Associations of growth trajectories in infancy and early childhood with later childhood outcomes\textsuperscript{1-4}

Kate Tilling, Neil M Davies, Emily Nicoli, Yoav Ben-Shlomo, Michael S Kramer, Rita Patel, Emily Oken, and Richard M Martin

\textbf{ABSTRACT}

\textbf{Background:} Weight and length at birth (which represent fetal growth) and weight and length or height gain during childhood (which potentially represent catch-up growth) may be related to later health outcomes. However, methods for the assessment of such relations are complex and underdeveloped.

\textbf{Objectives:} We aimed to describe childhood weight and length or height trajectories and to relate these to later outcomes by using rash at age 6.5 y as an example.

\textbf{Design:} The data came from a prospective cohort study in Belarus in 10,494 children born in 31 hospitals that participated in a cluster randomized trial of breastfeeding promotion. Weight and length or height were measured at birth, at scheduled clinic visits up to 1 y, and at 6.5 y; intermediate measures were obtained from routine child health records. Linear spline multilevel models for weight and length or height were used to estimate each child’s deviance from average birth weight, birth length, weight, and length or height gain velocity in each time period. Logistic regression was used to relate the outcome (parental report of rash at 6.5 y) to these weight and length or height estimates.

\textbf{Results:} The best-fitting splines for length or height and weight had knots at 3 and 12 mo, with another knot at 34 mo for height. The only relation between weight and length or height and reported rash was a positive association with weight gain velocity between 12 and 34 mo (odds ratio per SD increase in weight gain velocity: 1.11; 95% CI: 1.01, 1.22).

\textbf{Conclusion:} Advantages of multilevel models include no restriction to measures at arbitrary times or to individuals with complete data and allowance for measurement error. This trial was registered at ISRCTN.org as ISRCTN37687716. \textit{Am J Clin Nutr} doi: 10.3945/ajcn.110.001644.

\textbf{INTRODUCTION}

There is increasing emphasis in medical research on fetal and childhood antecedents of disease and how these interact with other exposures throughout the life course to influence later-life conditions (1). To answer questions about the relative importance of fetal programming and childhood development requires appropriate analyses of longitudinal data. For example, we might ask how fetal development (represented by birth weight and length) and subsequent weight or length/height changes during childhood relate to health outcomes in early adulthood. Low birth weight has been associated with increased risk of asthma (2) and decreased risk of atopic dermatitis (3) in late childhood. Shorter length/height (4) and higher body mass index (BMI; in kg/m\textsuperscript{2}) in childhood have also been related to increased incidence of asthma (4, 5). There is some evidence that length gain during the first year may also be positively related to asthma (particularly in those of low birth length) but not to atopy (6).

Analysis of life course data poses statistical problems. Analyses must account for dependencies between repeated observations on the same person: methods to do this (eg, multilevel models) are now widely available in standard statistical software packages (7). Where there are repeated measures of exposures related to a later-life outcome, standard regression models may be affected by multicollinearity because of both the lack of independence and the serial autocorrelation. Measurement error may vary over time (eg, absolute measurement error in weight will be larger in adulthood than in childhood). There will also usually be dropout over time (due to death, illness, emigration, nonresponse, and other reasons).

Commonly used approaches in the programming literature include z score plots (8) and “conditional regression” (also known as “life course plots”) (9). The z score plots do not allow for correlation between successive exposures, nor do they describe how the exposure for a given individual changes with time, and because this method conditions on an effect (the disease status), it can induce associations in the absence of causal effects (10). Life course plots are often interpreted as providing information about relations between the outcome and both current exposure and change in exposure, in which a change of sign between 2 times indicates that the change between those times is related to the

\textsuperscript{1} From the Department of Social Medicine (KT, EN, YB-S, RP, and RMM) and the Causal Analyses in Translational Epidemiology Centre (NMD), University of Bristol, Bristol, United Kingdom; the Departments of Pediatrics, and Epidemiology, Biostatistics and Occupational Health, McGill University Faculty of Medicine, Montreal, Canada (MSK); and the Department of Population Medicine, Harvard Medical School and Harvard Pilgrim Health Care Institute, Boston, MA (EO).

\textsuperscript{2} Presented at the conference “The Power of Programming: Developmental Origins of Health and Disease,” held in Munich, Germany, 6–8 May 2010.

\textsuperscript{3} Supported by a grant from the Canadian Institutes of Health Research and by grant FOOD-DT-2005-007036 from the European Union’s project on Early Nutrition Programming: Long-term Efficacy and Safety Trials (to RMM).

\textsuperscript{4} Address correspondence to K Tilling, University of Bristol, Department of Social Medicine, Canynge Hall, 39 Whatley Road, Bristol BS8 2PS, United Kingdom. E-mail: kate.tilling@bristol.ac.uk. doi: 10.3945/ajcn.110.001644.
outcome (9). Because successive exposures are autocorrelated, the regression coefficients are affected by collinearity; this problem increases with the number of prior exposures conditioned on. These methods share 2 major disadvantages: the samples at each time point must be the same (or random subsamples of each other) and measurements for all individuals must be taken simultaneously.

It is therefore critical to establish a general framework in which such lifecourse data can be analyzed flexibly and robustly. We aimed to use multilevel models to describe the trajectory of weight and length/height gain in early childhood, and to show how these trajectories can be related to later health outcomes, with the use of rash in later childhood as an example to show the principles.

METHODS

A detailed description of the original Promotion of Breast-feeding Intervention Trial (PROBIT) has been published (11). Briefly, 31 maternity hospitals in the Republic of Belarus and one each of their affiliated polyclinics (outpatient clinics for routine health care follow-up) were randomly assigned to a breastfeeding promotion intervention or to usual practice. Overall, 17,046 full-term (>37 wk gestation) healthy singletons who weighed ≥2500 g, with Apgar score ≥ 5 at 5 min, and who initiated breastfeeding, were recruited during their postpartum hospital stay between June 1996 and December 1997. These children were followed up at polyclinic visits at 1, 2, 3, 6, 9, and 12 mo; home visits were made when polyclinic visits were missed. At 6.5 y of age, 13,889 (82%) children attended a follow-up visit (12). The institutional review board of the Montreal Children’s Hospital approved both the original PROBIT trial and the 6.5-y follow-up, and the mothers who participated signed consent forms (in Russian) before entry into the PROBIT and the 6.5-y follow-up. Our study sample comprises the 10,495 (76%) participants at 6.5 y who had complete data on all potential confounders (see below) and at least one measure of weight between birth and 5 y.

Birth weight, gestational age, mother’s smoking in pregnancy, and mother’s and father’s education were prospectively recorded during the postpartum stay (11). Weight and length were measured by pediatricians at scheduled study visits at 1, 2, 3, 6, 9, and 12 mo (11), and >93% of the children in our sample attended all 6 visits. Because differences in weight and length/height gain were not major hypotheses of the PROBIT, measurements of weight and length/height during infancy were not standardized.

Weights and heights of children are measured at routine checkups in Belarus. Data on these weights and heights between 12 mo and 5 y were abstracted retrospectively from medical records by the child’s pediatrician. Follow-up interviews and examinations at 6.5 y of age were performed by 1 pediatrician each at 24 of the 31 polyclinics; in the 7 high-volume clinics that remained, the follow-up visits were shared by 2 pediatricians. Standing height was measured with a wall-mounted stadiometer and weight with an electronic digital scale (Bella 840; Seca Corporation, Hamburg, Germany) (12). Weight and height were each measured twice, and the average of the 2 measurements was used (12). An audit of the measures for 190 children performed an average of 18 mo after the study follow-up visit showed test–retest correlations of 0.84 for height and 0.89 for BMI (12). Parental height and weight were obtained by interview with the parent who accompanied the child. For most children (91%), the mother reported height and weight for both herself and the child’s father; in a minority (8%), the father (or occasionally a legal guardian) reported for both parents. Overall, 1379 (10%) of the fathers and 161 (1%) of the mothers did not have data on their reported height and weight.

Allergy symptoms and diagnoses were ascertained at 6.5 y of age (13) with the International Study of Asthma and Allergy in Childhood (ISAAC) questionnaire, which had already been translated into Russian and validated by the ISAAC investigators (14). The outcome used as an example here is the binary outcome “rash,” defined as a positive response to the question “Has your child ever had an itchy rash which was coming and going for at least 6 months?”

STATISTICAL METHODS

Summary of individual weight gain trajectories

Models for changes in exposure over part or all of an individual’s lifetime are often known as “growth trajectory” models. Trajectories can be modeled with multilevel models, in which random effects or individual-level residuals represent an individual’s underlying growth pattern (7). These strategies model growth data efficiently because, provided that data are missing at random, conditional on age and other observed variables, they use all available data, and are not restricted to individuals with complete data at all times, or data measured at the same times for all individuals. The occasion-level residuals represent the measurement error in exposure.

When exposures exhibit nonlinear patterns of change, it is necessary to choose a best-fitting function of time. Fractional polynomials compare fit among a family of flexible polynomial functions (15). However, the polynomial terms are not easily interpretable, especially if they are to be used to relate characteristics of growth to later outcomes. An alternative, more interpretable, approach is to use a series of linear [or nonlinear (16)] splines to model the trajectory. For example, weight could be allowed to have different linear slopes from 0 to 3 mo, 3 to 6 mo, and so forth, and the slopes would vary between individuals. For simple regression models, guidance on the choice of number and position of “knots” (the points at which spline curves join) includes subjective methods such as the use of a large number of knots and a decrease in the number until a “smooth” curve is reached, or more objective methods such as the placement of knots at centiles of the distribution of the time variable or stepwise regression to select those knots that are “significant” (17).

We decreased the dimensionality of the data by the estimation of individual weight gain trajectories with the use of a multilevel model, fitted with the use of MLWiN software, version 2.10 (18). We used fractional polynomials to find the best-fitting weight and length/height gain trajectories for boys and girls separately. We then fitted knot points at the planned clinic times and chose the model that best fitted the fractional polynomial. Our criterion for “best fit” was the smallest average absolute residuals over 5 periods: 0–5.99 mo, 6–11.99 mo, 1–1.99 y, 2–2.99 y, 3–3.99 y, and 4–5 y. We chose this criterion to identify a model that fitted best over the span of the exposure period, rather than one that fitted best where there was a higher concentration of measurements. Other criteria, such as fit of the linear spline model to the fractional polynomials, have been used (19), and we showed similar results if these criteria were applied.
Estimating associations between rash and weight and length/height gains

Logistic regression models related reported rash (outcome) to the individual estimates for each child of birth weight, early-infant weight gain, late-infant weight gain, and early-childhood weight gain (exposures), and the equivalent length/height measures. Models adjusted for intervention arm and clustering by polyclinic, potential baseline confounders (maternal and paternal heights, BMI, and highest educational levels), and the preceding weight or length/height measures (eg, the model for the effect of late-infant weight gain on rash included birth weight and early-infant weight gain as covariates). A further model (weight gain exposures only) also adjusted for length/height at the time of the weight measurement and for the length/height measure that preceded it.

We used path analysis to distinguish the direct effects of each weight and height gain measure on rash at 6.5 y (those effects not mediated through subsequent weight or height) from the indirect effects (those effects mediated through subsequent weight or height) (20). We also repeated all analyses without the inclusion of the potential confounding variables (parental height, BMI, and educational level) to examine the evidence for confounding by these variables. All analyses that related weight and length/height trajectories to outcome were carried out with the use of Stata software, version 11.0 (Stata Corp, College Station, TX).

RESULTS

The sample characteristics of the 10,502 children are summarized in Table 1. In total, 9766 (93%) of the children had all 6 scheduled weight and height measures (≤1 y). The median number of routine weight and height measures (between 1 y and the 6.5-y follow-up) was 4 (min 0, max 6). Of the sample included here, 10,494 (99%) had weight and height measured at the 6.5-y follow-up, at an average age of 6.6 (SD: 0.27) y.

The best-fitting splines for length/height and weight, in both boys and girls, had knots at 3 mo and 12 mo, with an extra knot at 34 mo for height. For simplicity and comparability between weight and length/height, we included the extra knot point at 34 mo in the weight model. Thus, for both boys and girls, we fitted a multilevel model for length/height, with linear gain between 0 and 3 mo (“early-infant length gain”), 3 and 12 mo (“late-infant length gain”), 12 and 34 mo (“early-childhood length gain”), and 34 and 60 mo (“mid-childhood height gain”). An equivalent model was fitted for weight. For each weight and length/height, there was evidence of differences in velocity between boys and girls (Table 2). From these models, individual estimates of birth weight and birth length and the 3 weight gain parameters and 4 length/height gain parameters were obtained for each child. The average patterns of change in weight and length/height (up to age 5 y) predicted by these linear spline models, for boys and girls in this population, together with 95% reference ranges, are shown in Table 2. For both length/height and weight, the rate of increase is high initially and then decreases with age. For weight, there is evidence of a slight increase in weight gain velocity after 34 mo.

The fit of the spline model for weight to the observed weight data is shown in Table 3. The 95% limits of agreement show that for birth weight, we would expect the difference between actual and predicted birth weight to lie between −0.63 kg and 0.78 kg for 95% of the children. These limits widen (as the scale of weight increases) with age, so that by 12 mo the difference would lie between −0.69 kg and 0.37 kg for 95% of children, and between −3.51 kg and 5.61 kg for 95% of children at 6.5 y. The length/height model showed similarly good fit to the observed length/height data (Table 3).

The weight and height exposures were then standardized (z scored) by sex. Birth weight and birth length were additionally standardized (z scored) by gestational age (in completed weeks, from 37 to 43). Rash was reported in 203 girls (4.0%) and 245 boys (4.5%). The odds ratios for rash for birth weight and weight gain in each period are shown in Table 4. There was no evidence of any relation between length at birth or length velocity during childhood and rash at 6.5 y. There was some

<table>
<thead>
<tr>
<th>Mother’s education [%]</th>
<th>Girls (n = 5052)</th>
<th>Boys (n = 5450)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Completed university</td>
<td>723 (14.3)</td>
<td>743 (13.6)</td>
</tr>
<tr>
<td>Advanced secondary or partial university</td>
<td>2633 (52.1)</td>
<td>2823 (51.8)</td>
</tr>
<tr>
<td>Common secondary</td>
<td>1532 (30.3)</td>
<td>1713 (31.4)</td>
</tr>
<tr>
<td>Incomplete secondary</td>
<td>164 (3.3)</td>
<td>171 (3.1)</td>
</tr>
<tr>
<td>Mother’s anthropometric measures</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Weight (kg)</td>
<td>66.39 ± 12.66²</td>
<td>66.21 ± 12.40</td>
</tr>
<tr>
<td>Height (cm)</td>
<td>164.44 ± 5.62</td>
<td>164.29 ± 5.65</td>
</tr>
<tr>
<td>Father’s education [%]</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Completed university</td>
<td>652 (12.9)</td>
<td>729 (13.4)</td>
</tr>
<tr>
<td>Advanced secondary or partial university</td>
<td>2440 (48.3)</td>
<td>2553 (46.8)</td>
</tr>
<tr>
<td>Common secondary</td>
<td>1857 (36.8)</td>
<td>2042 (37.5)</td>
</tr>
<tr>
<td>Incomplete secondary</td>
<td>103 (2.0)</td>
<td>126 (2.3)</td>
</tr>
<tr>
<td>Father’s anthropometric measures</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Weight (kg)</td>
<td>79.47 ± 11.52</td>
<td>79.88 ± 11.39</td>
</tr>
<tr>
<td>Height (cm)</td>
<td>175.97 ± 6.62</td>
<td>176.18 ± 6.61</td>
</tr>
<tr>
<td>Treatment arm [%]</td>
<td>2509 (49.7)</td>
<td>2701 (49.6)</td>
</tr>
</tbody>
</table>

² Mean ± SD (all such values).
DISCUSSION

We have used multilevel models to estimate growth trajectory information for each individual, and then used summary characteristics of these trajectories in a second model for the association of trajectory characteristics with the outcome. Individual-level estimates from multilevel growth models have been used in simple linear regression models to relate birth weight and childhood growth to outcomes, which include later obesity (19) and blood pressure (21). Individual trajectories can also be used to quantify other aspects of change. For example, multilevel models have been used to model BMI trajectories, and so derive individuals’ age and BMI at BMI peak, which were then related to BMI later in childhood (22). Estimates of individuals’ class membership from growth mixture models have also been related to risk factors and later outcomes (23).

Recent developments in methods for causal inference have clarified the problems that can arise when estimates of the effect of early exposure on an outcome are conditioned on later exposure (10, 24). For example, many studies of the relation between birth weight and adult blood pressure used regression models that also included adult weight (25). However, the correlation between birth weight and later weight can induce a negative relation between birth weight and adult blood pressure, even if the only true relation is between adult weight and blood pressure (26). Directed acyclic graphs can be used to help decide which variables should be included in the model that relates outcome to exposure (10). Path analysis may then be used to distinguish the direct effects of an exposure (those not mediated through any other measured variable) from indirect effects (20).

Multilevel models take longer to fit than simpler regression, although once derived, the exposures may be used for more than one outcome (19, 21). The added value from fitting such complex models is likely to depend on the number of growth measures to be used, how closely together the growth measures were obtained, and how many subjects are missing one or more measure, and how close to the planned clinic time each measure was actually taken. The

TABLE 3

Fit of estimated weight and length/height from multilevel spline models for weight and length/height with age for 10,502 children in the PROBIT (Promotion of Breastfeeding Intervention Trial) 1996–2003

<table>
<thead>
<tr>
<th></th>
<th>Weight</th>
<th>Length/height</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Observed minus predicted</td>
<td>95% limits of agreement</td>
</tr>
<tr>
<td></td>
<td>kg</td>
<td>cm</td>
</tr>
<tr>
<td>Birth</td>
<td>3.5 ± 0.42†</td>
<td>3.4 ± 0.32</td>
</tr>
<tr>
<td>3 mo</td>
<td>6.1 ± 0.66</td>
<td>6.3 ± 0.58</td>
</tr>
<tr>
<td>12 mo</td>
<td>10.6 ± 0.99</td>
<td>10.8 ± 0.93</td>
</tr>
<tr>
<td>6.5 y</td>
<td>22.8 ± 3.61</td>
<td>22.3 ± 2.93</td>
</tr>
</tbody>
</table>

† Mean ± SD (all such values).
larger the number of growth measures available, the greater will be the benefit in the decrease of dimensionality, and the more power there will be to fit the multilevel model. Measures taken at frequent time intervals (eg, 1, 2, 3, 4, 5, and 6 mo of age) are likely to be highly collinear, so a model that includes all of them as exposures would be unstable and thus would benefit from the use of multilevel models to decrease the number and collinearity of the exposures. If many subjects are missing one or more of the exposures, then the power for an analysis with the use of all exposures will be low, and thus the multilevel methods would be beneficial. Finally, where there is a wide spread of actual measurement times (eg, for a clinic planned for 2 mo of age, actual age varies between 1 and 3 mo), our methods will help decrease the associated measurement error.

Once multilevel models have been used to summarize growth trajectories, the usual regression methods can be used to relate these exposures to the outcomes of interest. These models will have to undergo the usual checks for validity of assumptions, and the usual issues of choice of confounding variables and probability of residual confounding will need to be considered. Here, we included parental heights and BMIs as confounders in the model; they could also be considered as proxies for the genetic components of weight and length/height. Inclusion of these confounding variables, plus highest maternal and paternal education levels, made little material difference to these models. However, there could be residual confounding by variables not measured here. Additionally, there is likely to be measurement error in maternal and paternal height and weight, with greater measurement error for paternal than for maternal measures, because the mother was usually the main informant. It may be that measured parental heights and weights would have a greater confounding effect than the self-reported heights and weights used here. However, given the lack of evidence for a confounding effect of self-reported measures, it seems unlikely that exact measures would have a greater effect. We have assumed here that the association between growth measures and later rash is linear: we examined this assumption with the use of categories of growth and observed no evidence of a nonlinear association (results available from the authors).

We have focused here on the development and application of methods to overcome some of the disadvantages of commonly used methods in the examination of evidence for an effect of fetal programming or childhood growth on later outcomes. The outcome used in this example, a rash that lasts for >6 mo, will include rashes with durations and causes that vary. This measurement error in the outcome is likely to be nondifferential between growth patterns and will thus bias estimates toward the null.

Here, multilevel models allowed us to decrease the dimensionality of the data from up to 13 measures per person to 5 summary measures for each of weight and length/height. We observed a small positive relation between weight gain between 12 and 34 mo and reported rash at age 6.5 y. This relation was not mediated through weight and height at 6.5 y. The summary of weight and length/height with the use of data-derived velocities can help identify relationships between early weight gain and length/height gain velocities and later outcomes. Multilevel models have several advantages over usual methods; they can access biological patterns of growth without restriction to measures at arbitrary ages or to individuals with complete data, and allow for within-subject variation and collinearity.

The authors’ responsibilities were as follows—RMM and YB-S: developed the hypotheses; MSK: contributed to obtaining funding for PROBIT fieldwork; NMD and EN: performed the statistical analysis under the supervision of KT; RMM: wrote the first draft; KT: coordinated completion of the article; and KT, RMM, NMD, and EN: had full access to all the data in the study and take responsibility for the integrity of the data and the accuracy of the data analysis. All authors critically commented on and edited earlier drafts of the article and approved the final version. The authors had no conflicts of interest.

REFERENCES