta}_i = 1.77$ months; $\hat{\gamma}_i = 0.02$; — — —, population predicted curve; — — —, individual predicted curve; - - - - - -, removing effect of $\hat{\beta}_i$ only) or (b) on the weeks-since-conception timescale ($\hat{\alpha}_i = -0.04$; $\hat{\beta}_i = 0.30$ months; $\hat{\gamma}_i = -0.05$; — — —, population predicted curve; — — —, individual predicted curve; ———, removing effect of $\hat{\alpha}_i$ only)

and therefore did not allow the identification of an APWV, if such a peak was actually present. This was not an issue with the SITAR model even when a transformation was required to meet the distributional assumptions, because the peak in the weight velocity curve was still identifiable. Hence, if the APWV is of interest, and if data include measurements at birth, the SITAR model, fitted on the appropriate growth scale and timescale to meet the assumptions, should be preferred.

Range, frequency and regularity of the observations over time are other important aspects to consider. Spline functions such as that used by the SITAR model are necessarily influenced by data points that are isolated and this was so with the analyses of the NINFEA 0–4-years data.

5.2. *Interpretation of the child-specific parameters*

By comparing the distribution of random effects that was predicted for males and females by the three models, when fitted on the combined data, we aimed to examine the discriminatory values of these parameters. Those predicted from the JB and Reed models that correspond to their intercepts ($\hat{c}_{1i}$ and $\hat{a}'_{1i}$), or correction to the intercept ($\hat{a}_{1i}$), did show differences between the genders. However, on their own these parameters are unlikely to discriminate the growth trajectories of subgroups defined by weaker predictors of growth than gender (for more discussion of this see the on-line appendix). In contrast each of the three SITAR random effects appeared to have distinct means and distributions between the genders, especially when their intercorrelations were taken into account (note that accounting for correlations between the parametric models' random effects was not possible because of extreme collinearity).

The interpretability of the SITAR model, however, requires additional care. In particular the interpretation of the tempo-parameter, which represents the shift on the time axis that is necessary to synchronize the curves which are centred on the tempo-milestone (the APWV in this setting), depends on the timescale that is used and the setting analysed. When changed by setting the time origin at conception, we found that gender differences with respect to tempo were slightly reduced. Thus this parameter represents an adjustment that is necessary to proxy true biological age better (hence measuring growth adjusted for maturation or developmental status). This is in line with results that were obtained for the cohort including only term babies with birth weight between 2500 and 4500 g (the GOCS), for which we observed that estimates of tempo did not vary across genders when using chronological age as the timescale, and it is also in line with the fact that the gender effect on tempo decreased when estimated by using only term babies in the Geracão XXI study compared with using the whole cohort. In other settings, with different growth variables and especially with different age ranges, the tempo-parameter represents adjustments of the timescale by other factors, such as hormonal levels in puberty; Cole *et al.* (2010).

A final note of caution with regard to the interpretability of the SITAR parameters derives from its extreme flexibility. Unlike the JB and Reed models, the SITAR model is not restricted to an age range; however, the wider the range of ages included in a SITAR model, the less clear will be the interpretation of its parameters. This would affect the tempo- and velocity-parameters in particular because individual trajectories over longer age ranges may include several inflection points, leading to multiple changes in velocity and therefore to the parameters representing a complex average of several, and possibly very different, departures from the population timescale. This would happen for example if a SITAR model were used to model the body mass index BMI (weight/height^2) from age 0 to 10 years, say. During this time children would experience both a peak velocity at around 1 year and a minimum (negative) velocity at about 5 years; Rolland-Cachera *et al.* (1987). Moreover, a single parameter tempo would represent an average of the child's departure from possibly two separate biological timescales: that linked to the infancy BMI-peak and that linked to the childhood BMI-rebound.

5.3. Issues related to life course epidemiology

Much interest in modelling growth data comes from research in life course epidemiology (Kuh and Ben-Shlomo, 2004), where summary growth parameters such as those described in this paper are used as explanatory variables for the onset of a later outcome. Thus analyses consist of two components, carried out in stages: a first stage focused on modelling the growth data, and a second stage aimed at relating parameters from the growth model to the distal health outcome. For such analyses to succeed it is crucial that salient features of growth are estimated for all children in the study and that the statistical uncertainties of the parameters extracted in the first stage are properly accounted for; Molinari and Gasser (2004). The former may not be possible if parameters used to summarize the growth data are not available for all children (for example see Silverwood *et al.* (2009)), with the second-stage analyses consequently being affected by selection bias. Adopting the SITAR modelling approach, and using the three random parameters of size, tempo and velocity as explanatory variables, would not suffer from this. Moreover, using all three SITAR parameters instead of a single growth indicator, such as the APWV, would lead to a more comprehensive summary of the individual growth patterns.

The additional statistical uncertainties arising from using estimated growth parameters in the second stage of the analyses are rarely addressed in applications. However, they can be dealt with, at least approximately, by correcting the standard errors of the second-stage estimates by using the bootstrap (Efron and Tibshirani, 1993), or weighting the second stage according to the amount of child-specific information available in the first stage (e.g. the number of observations; see Tzoulaki *et al.* (2010)). The former approach would be quite cumbersome given the complexities of estimating these models. A one-step estimation of the two model components, growth and distal outcome modelling, would not require such inferential adjustments. This is where specifying the growth model by using random-effects polynomial curves would be advantageous as they could be estimated jointly with the distal outcome model by using structural equation models; De Stavola *et al.* (2006). This approach, however, would suffer from a lack of biological interpretability of the random-effects parameters. Moreover it is not easily implemented when fitting more complex growth models, at least by using standard software.

Another aspect of growth modelling that is relevant to life course epidemiology is that focused on identifying factors that influence childhood growth. Again such analyses could be carried out in two stages, where some growth parameters are first estimated and then modelled in terms of exposures of interest (as we did in Section 4.4). Alternatively models could be specified so that estimation is carried out in a single step. This is easily achieved within the framework of random-effects polynomial models (for example see dos Santos Silva *et al.* (2002)). Of the models that were reviewed here, however, only the SITAR model has been generalized to include explanatory factors (Beath, 2007; Hui *et al.*, 2010; Johnson *et al.*, 2011), although such specification allows only the inclusion of explanatory variables for size and for the shifted and rescaled timescale (which was denoted by Beath (2007) as the 'growth rate'). Thus the interpretation of estimated effects of explanatory variables for the SITAR growth parameters needs particular care. Moreover, given the non-linearity of the SITAR model, the explanatory variable parameters should not be used to infer population-average effects; Fitzmaurice *et al.* (2009).

Finally in this field, there is much interest in comparing and summarizing evidence from different studies. However, not one but several study-specific parameters for each growth dimension would need to be compared, i.e. size, tempo and velocity. Simultaneous analysis of all parameters could be carried out by fitting multivariate metaregression models that account for the covariance structure of the parameters estimated within each study; van Houwelingen *et al.* (2002).

5.4. Conclusions

In this paper we have examined alternative modelling approaches to the estimation of the salient features of growth in weight in infancy and early childhood and discussed the difficulties, advantages and disadvantages of choosing each model. Our conclusion is that a range of options is available to researchers but that each necessitates careful understanding of assumptions and parameterizations. Of the three models that were discussed, the SITAR model is certainly the most flexible and the most useful for life course enquiries, as it allows the identification of important features of the children's growth trajectories. Its application, however, requires careful selection and substantive understanding of the model parameterization. This is true for the other models also but possibly to a lesser extent. However, the potential for extracting the most salient features of the growth patterns will depend on the richness and quality of the data, and this is where researchers should make the greatest investment.

6. Funding

The NINFEA cohort was partially funded by the Compagnia San-Paolo Foundation, Piedmont Region and Italian Ministry of University and Research. The Geracão XXI cohort was funded by Programa Operacional de Sade—Saúde XXI, Quadro Comunitário de Apoio III and by Administração Regional de Saúde Norte. The GOCS cohort was funded by Fondecyt projects 1060785, 1090252 and 1120326.

Acknowledgements

The authors thank Milena Maule, University of Turin, for many relevant insights. We are also grateful to all participants of the three cohorts.

References

Asefa, M., Drewett, R. and Hewison, J. (1996) An Ethiopian birth cohort study. *Paediatr. Perinat. Epidem.*, **10**, 443–462.

Baird, J., Fisher, D., Lucas, P., Kleijnen, J., Roberts, H. and Law, C. (2005) Being big or growing fast: systematic review of size and growth in infancy and later obesity. *Br. Med. J.*, **331**, 929.

Barker, D. J. P. (1998) *Mothers, Babies and Health in Later Life*. Edinburgh: Churchill Livingstone.

Beath, K. J. (2007) Infant growth modelling using a shape invariant model with random effects. *Statist. Med.*, **26**, 2547–2564.

Berkey, C. S. (1982) Comparison of two longitudinal growth models for preschool children. *Biometrics*, **38**, 221–234.

Berkey, C. S. and Reed, R. B. (1987) A model for describing normal and abnormal growth in early childhood. *Hum. Biol.*, **59**, 973–987.

Box, G. E. P. and Cox, D. R. (1964) An analysis of transformations (with discussion) *J. R. Statist. Soc.* B, **26**, 211–252.

Brown, R. C., Dwyer, T., Kasten, C., Krotoski, D., Li, Z., Linet, M. S., Olsen, J., Scheidt, P., Winn, D. M. and International Childhood Cancer Cohort Consortium (2007) Cohort profile: the International Childhood Cancer Cohort Consortium (I4C). *Int. J. Epidem.*, **36**, 724–730.

CHICOS (2010) CHICOS: developing a child cohort research strategy for Europe. Centre for Research in Environmental Epidemiology, Barcelona. (Available from `http://www.chicosproject.eu/`.)

Cole, T. J., Donaldson, M. D. and Ben-Shlomo, Y. (2010) SITAR—a useful instrument for growth curve analysis. *Int. J. Epidem.*, **39**, 1558–1566.

Count, E. W. (1943) Growth patterns of the human physique: an approach to kinetic anthropometry. *Hum. Biol.*, **15**, 1–32.

Deming, J. and Washburn, A. H. (1963) Application of the jenss curve to the observed pattern of growth during the first eight years of life in forty boys and forty girls. *Hum. Biol.*, **35**, 484–506.
De Stavola, B. L., Nitsch, D., dos Santos Silva, I., McCormack, V., Hardy, R., Mann, V., Cole, T. J., Morton, S. and Leon, D. A. (2006) Statistical issues in life course epidemiology. *Am. J. Epidem.*, **163**, 84–96.
van Dommelen, P., van Buuren, S., Zandwijken, G. R. and Verkerk, P. H. (2005) Individual growth curve models for assessing evidence-based referral criteria in growth monitoring. *Statist. Med.*, **24**, 3663–3674.
Efron, B. and Tibshirani, R. (1993) *An Introduction to the Bootstrap*. New York: Chapman and Hall.
Fitzmaurice, G., Davidian, M., Verbeke, G. and Molenberghs, G. (eds) (2009) *Longitudinal Data Analysis*. Boca Raton: Chapman and Hall–CRC.
Goldstein, H. (2011) *Multilevel Statistical Models*, 4th edn. London: Arnold.
van Houwelingen, H. C., Arends, L. R. and Stijnen, T. (2002) Advanced methods in meta-analysis: multivariate approach and meta-regression. *Statist. Med.*, **21**, 589–624.
Hui, L. L., Leung, G. M., Cowling, B. J., Lam, T. H. and Schooling, C. M. (2010) Determinants of infant growth: evidence from Hong Kong's "Children of 1997" birth cohort. *Ann. Epidem.*, **20**, 827–835.
Huxley, R. R., Shiell, A. W. and Law, C. M. (2000) The role of size at birth and postnatal catch-up growth in determining systolic blood pressure: a systematic review of the literature. *J. Hypertens.*, **18**, 815–831.
Jenss, R. M. and Bayley, N. (1937) A mathematical method for studying the growth of a child. *Hum. Biol.*, **9**, 556–563.
Johnson, L., Llewellyn, C. H., van Jaarsveld, C. H., Cole, T. J. and Wardle, J. (2011) Genetic and environmental influences on infant growth: prospective analysis of the Gemini twin birth cohort. *PLOS One*, **6**, no. 5, article e19918.
Kuh, D. and Ben-Shlomo, Y. (2004) *A Life Course Approach to Chronic Disease Epidemiology*. New York: Oxford University Press.
Lindstrom, M. L. and Bates, D. M. (1990) Nonlinear mixed effects models for repeated measures data. *Biometrics*, **46**, 673–687.
Manwani, A. H. and Agarwal, K. N. (1973) The growth pattern of indian infants during the first year of life. *Hum. Biol.*, **45**, 341–349.
Molinari, L. and Gasser, T. (2004) The human growth curve: distance, velocity and acceleration. In *Methods in Human Growth Research* (eds R. Hauspie, N. Cameron and L. Molinari). Cambridge: Cambridge University Press.
Pinheiro, C. and Bates, D. M. (2000) *Mixed-effects Models in S and S-PLUS*. New York: Springer.
Richiardi, L., Baussano, I., Vizzini, L., Douwes, J., Pearce, N. and Merletti, F. (2007) Feasibility of recruiting a birth cohort through the Internet: the experience of the NINFEA cohort. *Eur. J. Epidem.*, **22**, 831–837.
Richter, L. M., Victora, C. G., Hallal, P. C., Adair, L. S., Bhargava, S. K., Fall, C. H., Lee, N., Martorell, R., Norris, S. A., Sachdev, H. S., Stein, A. D. and COHORTS Group (2012) Cohort profile: the consortium of health-orientated research in transitioning societies. *Int. J. Epidem.*, **41**, 621–626.
Rolland-Cachera, M. F., Deheeger, M., Guilloud-Bataille, M., Avons, P., Patois, E. and Sempe, M. (1987) Tracking the development of adiposity from one month of age to adulthood. *Ann. Hum. Biol.*, **14**, 219–229.
Rubin, D. (1976) Inference and missing data. *Biometrika*, **63**, 581–592.
Ruppert, D., Wand, M. P. and Carroll, R. J. (2003) *Semiparametric Regression*. Cambridge: Cambridge University Press.
dos Santos Silva, I., De Stavola, B. L., Mann, V., Kuh, D., Hardy, R. and Wadsworth, M. E. (2002) Prenatal factors, childhood growth trajectories and age at menarche. *Int. J. Epidem.*, **31**, 405–412.
Silverwood, R. J., De Stavola, B. L., Cole, T. J. and Leon, D. A. (2009) BMI peak in infancy as a predictor for later BMI in the Uppsala Family Study. *Int. J. Obes. Lond.*, **33**, 929–937.
Simondon, K., Simondon, F., Delpeuch, F. and Cornu, A. (1992) Comparative study of five growth models applied to weight data from congolese infants between birth and 13 months of age. *Am. J. Hum. Biol.*, **4**, 327–335.
Sullivan, M., McGrath, M., Hawes, K. and Lester, B. (2008) Growth trajectories of preterm infants: birth to 12 years. *J. Ped. Hlth Care*, **22**, 83–93.
Tzoulaki, I., Sovio, U., Pillas, D., Hartikainen, A. L., Pouta, A., Laitinen, J., Tammelin, T. H., Jarvelin, M. R. and Elliott, P. (2010) Relation of immediate postnatal growth with obesity and related metabolic risk factors in adulthood: the northern Finland birth cohort 1966 study. *Am. J. Epidem.*, **171**, 989–998.
Verbeke, G. and Molenberghs, G. (2000) *Linear Mixed Models for Longitudinal Data*. Berlin: Springer.
Yang, S., Tilling, K., Martin, R., Davies, N., Ben-Shlomo, Y. and Kramer, M. S. (2011) Pre-natal and post-natal growth trajectories and childhood cognitive ability and mental health. *Int. J. Epidem.*, **40**, 1215–1226.

Supporting information

Additional 'supporting information' may be found in the on-line version of this article:

'Appendix: extreme growth trajectories'.