Early individual and family predictors of weight trajectories from early childhood to adolescence: results from the Millennium Cohort Study

Constança Soares dos Santos, MD\textsuperscript{a,b,c,*}, João Picoito, MD\textsuperscript{b,c,d}, Carla Nunes, PhD\textsuperscript{b,c}, Isabel Loureiro, MD, PhD\textsuperscript{b,c}

\textsuperscript{a}Department of Pediatrics, Centro Hospitalar Cova da Beira, Quinta do Alvito, 6200-251 Covilhã, Portugal.
\textsuperscript{b}Escola Nacional de Saúde Pública, Universidade NOVA de Lisboa, Avenida Padre Cruz, 1600-560 Lisbon, Portugal.
\textsuperscript{c}Centro de Investigação em Saúde Pública, Escola Nacional de Saúde Pública, Universidade NOVA de Lisboa, Avenida Padre Cruz, 1600-560 Lisbon, Portugal.
\textsuperscript{d}Department of Child and Adolescent Psychiatry, Centro Hospitalar e Universitário de Coimbra, Rua Doutor Afonso Romão, 3000-609 Coimbra, Portugal.

*Corresponding author: Constança Soares dos Santos
Email: csd.santos@ensp.unl.pt
Phone: +351 969285338
Postal address: Escola Nacional de Saúde Pública, Avenida Padre Cruz, 1600-560 Lisbon, Portugal.

João Picoito: jm.picoito@ensp.unl.pt
Carla Nunes: cnunes@ensp.unl.pt
Isabel Loureiro: isalou@ensp.unl.pt

Running title: Early childhood predictors of weight trajectories

Keywords: weight trajectories, early childhood, family determinants, Millennium Cohort Study, Growth Mixture Modeling

Word count (manuscript): 5000
Word count (abstract): 199
Number of tables and figures: 6
Abstract

Background: Early infancy and childhood are critical periods in the establishment of lifelong weight trajectories. Parents and early family environment have a strong effect on children’s health behaviors that track into adolescence, influencing lifelong risk of obesity.

Objective: We aimed to identify developmental trajectories of body mass index (BMI) from early childhood to adolescence and to assess their early individual and family predictors.

Methods: This was a secondary analysis of the Millennium Cohort Study and included 17,166 children. Weight trajectories were estimated using growth mixture modeling based on age- and gender-specific BMI Z-scores, followed by a bias-adjusted regression analysis.

Results: We found four BMI trajectories: Weight Loss (69%), Early Weight Gain (24%), Early Obesity (3.7%), and Late Weight Gain (3.3%). Weight trajectories were mainly settled by early adolescence. Lack of sleep and eating routines, low emotional self-regulation, child-parent conflict, and low child-parent closeness in early childhood were significantly associated with unhealthy weight trajectories, alongside poverty, low maternal education, maternal obesity, and prematurity.

Conclusions: Unhealthy BMI trajectories were defined in early and middle-childhood, and disproportionally affected children from disadvantaged families. This study further points out that household routines, self-regulation, and child-parent relationships are possible areas for family-based obesity prevention interventions.

Abbreviations in alphabetical order:

BIC – Schwarz’s Bayesian Information Criterion
BLRT – Bootstrapped likelihood ratio test
BMI – Body mass index
BMIZ – Bootstrapped likelihood ratio test
BMIZ – Body mass index
BMIZ – Age- and gender-specific BMI Z-scores
FIML – Full information maximum likelihood
GA – Gestational age
GMM – Growth Mixture Modeling
IOTF – International Obesity Task Force
LCGA – Latent Class Growth Analysis
MCS – Millennium Cohort Study
OR – Odds ratio
SD – Standard deviation
WHO – World Health Organization
Z-score – Standard deviation scores
Introduction

Obesity is a major Public Health issue worldwide. Over the last decades, obesity has been increasing at an alarming rate and appearing at progressively younger ages (1). In fact, between 1975 and 2016, among children and adolescents aged 5–19 years, the global prevalence of overweight has increased from 4% to 18% in girls and 19% in boys, and the global prevalence of obesity has risen from 0.7% to 5.6% in girls and from 0.9% to 7.8% in boys, summing up to 340 million children and adolescents (1). Besides, 41 million children under 5 years of age were estimated to suffer from overweight or obesity in 2016 (2).

In the short term, children with overweight and obesity not only have a higher risk of hypertension, diabetes, and sleep problems, they also have a higher risk of psychological distress such as negative body image, low self-esteem, depression, and peer problems (3,4). Furthermore, unhealthy weight tends to persist into adolescence and adulthood, increasing the lifelong risk of these non-communicable diseases (5).

Early infancy and childhood appear to be critical periods in the establishment of lifelong weight trajectories (6–8). Therefore, it is important to understand these trajectories and their early determinants in order to inform effective public health interventions.

From an ecological perspective, family plays a major role in every aspect of a child’s health and development, especially during infancy and early childhood (9). Both observational and experimental studies support the persistent effect of early family environment on health behaviors and weight status, highlighting the central role of parents in childhood obesity (10). Several studies suggest that general parenting, parenting styles and practices, and parent-child relationships can shape early eating, exercise, sleep, and screen use habits that track into adolescence (11,12). However, its influence on weight status is still debated.

An empirical way to investigate body mass index (BMI) trajectories over time is by using Growth Mixture Modeling (GMM), an extension of Growth Curve Modeling (13). GMM is a person-oriented approach that assumes that individuals do not belong to a single homogenous population but rather to distinct unobserved subpopulations with different developmental trajectories (14). GMM focuses on longitudinal change within each subpopulation and allows the classification of individuals into latent classes based on their growth trajectories.

Investigating BMI trajectories of over time requires large samples and accurate anthropometric measurements. Although several studies explored weight trajectories in childhood, some studies used categorical measures of overweight and obesity, which leads to classification bias, and others used “crude” BMI values, which do not account for growth (13). Since normal growth results in an expected increase in BMI, age- and gender-specific BMI standard deviation scores (Z-scores) are often considered the gold standard for the analysis of anthropometric data at an
epidemiological level (15). In fact, the Z-score has a linear scale that allows comparison between age groups and gender. Furthermore, it can be analyzed by summary statistics: the mean Z-score reflects the nutritional status of the entire population, and the standard deviation (SD) of the Z-score reflects the quality and accuracy of the data (15). On the other hand, most of these studies explored a small subset of covariates such as gender, ethnicity, and socioeconomic status, leaving out important factors like birthweight and breastfeeding duration that are well known to influence weight status, and even fewer studies have included early family environment, general parenting, and parenting practices.

Thus, the main objectives of our study were 1) to identify distinct subpopulations with different developmental trajectories of BMI in a general prospective cohort of British children, the Millennium Cohort Study; and 2) to examine the association between these trajectories and individual and family factors, focusing on those of early childhood.

In the literature, the most commonly found trajectory is the stable normative trajectory comprising the larger portion of the sample (13). While most studies report a decreasing BMI trajectory and an increasing BMI trajectory, other studies indicate an additional stable high trajectory (13). Therefore, we hypothesize 1) the existence of different weight trajectories, including a stable normative trajectory and a persistent high trajectory; 2) that individual and family factors influence the risk of belonging to a group trajectory; and 3) that an adverse early family context increases the risk of following an unhealthy weight trajectory.

Methods

Case Study

Data were drawn from the Millennium Cohort Study (MCS), a cohort study that follows children born between September 2000 and January 2002, and living and growing up in England, Scotland, Wales, and Northern Ireland. It also provides information about family circumstances and the broader socioeconomic context. MCS was designed to overrepresent specific subgroups of the population, namely children living in disadvantaged areas and those who are ethnic minorities. The sample is clustered by electoral wards stratified by country, ethnicity, and Child Poverty Index (16).

The study began with an original sample of 18,552 families, and at Sweep 2 (2003–04), it recruited 691 “new families” who were eligible but were missed at Sweep 1. Therefore, the total number of families ever interviewed comprises 19,243 families (19,517 children). Children were around 9 months at Sweep 1 and about 3, 5, 7, 11, and 14 years old at the subsequent sweeps. The study protocol meets the ethical requirements of the Helsinki Declaration, and it was approved by
the Northern and Yorkshire Research Ethics Committee (07/MRE03/32). Informed consent was obtained from parents or legal guardian before participation. Further information about the study design can be found in (16).

To the present study, singletons and the first-born child of the families with twins and triplets with valid data on BMI in at least one of Sweeps 2–6 were included, comprising 17,166 children.

**Measures**

**Anthropometric measures**

Weight and height were measured by trained interviewers using standardized instruments (Tanita HD-305 scales, Tanita UK Ltd; and portable stadiometers, Leicester Height Measure, Seca UK), with children wearing neither shoes nor outdoor clothes. Weight and height were used to calculate BMI (kg/m²).

We calculated age- and gender-specific BMI z-scores (BMIZ) using World Health Organization (WHO) Anthro (17) and Anthro Plus (18) software, having as reference the WHO Multicenter Growth Reference Study population. Observations with extreme values (below −5 SD or above +5 SD) were considered outliers and excluded(15). Four individuals presented BMIZ values only in the first two sweeps, with borderline BMIZ at age 3 (around −4 SD) corresponding to extreme thinness, and at age 5 in the overweight range (around +1.5 SD). Since there were no subsequent values to validate these observations, they were considered highly implausible and excluded.

**Covariates**

Predictors of nutritional status were selected based on previous research, according to Davidson and Birch’s ecological model of overweight (9). A conceptual framework of early life predictors of Overweight and Obesity is presented in Figure 1. Further details on how covariates were categorized and coded are presented in Table S1.

**Sociodemographic and economic covariates** were gender, ethnicity, family structure, family poverty, maternal age at birth of cohort member, maternal nutritional status, and maternal education measured at 9 months.

**Perinatal and early infancy covariates**

Gestational age (GA) was categorized as “extreme to moderate prematurity,” “late prematurity,” and “term,” according to the WHO/UNICEF definition (19).

Birthweight was categorized as “low birthweight,” “normal birthweight,” and “high birthweight,” according to the WHO definition (20).
Birthweight centiles for age and gender were calculated using Intergrowth21 software, and individuals were further classified in “Appropriate for GA,” “Small for GA,” and “Big for GA” (21).

Breastfeeding duration was assessed in Sweeps 1 and 2; it was re-categorized as in (22), with two more categories (“6 to 12 months” and “more than 12 months”), according to its distribution.

Introduction to solid food was classified as “early introduction” if it occurred before 4 months, and “late introduction” if it occurred after 6 months, according to European Society for Pediatric Gastroenterology Hepatology and Nutrition recommendations (23).

**Child temperament and self-regulation**

Child temperament was assessed at 9 months by 14 items from the Carey Infant Temperament Scale, capturing three dimensions: mood, adaptability/approach-withdrawal, and regularity. We created total scores for each dimension by summing the individual responses, as in (24). High scores on the first two dimensions indicated distress and withdrawal (α = .546 and .677, respectively), and high scores on the last dimension indicated regularity (α = .713).

Child self-regulation was assessed at 9 months by 10 items from the Child Social Behavior Questionnaire, with two dimensions: cognitive (α = .573) and emotional self-regulation (α = .632). Total scores for each dimension were calculated as the average of valid responses (25), so that higher scores in the first dimension indicated higher cognitive self-regulation, and higher scores in the second indicated emotional dysregulation.

**Household routines, parenting beliefs, and parenting activities**

Household routines were assessed at age 3, comprising sleep and feeding routines. A total sum score was created, with higher scores indicating consistent routines (α = .541).

Parenting beliefs were assessed at 9 months comprising 3 items derived from the European Longitudinal Study of Pregnancy and Childhood about the importance of stimulating, talking, and cuddling to a baby’s development. We summed the responses (26), so that higher scores indicated more positive parenting beliefs (α = .730).

Parenting activities were assessed at age 3 and comprised 5 items (reading, teaching the alphabet, teaching counting, teaching songs/rhymes, drawing). A total sum score was generated, so that higher scores indicated higher involvement in these activities (α = .610).

**Discipline practices and child-parent relationship**

Discipline practices were assessed at age 3 by 6 items from Straus’s Conflict Tactics Scale, measuring how often the mother used punishing (smack, shout, “tell off”) or withdrawal tactics (ignore, take away treats, send to bedroom/naughty chair) when the child misbehaved. We created
two scores: a harsh parenting score ($\alpha = .657$) by summing the responses in the punishing items, and a positive parenting score ($\alpha = .556$) by summing the responses in the withdrawal items (27).

**Parent-child relationship** was assessed at age 3 by Pianta Short Form, a 15-item self-administered scale based on attachment theory. This scale comprised two scores: closeness ($\alpha = .657$) and conflict ($\alpha = .787$).

**Statistical analysis**

Statistical analysis was performed using IBM Statistical Package for the Social Sciences, version 24.0 (SPSS Inc., Chicago, IL) and Mplus, version 8.3 (Muthén & Muthén 2017). Statistical significance was set to $p < 0.05$.

**Estimation of BMI trajectories**

Longitudinal BMI trajectories were analysed with GMM. GMM classifies children into latent classes based on their longitudinal change, so that individuals with similar BMI trajectories are assigned to the same class, and individuals in different classes follow significantly different BMI trajectories (14).

We tested multiple GMM models with different specifications before choosing the final model. First, we performed single-group models to identify the pattern (intercept only, linear, quadratic, cubic) that represented change over time the best. We also performed a latent basis model, where the pattern of change is not predefined but driven by the data, and loadings for the *slope* factor are estimated to represent the proportion of the total amount of growth that has occurred up to that point. Next, we performed GMM with *intercept* and *slope* variances fixed at zero for each class, a subtype of GMM called Latent Class Growth Analysis (LCGA), which assumes that all individual growth trajectories within a class are homogeneous (14). Then, we performed several GMMs with different growth factor variance specifications: first, with equal variances (homoscedastic model), and then freely estimated (heteroscedastic model). To determine the optimal number of classes, we started with a one-class solution and progressively increased the number of classes.

**Model selection**

The best model was chosen considering information criteria, theoretical justification, and interpretability (28,29). First, we looked at two model fit indices: Schwarz’s Bayesian Information Criterion (BIC) and the Bootstrapped Likelihood Ratio Test (BLRT) (28). BIC considers the likelihood of a model as well as the number of estimated parameters, with lower values of BIC indicating better fitting. BLRT compares a model with k classes to a model with k-1 classes, providing a $p$-value. We then looked at entropy, a measure of classification quality and separation between classes. Higher values of entropy (near 1) indicate more confidence in the
classification (30). Although values above 0.8 are considered good, there is no consensual definition of what constitutes low entropy (30). Finally, we considered the interpretability of the models and their practical meaningfulness, rejecting models with clinically uninterpretable classes and with classes representing less than 1% of the total sample.

**Association between BMI trajectories and early individual and family factors**

The association between covariates and class membership was calculated based on crude and adjusted odds ratio (OR) using a bias-adjusted three-step approach, which takes into account the classification error in class assignment (31).

**Missing values**

Missingness resulted from item non-response and attrition. Attrition is mainly related to sociodemographic characteristics, so the MCS study was designed to account for this bias and still provide representative information.

Missing values on BMI variables were handled with full information maximum likelihood (FIML) estimation under missing data theory. Considered the gold standard for handling missing data in latent variable indicators, FIML uses all available data points and is robust to non-normal distribution. (32)

Missing values on covariates were handled using multiple imputations carried out in Mplus. We used Bayesian estimation to create 25 imputed datasets, using the Markov chain Monte Carlo algorithm, and convergence criterion was set to 0.05. We included all covariates and also BMIz variables under the missing at random assumption, accounting for the complex sample design (32). Imputed values compare reasonably to those observed.

**Results**

**Body Mass Index**

Table 1 shows the summary statistics of BMIz across all sweeps and the nutritional status according to WHO BMIz cut-offs.

Comparing our results to those of the WHO Multicenter Growth Reference Study, at age 3 our study sample showed a slightly higher mean BMIz, implying an upward shift of our sample distribution, with a progressive decrease in subsequent sweeps. The BMIz SD superior to 1 indicated a slightly more dispersed distribution, but in all sweeps, it was below 1.3, suggesting good quality of the data. The overall prevalence of overweight and obesity decreased from 44.2% at age 3 to 29.9% at age 14.

**Covariates**
Sociodemographic characteristics of the analytical sample are presented in Table 2, and psychosocial covariates are presented in Table S2. Compared with children in the analytic sample, excluded individuals (who had no BMIzs in any of the sweeps) were more likely to be male, to be from ethnic minorities, to come from economically disadvantaged families and from monoparental families. Mothers of excluded children were younger and less educated. Complete bias analysis is available in Table S3.

BMI trajectory estimation

The single-group models that served as a basis for subsequent GMM are presented in Figure S2, showing that the quadratic and latent basis models appeared to better explain change in BMIz over time. Although the Quadratic GMM showed lower BICs, they provided uninterpretable trajectories; therefore, we decided to further analyze the latent basis GMM, presented in Table 3.

The LCGA showed much higher BICs than the other models, and since LCGA is based on the assumption that there is no within-class variability, we did not explore these models further.

Regarding the homoscedastic GMM, the 2-class model showed a big heterogeneous class (95.6%) and a small well-defined class (4.4%), corresponding to an Early Obesity trajectory class.

The 3-class model provided classes that were very disproportional (one big class comprising 91.9% and the two remaining classes comprising 4.3% and 3.8%). Furthermore, the two smaller classes were interpretable and meaningful, corresponding to an Early Obesity trajectory (4.3%) and a Late Weight Gain trajectory (3.8%); however, the bigger class (91.9%) was still heterogeneous, showing 20% of overweight and obese children across all sweeps, not corresponding to a normative class that we could use as a reference.

The 4-class model showed more proportional and meaningful classes; it provided an Early Obesity trajectory (3.7%) and a Late Weight Gain trajectory (3.3%) similar to the 3-class model, and Weight Loss (69.0%) and Early Weight Gain (24%) trajectories. The 5-class model showed an additional class that represented only less than 1% of the sample.

Regarding the heteroscedastic GMM models, they showed lower BIC, with BLRT favoring the 2-class model. This model provided a decreasing trajectory (77.7%) and an increasing trajectory (22.3%). However, this model showed low entropy (0.49) and provided poorly defined classes, while the decreasing trajectory still comprised around 20% of overweight and obese children at Sweep 6.

Therefore, we considered the 4-Class Latent Basis Homoscedastic GMM the best model to explain the change in BMIz.

BMI trajectory characterization
Figure 1 shows the BMIz trajectories from early childhood to adolescence. Latent growth factor means and variances are available in Table S4 for each class. The intercept and slope are inversely correlated ($r = -0.257, p < 0.001$), meaning that the higher the initial BMIz, the lower the growth. The time scores are 0, 0.326 ($p < 0.001$), 0.633 ($p < 0.001$), 0.924 ($p < 0.001$), and 1, meaning that individuals have reached 32.6% of the total change in BMIz at age 5, 63.3% at age 7, 92.4% at age 11, and 100% at age 14. Thus, the greatest amount of change occurs during early and middle-childhood, and there is little further change (7.6%) in BMIz in early adolescence.

Observing the BMI trajectories, two clearly distinct classes are seen: the Early Obesity and the Late Weight Gain. The Early Obesity class follows a clearly distinct trajectory across all sweeps, with a significantly higher mean BMIz at the starting point (mean intercept = 3.102, $p < 0.001$) than the other classes, that then decreases until the end of the study period (mean slope = -0.935, $p < 0.001$) but is still above the obesity cut-off. Although the Late Weight Gain class shows a lower starting point than the other classes (mean intercept = -0.716, $p < 0.001$), it then shows the greatest increase (mean slope = 1.889, $p < 0.001$) throughout early and middle-childhood, reaching the cut-off of overweight by age 11.

The other two bigger classes (Weight Loss and Early Weight Gain) have close starting points at age 3 (normal-high mean intercepts of 0.723 and 0.931, respectively; Wald test $p = 0.06$), but from then forward, they follow significantly opposite trajectories during childhood (Wald test $p < 0.001$): the Weight Loss Class steadily decreases (mean slope = -0.781), remaining in the normative BMIz range, while the Early Weight Gain increases (mean slope = 0.428) and plateaus in the overweight range.

The proportions of overweight and obese children in each class is represented in Figure 2.

**Association between BMI trajectories and individual and family covariates**

The association between BMI trajectories and individual and family covariates are shown in Figure 3, and unadjusted and adjusted OR are available in Table S5.

**Discussion**

We used GMM to capture the developmental change in BMI from early childhood to adolescence. We found four weight trajectories: Weight Loss (69%), Early Weight Gain (24%), Early Obesity (3.7%), and Late Weight Gain (3.3%). Using data from the MCS, a recent study applying LCGA to raw BMI data from 3 to 11 years also found four trajectories: Stable (83.8%), Moderate Increasing (13.1%), High Increasing (2.5%), and Decreasing (0.6%) (33). In addition, a similar study using the same period and that applied GMM instead of LCGA to raw BMI values found four trajectories, but with a significantly different interpretation: Low normal (boys, 49%; girls,
42%), Mid normal (boys, 36%; girls, 38%), Overweight (boys, 12%; girls, 16%), and Obesity (boys, 2%; girls, 3%) (34). In the latter study, there were no increasing or decreasing trajectories, meaning that children did not change to different weight categories during the study period (no mobility from normal to overweight/obesity categories or vice-versa) (34). Based on a methodologically different approach, our study further builds on the contrasting results of these previous studies, applying a homoscedastic latent basis GMM on age- and gender-specific BMIz to better capture BMI variation with growth and expanding the analysis from age 3 to age 14.

**BMI trajectories are mainly settled by early adolescence**

The greatest amount of change in BMIz occurred during early and middle-childhood, and there was little further change after age 11. In our study sample, by the time children enter school, they have reached one third of the total amount of change in their BMIz; further, 59.8% of the total amount of change occurs between ages 5 and 11, suggesting that early and middle-childhood are two different critical periods to intervene in weight trajectories. The importance of preschool years in weight gain has been established in other studies (6). In contrast, in one study using data from the Avon Longitudinal Study of Parents and Children (7) and in another in the USA (8), BMIzs steadily increased during childhood, with the greatest change occurring after school entry, suggesting that excess weight gain and obesity also develop in middle-childhood. In both studies, there was little further change in early adolescence.

**BMI trajectories are influenced by different early childhood factors**

The Early Obesity and Late Weight Gain classes follow clearly distinct and extreme trajectories across all sweeps and represent smaller groups of individuals (3.7% and 3.3%, respectively). The other two classes, Weight Loss and Early Weight Gain, have close starting points at age 3 and then follow less extreme but opposite trajectories during childhood, which represent the majority of individuals (69% and 24%, respectively). Although poverty, ethnic minority, single-parenthood, and low maternal education formed a common core of risk factors for unhealthy weight trajectories, we found different associations with early biological, psychological, and family factors.

**Early socioeconomic context**

In our study, children living in economically disadvantaged, ethnic, and single-parent families at 9 months of age were at greater risk of following unhealthy weight trajectories compared to their peers living in white advantaged families. Poverty has been consistently associated with obesity in both children and adults. Indeed, several studies support that early childhood poverty has an enduring association with obesity (35) and that its effect persists despite subsequent improvement in socioeconomic status (36).
However, in our study, this association disappeared after adjusting to other factors like maternal nutritional status and maternal education. In fact, maternal education appears to play a major role in childhood obesity (37). A large prospective study including data from 11 European cohorts concluded that low maternal education substantially increased the risk of early childhood adiposity in all countries (37). Therefore, maternal education might moderate the effect of poverty on obesity.

**Child factors**

**Gestational age, birthweight, breastfeeding, and complementary feeding**

In our study, children with high birthweight and who were big for GA were at greater risk for Early Weight Gain and Early Obesity. Interestingly, late prematurity was associated with the Early Weight Gain and Early Obesity trajectories, but extreme to moderate prematurity and being small for GA increased the risk for Late Weight Gain.

Early life context, including pregnancy, has a strong effect on later risk of obesity. Two cross-sectional studies of young adults found that those who had been born before 33 weeks GA had higher adiposity and cardiometabolic risk than those born at term (38,39). Moreover, it has been proposed that both nutrient overabundance and scarcity during pregnancy and infancy lead to a metabolic programming that results in an increased obesity risk throughout the lifespan (40). In fact, in a longitudinal cohort in Rotterdam, individuals with fetal growth restriction followed by infant weight acceleration had higher visceral and liver adiposity than those with normal fetal and infant growth (41).

In our study, being breastfed for more than 2 months lowered the risk of Early Weight Gain and Early Obesity, but not of Late Weight Gain, and early solid food introduction had an opposite effect. These associations were moderated by socioeconomic and biological factors. Breastfeeding has many demonstrated benefits for both mother and child (42). The mechanisms underlying the relationship between breastfeeding and obesity are still debated and include human milk polysaccharides, modulation of gut microbiota, and promotion of sensitive feeding and self-regulation (43,44). The protective effect of exclusive breastfeeding on rapid weight gain is seen mainly in early childhood, but its long-lasting effect has been debated (45). Other authors even argue that this relationship may stem from the more favorable socioeconomic and educational background of breastfeeding mothers (46).

**Child temperament and self-regulation**

In our study, emotional dysregulation in infancy was associated with Early Obesity, even after adjusting for socioeconomic and biological factors.
The effect of infant self-regulation on adolescent obesity has already been demonstrated in a study using the same MCS cohort (25). A growing body of evidence indicates that self-regulation has an important role in eating behavior, with children who show a lower capacity to self-soothe and to self-regulate being at greater risk of using food as a consolation tool and of obesity (47). The ability to regulate emotions begins in infancy and develops in the context of early mother-child reciprocal interactions. Breastfeeding has been shown to promote a mother’s sensitivity and capability of attributing mental states to their babies, and to predict more positive and sensitive behaviors during feeding at 12 months (44).

**Family factors**

*Household routines, parenting beliefs, and parenting activities*

In our study, children with consistent sleep and eating routines in early childhood were less likely to follow unhealthy weight trajectories, even after adjusting for socioeconomic and biological factors. The effect of household routines on obesity has already been consistently demonstrated in a study using the MCS and in the USA (25,48).

In our study, more positive parenting beliefs at age 9 months decreased the risk of Late Weight Gain, but this association disappeared after adjusting to other covariates. Although parental engagement in children's daily activities has been associated with lower obesity risk (49), we found no association between parenting activities and weight trajectories.

*Discipline practices and child-parent relationship*

Children who had a close relationship with their mothers in early childhood were less likely to follow the Late Weight Gain trajectory, and child-parent conflict in early childhood was associated with Early Obesity. There is a strong association between parent-child relationship and unhealthy eating and sedentary behaviors, factors well known to promote weight gain. A recent review highlights that a secure parent-child bond and high parental connectedness are associated with better eating and general health behaviors, while an insecure attachment and difficult parent-child relationship were associated with disordered eating and sedentary behaviors, mediated by temperament and self-regulation (50). Regarding weight status, in a longitudinal study, individuals with poor mother-child relationships in early infancy, assessed by maternal sensitivity and attachment, were at greater odds of obesity during adolescence (51).

In our study, positive parenting showed a small protective association with Early and Late Weight Gain that disappeared after adjusting to other factors. Harsh parenting has been demonstrated to increase the risk of obesity, while the appropriate use of family rules has been associated with a deceleration of obesity risk in adolescence (49,52).

**Strengths and limitations**
This study provided new evidence about the relationship between nutritional status trajectories and child and family factors. Its prospective longitudinal design allowed the exploration of BMI trajectories from early childhood to adolescence, and its large sample size allowed the exploration of multiple important individual and family covariates. To date, this is one of few studies that comprehensively included psychological factors like child temperament, self-regulation, parenting, and child-parent relationship alongside socioeconomic and biological factors.

Also, this study further explored the methodological aspects of investigating BMI trajectories. Anthropometric measures were directly collected according to the best standards, providing reliable and accurate data for subsequent analysis. Furthermore, trajectories were based on a continuous measure of BMI rather than on categorical measures of overweight and obesity. Especially, trajectories were based on BMI Z-scores, considered the gold standard for evaluating anthropometric measures. Moreover, it included 5 repeated measures of BMI over time and explored different GMM specifications.

Nevertheless, this study is not without limitations. We found it challenging to choose the best model. Although there is a common understanding on how model selection should be guided, there is some debate about the best indicators and their cut-offs. Furthermore, we opted for a homoscedastic GMM because the heteroscedastic GMM provided less substantive and poorly defined classes. Nevertheless, we must acknowledge that homoscedasticity might influence our results. Also, non-invariant models, which are more flexible and which have more parameters to estimate, tend to have superior information criteria at the expense of decreased interpretability and entropy, and one might question what would be the use of a better fitting model if it had lost its classification accuracy. Therefore, meaningfulness has a significant role in model selection. Although the entropy of our final model is not ideal, the subsequent regression analysis was adjusted to this classification bias.

Attrition is a major problem in longitudinal studies. GMM uses FIML to estimate the BMI trajectories, including all available data in the analysis, and therefore minimizes the bias effect of attrition.

Child and family covariates showed different percentages of missing values, and a complete case analysis would have substantially reduced the available sample; thus, we needed to impute missing values, which might influence our associations. Multiple imputation was performed according to best practice, and analysis in imputed and non-imputed data yielded concordant results. Although we recategorized covariates according to their distributions, previous research, and, whenever possible, international standards and recommendations, this recategorization might affect the association between BMI trajectories and covariates.
Conclusion

This study focused on BMI trajectories and looked at their early childhood predictors. In our study, 31% of the children followed an unhealthy weight trajectory that was mainly set by the time they reached early adolescence. Therefore, obesity prevention interventions should target children in early and middle-childhood, particularly those living in disadvantaged families. This study further points out that the lack of routines, low emotional self-regulation, low child-parent closeness, and child-parent conflict are significantly associated with unhealthy BMI trajectories, even after adjusting to other contextual factors; therefore, these are possible areas for family-based, health-promotion interventions. Further studies should focus on how different family and parenting factors interplay and influence weight trajectories, and on possible short- and long-term consequences on health status and well-being, following a developmental perspective.

Declarations

Conflicts of interest statement: The authors declare that they have no conflicts of interest to disclose. There is no funding source.

Acknowledgements: The Millennium Cohort Study (MCS), which began in 2000, is conducted by the Centre for Longitudinal Studies (CLS). It aims to chart the conditions of social, economic and health advantages and disadvantages facing children born at the start of the 21st century. The Principal Investigator and Director of the MCS is Prof. Emla Fitzsimons, UCL Institute of Education, University College London. For details, see https://cls.ucl.ac.uk/clt-studies/millennium-cohort-study/

The authors are grateful to the Centre for Longitudinal Studies (CLS), UCL Institute of Education, for the use of these data and to the UK Data Service for making them available, as well as to all the families who have participated in the MCS. However, neither CLS nor the UK Data Service bear any responsibility for the analysis or interpretation of these data.

Authors’ contributions: All authors contributed to the study conception and design. Data analysis were performed by CS. The first draft of the manuscript was written by CS and all authors commented on previous versions of the manuscript and revised it critically for important intellectual content. All authors read and approved the final manuscript and agree to be accountable for all aspects of the work in ensuring that questions related to the accuracy or integrity of any part of the work are appropriately investigated and resolved.

Data accessibility statement: The data that support the findings of this study are available through the UK Data Service [https://beta.ukdataservice.ac.uk/datacatalogue/series/series?id=2000031], but restrictions apply to the availability of these data, which were used under Special Licence for the current study. Data are available through the UK Data Service after approval by the CLS Data Access Committee.


Data references


<table>
<thead>
<tr>
<th>Sweep</th>
<th>Age 3</th>
<th>Age 5</th>
<th>Age 7</th>
<th>Age 11</th>
<th>Age 14</th>
</tr>
</thead>
<tbody>
<tr>
<td>n</td>
<td>14176</td>
<td>14922</td>
<td>13603</td>
<td>12864</td>
<td>10958</td>
</tr>
<tr>
<td>Mean (SD)</td>
<td>0,88 (1,04)</td>
<td>0,62 (1,02)</td>
<td>0,48 (1,14)</td>
<td>0,53 (1,23)</td>
<td>0,39 (1,20)</td>
</tr>
<tr>
<td>Min</td>
<td>-4,35</td>
<td>-4,60</td>
<td>-4,94</td>
<td>-4,54</td>
<td>-4,51</td>
</tr>
<tr>
<td>Max</td>
<td>5,00</td>
<td>4,98</td>
<td>4,96</td>
<td>4,76</td>
<td>4,58</td>
</tr>
<tr>
<td>Range</td>
<td>9,35</td>
<td>9,58</td>
<td>9,90</td>
<td>9,30</td>
<td>9,09</td>
</tr>
<tr>
<td>Severe thinness</td>
<td>17 (0,1)</td>
<td>16 (0,1)</td>
<td>26 (0,2)</td>
<td>30 (0,2)</td>
<td>37 (0,3)</td>
</tr>
<tr>
<td>Thinness</td>
<td>61 (0,4)</td>
<td>55 (0,4)</td>
<td>101 (0,7)</td>
<td>200 (1,6)</td>
<td>208 (1,9)</td>
</tr>
<tr>
<td>Normal weight</td>
<td>7828 (55,2)</td>
<td>10160 (68,1)</td>
<td>9655 (71,0)</td>
<td>8018 (62,3)</td>
<td>7434 (67,8)</td>
</tr>
<tr>
<td>Overweight</td>
<td>4511 (31,8)</td>
<td>3437 (23,0)</td>
<td>2522 (18,5)</td>
<td>2931 (22,8)</td>
<td>2180 (19,9)</td>
</tr>
<tr>
<td>Obesity</td>
<td>1759 (12,4)</td>
<td>1254 (8,4)</td>
<td>1299 (9,5)</td>
<td>1685 (13,1)</td>
<td>1099 (10)</td>
</tr>
</tbody>
</table>

% excluded extreme values (< -5 SD or > +5 SD): at S2 - 1%; at S3 – 0,4%; at S4 – 0,2%; at S5 – 0,05 %; at S6 – 0,01%

SD = Standard deviations
### Table 2 – Characteristics of the analytical sample (n= 17166)

<table>
<thead>
<tr>
<th>Characteristic</th>
<th>n</th>
<th>n (%)</th>
<th>Missing (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Gender</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Male</td>
<td>8775</td>
<td>(51,1)</td>
<td>0</td>
</tr>
<tr>
<td>Female</td>
<td>8391</td>
<td>(48,9)</td>
<td>0</td>
</tr>
<tr>
<td><strong>Ethnicity</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>White</td>
<td>1405</td>
<td>(82,3)</td>
<td>0.6</td>
</tr>
<tr>
<td>Other</td>
<td>3016</td>
<td>(17,7)</td>
<td>0</td>
</tr>
<tr>
<td><strong>Gestational age</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Extreme to moderate Prematurity</td>
<td>319</td>
<td>(2,0)</td>
<td>5.2</td>
</tr>
<tr>
<td>Late Prematurity</td>
<td>892</td>
<td>(5,5)</td>
<td></td>
</tr>
<tr>
<td>Term</td>
<td>15066</td>
<td>(92,6)</td>
<td>0</td>
</tr>
<tr>
<td><strong>Birthweight</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Low</td>
<td>1213</td>
<td>(7,1)</td>
<td>0.7</td>
</tr>
<tr>
<td>Normal</td>
<td>14010</td>
<td>(82,2)</td>
<td></td>
</tr>
<tr>
<td>High</td>
<td>1816</td>
<td>(10,7)</td>
<td></td>
</tr>
<tr>
<td><strong>Birthweight for gestational age</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Small</td>
<td>1237</td>
<td>(7,8)</td>
<td>7</td>
</tr>
<tr>
<td>Appropriate</td>
<td>11884</td>
<td>(74,5)</td>
<td></td>
</tr>
<tr>
<td>Large</td>
<td>2838</td>
<td>(17,8)</td>
<td></td>
</tr>
<tr>
<td><strong>Breastfeeding duration</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Never</td>
<td>5706</td>
<td>(33,4)</td>
<td>0.5</td>
</tr>
<tr>
<td>&lt;=2 months</td>
<td>4951</td>
<td>(29)</td>
<td></td>
</tr>
<tr>
<td>2 to 4 months</td>
<td>2067</td>
<td>(12,1)</td>
<td></td>
</tr>
<tr>
<td>4 to 6 months</td>
<td>1270</td>
<td>(7,4)</td>
<td></td>
</tr>
<tr>
<td>6 to 12 months</td>
<td>1668</td>
<td>(9,8)</td>
<td></td>
</tr>
<tr>
<td>&gt;=12 months</td>
<td>1413</td>
<td>(8,3)</td>
<td></td>
</tr>
<tr>
<td><strong>Age at complementary feeding</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Before 4 months</td>
<td>5146</td>
<td>(31,2)</td>
<td>4.1</td>
</tr>
<tr>
<td>Between 4 and 6 months</td>
<td>10836</td>
<td>(65,8)</td>
<td></td>
</tr>
<tr>
<td>After 6 months</td>
<td>486</td>
<td>(3)</td>
<td></td>
</tr>
<tr>
<td><strong>Family structure (Sweep 1)</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Two parents/carers</td>
<td>13766</td>
<td>(83,5)</td>
<td>4</td>
</tr>
<tr>
<td>Single parenthood (One parent/carer)</td>
<td>2720</td>
<td>(16,5)</td>
<td></td>
</tr>
<tr>
<td><strong>OECD Median Poverty Indicator (Sweep 1)</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Above 60% median</td>
<td>10528</td>
<td>(64,1)</td>
<td>4.3</td>
</tr>
<tr>
<td>Below 60% median</td>
<td>5901</td>
<td>(35,9)</td>
<td></td>
</tr>
<tr>
<td><strong>Mother age at cohort member’s birth</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>12 to 19 years</td>
<td>1476</td>
<td>(8,6)</td>
<td>0.4</td>
</tr>
<tr>
<td>20 to 39 years</td>
<td>15257</td>
<td>(89,2)</td>
<td></td>
</tr>
<tr>
<td>40 plus</td>
<td>368</td>
<td>(2,2)</td>
<td></td>
</tr>
<tr>
<td><strong>Mother nutritional status (Sweep 1)</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Underweight</td>
<td>1825</td>
<td>(12,8)</td>
<td>16.7</td>
</tr>
<tr>
<td>Normal weight</td>
<td>6872</td>
<td>(48,1)</td>
<td></td>
</tr>
<tr>
<td>Overweight</td>
<td>3721</td>
<td>(26)</td>
<td></td>
</tr>
<tr>
<td>Obesity</td>
<td>1879</td>
<td>(13,1)</td>
<td></td>
</tr>
<tr>
<td><strong>Maternal education (Sweep 1)</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Low (NVQ&lt;3 / none/ overseas)</td>
<td>11492</td>
<td>(69,9)</td>
<td>4.3</td>
</tr>
<tr>
<td>High (NVQ=&gt;4)</td>
<td>4940</td>
<td>(30,1)</td>
<td></td>
</tr>
</tbody>
</table>

NVQ – National Vocational Qualifications

Data are given as n (%)
Table 3 - Growth Mixture Models

<table>
<thead>
<tr>
<th>No of classes</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
</tr>
</thead>
<tbody>
<tr>
<td>Log-likelihood</td>
<td>-101683.801</td>
<td>-90860.327</td>
<td>-85782.504</td>
<td>-83534.897</td>
<td></td>
</tr>
<tr>
<td>AIC</td>
<td>203387.602</td>
<td>181746.654</td>
<td>171597.007</td>
<td>167107.794</td>
<td></td>
</tr>
<tr>
<td>BIC</td>
<td>203465.108</td>
<td>181847.412</td>
<td>171721.017</td>
<td>167255.056</td>
<td></td>
</tr>
<tr>
<td>aBIC</td>
<td>203433.329</td>
<td>181806.099</td>
<td>171670.170</td>
<td>167194.675</td>
<td></td>
</tr>
<tr>
<td>Entropy</td>
<td>0.74</td>
<td>0.78</td>
<td>0.77</td>
<td>1.000</td>
<td></td>
</tr>
<tr>
<td>BLRT p-value</td>
<td>&lt;0.001</td>
<td>&lt;0.001</td>
<td>&lt;0.001</td>
<td>1.000</td>
<td></td>
</tr>
<tr>
<td>Parameters</td>
<td>10</td>
<td>13</td>
<td>16</td>
<td>19</td>
<td></td>
</tr>
<tr>
<td>n</td>
<td>17166</td>
<td>10352/6813</td>
<td>2895/4803/9467</td>
<td>2484/1393/5195/8093</td>
<td></td>
</tr>
</tbody>
</table>

Latent Basis Homoscedastic Growth Mixture Model
growth factor variances equal between classes

<table>
<thead>
<tr>
<th>No of classes</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
</tr>
</thead>
<tbody>
<tr>
<td>Log-likelihood</td>
<td>-79740.295</td>
<td>-79489.756</td>
<td>-79363.766</td>
<td>-79334.793</td>
<td>-79306.347</td>
</tr>
<tr>
<td>AIC</td>
<td>159506.591</td>
<td>159011.511</td>
<td>158765.531</td>
<td>158713.586</td>
<td>158662.694</td>
</tr>
<tr>
<td>BIC</td>
<td>159607.349</td>
<td>159135.521</td>
<td>158912.793</td>
<td>158884.100</td>
<td>158856.460</td>
</tr>
<tr>
<td>aBIC</td>
<td>159566.035</td>
<td>159084.674</td>
<td>158852.413</td>
<td>158814.185</td>
<td>158777.011</td>
</tr>
<tr>
<td>Entropy</td>
<td>0.88</td>
<td>0.82</td>
<td>0.60</td>
<td>0.65</td>
<td></td>
</tr>
<tr>
<td>BLRT p-value</td>
<td>&lt;0.001</td>
<td>&lt;0.001</td>
<td>&lt;0.001</td>
<td>&lt;0.001</td>
<td></td>
</tr>
<tr>
<td>Parameters</td>
<td>13</td>
<td>16</td>
<td>19</td>
<td>22</td>
<td></td>
</tr>
<tr>
<td>n</td>
<td>17166</td>
<td>16682/476</td>
<td>519/408/16238</td>
<td>12600/3657/505/403</td>
<td>12728/3539/510/357/31</td>
</tr>
</tbody>
</table>

Latent Basis Heteroscedastic Growth Mixture Model
growth factor variances variant between classes

<table>
<thead>
<tr>
<th>No of classes</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
</tr>
</thead>
<tbody>
<tr>
<td>Log-likelihood</td>
<td>-79740.295</td>
<td>-79341.995</td>
<td>-79272.012</td>
<td>-79248.190</td>
<td></td>
</tr>
<tr>
<td>AIC</td>
<td>159506.591</td>
<td>158721.990</td>
<td>158594.023</td>
<td>158558.380</td>
<td></td>
</tr>
<tr>
<td>BIC</td>
<td>159607.349</td>
<td>158869.252</td>
<td>158787.789</td>
<td>158798.649</td>
<td></td>
</tr>
<tr>
<td>aBIC</td>
<td>159566.035</td>
<td>158808.871</td>
<td>158708.341</td>
<td>158700.133</td>
<td></td>
</tr>
<tr>
<td>Entropy</td>
<td>0.49</td>
<td>0.48</td>
<td>0.46</td>
<td></td>
<td></td>
</tr>
<tr>
<td>BLRT p-value</td>
<td>0.0000</td>
<td>1.0000</td>
<td>1.0000</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Parameters</td>
<td>13</td>
<td>19</td>
<td>25</td>
<td>31</td>
<td></td>
</tr>
<tr>
<td>n</td>
<td>17166</td>
<td>14482/2683</td>
<td>13870/2943/352</td>
<td>12981/2567/915/702</td>
<td></td>
</tr>
</tbody>
</table>

BLRT = Bootstrapped Log-Likelihood Ratio Test
All Growth Mixture Models were estimated without covariates using MPlus, version 8.3 (ANALYSIS = MIXTURE; ALGORITHM = INTEGRATION). The number of random starts was increased to 2000 with 500 final stage optimizations (STARTS = 2000 500). To compute the BLRT, the number of random starts for the k class was increased to 200 with 40 final stage optimizations (LRTSTARTS = 0 0 200 40). Then, we repeated the analysis accounting for the complex sample design (ANALYSIS = MIXTURE COMPLEX).
Figure 1 - BMI z-score trajectories

- **Obesity**
- **Overweight**

- **Weight Loss (69%)**
- **Early Weight Gain (24%)**
- **Early Obesity (3.7%)**
- **Late Weight Gain (3.3%)**
Figure 2 - Class characterization – nutritional status

- Weight Loss Class (69%)
- Early Weight Gain Class (24%)
- Early Obesity Class (3.7%)
- Late Weight Gain Class (3.3%)

The diagrams show the distribution of weight status across different age groups (3, 5, 7, 11, 14 years) for each class. The colors represent:
- Green: Normal weight
- Yellow: Overweight
- Red: Obesity
Figure 3 – Early Life Predictors of BMI trajectories

a) **Undeajusted OR (Reference Class: Weight Loss)**
b) Adjusted OR (Reference Class: Weigh Loss)

CSBQ – Child Social Behavioural Questionnaire; GA – Gestational Age; PT - prematurity