BMJ Open

Relative importance of prenatal and postnatal determinants of stunting: data mining approaches to the MINIMat cohort, Bangladesh

Pernilla Svefors,[1,2] Oleg Sysoev,[3] Eva-Charlotte Ekstrom,[1] Lars Ake Persson,[4] Shams E Arifeen,[5] Ruchira T Naved,[6] Anisur Rahman,[5] Ashraful Islam Khan,[7] Katarina Selling[1]

For numbered affiliations see end of article.

Correspondence to
Dr Pernilla Svefors;
pernilla.svefors@kbh.uu.se

ABSTRACT

Introduction WHO has set a goal to reduce the prevalence of stunted child growth by 40% by the year 2025. To reach this goal, it is imperative to establish the relative importance of risk factors for stunting to deliver appropriate interventions. Currently, most interventions take place in late infancy and early childhood. This study aimed to identify the most critical prenatal and postnatal determinants of linear growth 0–24 months and the risk factors for stunting at 2 years, and to identify subgroups with different growth trajectories and levels of stunting at 2 years.

Methods Conditional inference tree-based methods were applied to the extensive Maternal and Infant Nutrition Interventions in Matlab trial database with 309 variables of 2723 children, their parents and living conditions, including socioeconomic, nutritional and other biological characteristics of the parents; maternal exposure to violence; household food security; breast and complementary feeding; and measurements of morbidity of the mothers during pregnancy and repeatedly of their children up to 24 months of age. Child anthropometry was measured monthly from birth to 12 months, thereafter quarterly to 24 months.

Results Birth length and weight were the most critical factors for linear growth 0–24 months and stunting at 2 years, followed by maternal anthropometry and parental education. Conditions after birth, such as feeding practices and morbidity, were less strongly associated with linear growth trajectories and stunting at 2 years.

Conclusion The results of this study emphasise the benefit of interventions before conception and during pregnancy to reach a substantial reduction in stunting.

Strengths and limitations of this study

► Includes high-quality longitudinal data with low rates of missing data on child growth and a wide range of prenatal and postnatal household, family and environmental factors, child characteristics at birth, infant feeding and morbidity.

► Employs decision tree-based methods that permit the inclusion of a high number of predictor variables, variables of different types and automatically discover complex interactions between predictor variables and include them in the model.

► Some potentially important determinants of linear growth were not present in the database.

► The study does not include stratified analyses for girls and boys.

adult chronic diseases.[3] In 2012, WHO adopted a resolution on maternal and child undernutrition, targeting a reduction of stunting by 40% by 2025.[4] Linear growth is most susceptible to environmentally modifiable factors from conception up to 2 years of age, that is, the first 1000 days when most of the growth faltering takes place.[5 6] To develop and deliver appropriate interventions, it is imperative to establish the relative importance of stunting risk factors. In addition, the sustainable development health goal has emphasised the personalised perspective under the universal coverage of healthcare. Precision public health interventions by identifying and targeting high-risk subgroups can be one of the strategies to reach this goal.[7]

Previous studies employing classical statistical methods have identified a wide range of prenatal and postnatal factors associated with impaired growth.[8–13] Low birth weight, maternal height, maternal education, poverty and inadequate complementary feeding practices have been recognised as important risk

INTRODUCTION

Linear growth is considered to be the best overall indicator of children's present and future health[1 2] and the reduction of growth failure is one of the targets within the sustainable development agenda. Stunted growth is associated with short-term morbidity and mortality, impaired cognitive development, lower future productivity and increased risk of

factors.[14–16] Some analyses emphasise the importance of fetal growth restriction for later stunted growth, but rarely is the relative importance of prenatal and postnatal factors assessed.[17] Despite these findings, policy documents and recommendations emphasise interventions especially after birth, and prenatal recommendations are usually limited to routine micronutrient supplementation for pregnant women.[18–20]

Despite a wealth of literature relating to the determinants of stunting, studies with a holistic approach, which concurrently account for household, environmental, nutritional, biological and socioeconomic influences, are few. Moreover, individuals and groups may be stunted for various reasons and thus respond differently to interventions. Studies that identify risk groups with different probabilities of stunting are, to the best of our knowledge, not yet available. The available studies with a multifactorial approach have frequently had a cross-sectional design and have applied traditional statistical methods. As visualised in WHO's conceptual framework on childhood stunting,[21] the causes of stunted linear growth are complex. The number of risk factors and the complexity of the associations of these risk factors with linear growth restriction make traditional statistical models ineffective from a predictive perspective. Moreover, classical statistical methods do not have the capacity to identify groups with different risks based on the combinations of predictors. Decision trees are popular data mining (DM) methods, which allows for the inclusion of a high number of predictor variables, handling variables of different types, automatically discovering complex interactions between predictor variables and including them in the model.[22] Decision tree-based algorithms can be used to rank a high number of predictors according to their relative importance for the outcome and to identify subgroups with different risk patterns.

The Maternal and Infant Nutrition Interventions in Matlab (MINIMat) was a randomised prenatal food and multiple micronutrient trial carried out in rural Bangladesh. The frequent follow-up of mothers and children participating in this trial resulted in an extensive database, including frequent prenatal and postnatal anthropometric assessments, socioeconomic and biological characteristics of the mother and father, information on maternal exposure to violence, household food security, breastfeeding and infant-feeding practices, and measurement of morbidity of the mothers during pregnancy and repeatedly of children up to 24 months of age. The aim of this study is to, within this Bangladeshi cohort, assess the relative importance of determinants of linear growth from 0 to 24 months and risk factors for stunting at 2 years, and to identify risk groups with negative growth trajectories and high prevalence of stunting at 2 years.

METHODS
Study setting, participants and study design
The MINIMat trial (isrctn.org identifier: ISRCTN16581394) was carried out in Matlab, Bangladesh,

a rural delta region located 57 km southeast of the capital Dhaka. In this area, a health and demographic surveillance system enables early pregnancy identification and longitudinal follow-up. Pregnant women were enrolled in the MINIMat trial and the follow-up included their offspring. MINIMat was a factorial randomised trial primarily evaluating the effect of an early invitation to prenatal food supplementation (vs usual timing) combined with multiple micronutrient supplementation (vs usual programme iron-folate) to pregnant women on maternal haemoglobin, birth weight, gestational age at birth and infant mortality.[23] Further, the participating women were randomly assigned to either counselling for exclusive breast feeding or a different health education message of equivalent intensity.[24] The MINIMat trial recruited pregnant women from November 2001 to October 2003. When a woman reported to a community health worker that her menstruation was delayed by more than 14 days, she was offered a pregnancy test and her date for the last menstrual period (LMP) was recorded. If LMP date was missing, the gestational age assessment was based on ultrasound examination. In total, 4436 pregnant women participated, giving birth to 3625 live-born infants from April 2002 to June 2004. The pregnant women were enrolled at around gestational week 8. In this analysis, the mothers and children were followed through pregnancy, birth and up to 2 years of age.

Written and oral informed consent was obtained from all participating women and from the parents of the participating children. The Ethical Review Committee at the International Centre for Diarrhoeal Disease Research, Bangladesh, approved the study (approval registration numbers 2000-025; 2002-031; 2005-004).

Data collection
Predictor and outcome variables are presented in figure 1, grouped according to WHO conceptual framework of stunting.[21] Data were collected using questionnaires, physical examinations and laboratory analyses. At enrolment, well-trained field workers collected information on women's age, parity, marital status, educational level, occupation, maternal morbidity, socioeconomic characteristics and household food security. Socioeconomic status was assessed based on a range of household assets, and a continuous household asset score, with a mean value of zero, was constructed based on a principal component analysis.[25] A validated household food security scale was created from 11 items with data on frequency of food purchased, cooked, borrowed or lent (food and money), and whether there was ready access to adequate meals and snacks.[26] The participating women were also asked whether they had suffered any of 30 morbidity symptoms from 12 different categories, including airway, urinary tract, fever, circulation, bowel or pain symptoms during the last month. A sum score ranging from 0 to 12 was created based on the absence of symptoms or those not recorded for each category.
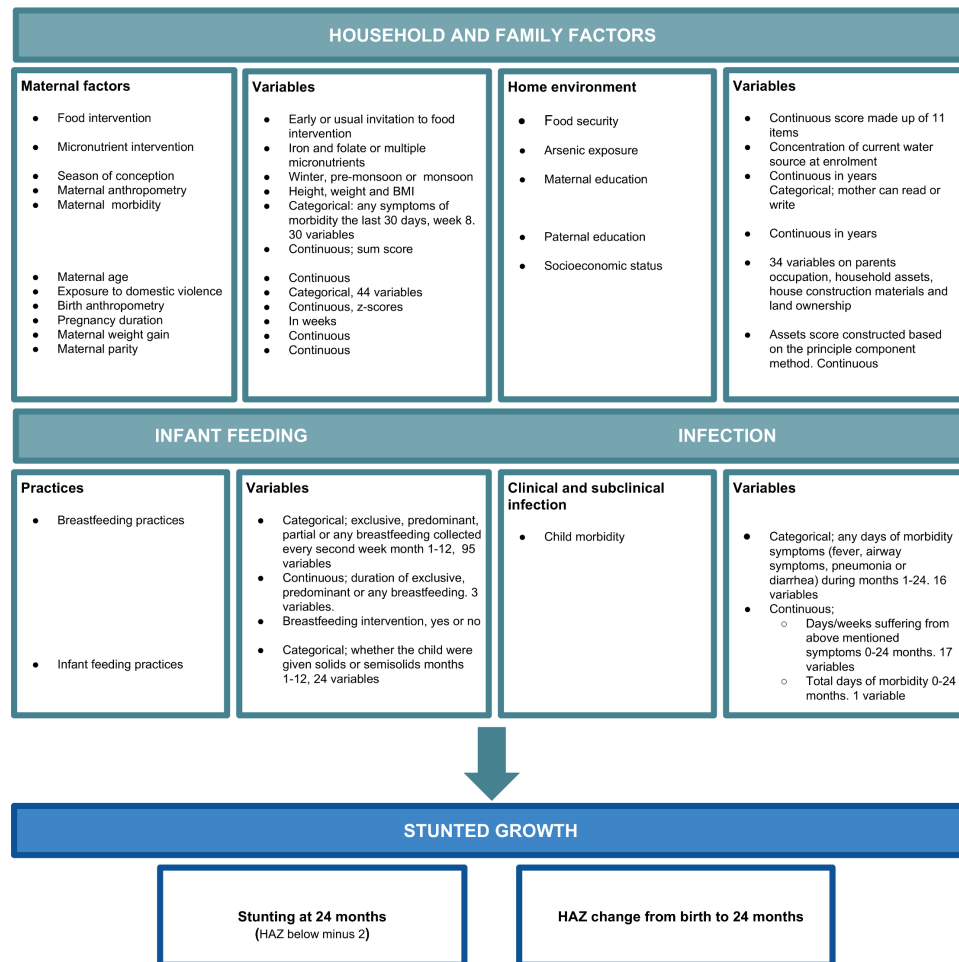
Figure 1 Factors, variables and outcomes included in the analysis of data from the Maternal and Infant Nutrition Interventions in Matlab cohort, Bangladesh. grouping according to WHO conceptual framework on childhood stunting.[21] BMI, body mass index; HAZ, height-for-age z-scores.

Home visits were followed by clinic visits at local health subcentres. Maternal height and weight were measured at around 8 weeks of gestation using a stadiometer to the nearest 0.1 cm and an electronic scale (Uniscale, SECA, Hamburg, Germany) with a precision of 0.10 kg. In the third trimester, paramedics interviewed the participating women in privacy regarding their experiences of domestic violence. A modified version of WHO collaborative study questionnaire was used,[27 28] based on the Conflict Tactic Scale covering physical, sexual and emotional violence and controlling behaviour.[29] Household drinking water was analysed for arsenic concentration.[30]

A birth notification system allowed birth anthropometry to be measured within 72 hours. In the few cases where the newborns were reached after 72 hours, the measurements were adjusted to the time of birth using an SD score transformation, assuming that the infants remained in the same relative position in the anthropometric distribution during this period.[31] At birth, data on sex, birth weight, length and breastfeeding practices were collected. During the subsequent 2-year study period, the mother-and-child pairs were visited monthly in their homes during the first year, and every 3 months during the second year. On these occasions, data on infant feeding practices, child morbidity and anthropometry were collected. The mothers were interviewed about breast feeding and complementary feeding practices. Breastfeeding practices were categorised into exclusive, predominant, partial or any breast feeding for each month from 1 to 12 months. The total time for exclusive, predominant and any breast feeding was calculated. WHO recommendations guided the breastfeeding assessment[32] and results were validated with a stable-isotope technique. The classification of exclusive breast feeding was found to suffer from limited misclassification in both directions and to be accurate at the group level.[33] The food given to the infant was categorised into semisolids and solids each month from 1 to 12 months. The data collection did not include full dietary assessments or classification of dietary diversity and meal frequency.

The mothers were also asked whether the child had any of the following symptoms during the last week: fever, cough, difficult breathing, chest in-drawing, rapid breathing, diarrhoea, bloody diarrhoea and the duration of these symptoms.[34] Categories were created based on whether the child had suffered from fever, respiratory

symptoms, suspected pneumonia or diarrhoea, and the sum of days with each symptom and total morbidity calculated from birth to 24 months. To reduce the risk of recall bias, the mothers were visited monthly with an interview recall period of 7 days for child morbidity. One week has been found to be optimal for this kind of morbidity recall assessment.[35]

Children's weight was measured by SECA beam and electronic scales (UNICEF Uniscale, SECA Gmbh & Co) with a precision of 0.01 kg. The length at birth and up to 1.5 years was measured with a collapsible, locally manufactured length board with a precision of 0.1 cm. From 1.5 to 2 years, height was measured to the nearest 0.1 cm, using a freestanding stadiometer. Head and chest circumference was measured with a measuring tape. Two measurements were recorded on each occasion and the mean was calculated. The equipment was calibrated daily and refresher training on data collection methods, including the standardisation of anthropometric measurements, was conducted periodically.

Outcomes

Height-for-age z-scores (HAZ) were calculated from the measured length and height data using the programme WHOAnthro, based on WHO growth reference for children.[36] Children with a HAZ below minus two SD-scores were classified as stunted. Two outcomes were analysed: stunting at 24 months and the change in HAZ from birth to 24 months, referred to as Δ HAZ and calculated by subtracting HAZ at birth from HAZ at 24 months, that is, ΔHAZ=HAZ at 24 months – HAZ at birth.

Statistical analysis

A database was created with 309 variables characterising mothers and children in the MINIMat cohort from enrolment in early pregnancy up to the time when the children were 24 months of age. The subset of records that had height measurements at birth and 24 months was selected (n=2723). The average per cent of missing values among all the predictors was 4%. The highest per cent missing was among maternal morbidity data during pregnancy (22%) and categorical monthly child morbidity data (ill or not), ranging from 0% to 35% with the highest number of missing observations in the first months. The continuous child morbidity data however (sum of days with different types illnesses) had no missing values. The most important variables identified by the random forest analyses and the variables included by the conditional inference trees (CIT) had less than 1% missing values. The missing values of the predictor variables were imputed. To find the best method to impute the missing data, we made a simulation study of the performance of the following imputation methods: imputation by variable mean, K-nearest neighbour imputation[37] and random forest imputation.[38] The design of the study followed a procedure similar to the strategy described in Jönsson and Wohlin[37]; see online supplementary appendix. Accordingly, we imputed the data by use of the random

forest as the simulation study revealed that this method provided the most accurate imputations.

Decision trees[22] are DM methods that allow for specifying an arbitrarily high number of predictor variables, handle variables of different types, automatically discover complex interactions between predictor variables and include them in the model. Traditional decision trees, such as classification and regression trees, have been shown to be biassed.[39] This motivated us to select the CIT framework, a method that embeds a statistical hypothesis-testing framework into a recursive partitioning algorithm used for model building.[39] CIT were used in order to identify subgroups characterised by combinations of levels of certain predictors with distinct values of Δ HAZ or prevalence of stunting at 24 months. Cross validation, a well-established model selection method that selects a tree with an optimal predictive performance for new unseen data, was applied. Cross validation splits the dataset into different train and test sets repeatedly, estimates the model in one set and validates the prediction on another set, followed by an aggregation of the predictions.[40] To ensure public health relevance, the minimum number of observations in each terminal node (subgroup) was set to 250.

Conditional random forest (CRF) analyses were performed to assess and rank the importance of predictors with regard to their ability to explain the variation of the continuous outcome of the change in HAZ from birth to 24 months and the presence of stunting at 24 months of age. In conditional random forest analysis, an ensemble of CIT is created by means of drawing subsamples from the original data and fitting a unique randomised CIT to each sample. Possible predictors at each split are selected randomly from the complete set of predictors, which leads to a better predictive performance of the tree ensemble.[40] The importance of a variable is computed by comparing the predictive mean squared error (MSE) from the original data and a dataset where the corresponding variable values are specified incorrectly, which makes the variable irrelevant for the prediction. If the variable does not contribute to the prediction, the MSE is expected to be small when the values of the variable are permutated. An aggregated difference between the MSE values over the given ensemble of trees makes up the relative importance of a variable. The random forests analyses were created based on 3000 trees, and the 30 variables with the highest importance measure are presented. The programming language R V.3.2.4[41] and the 'party' package[42] were used for all analyses.

Patient and public involvement

No participants were involved in developing the hypothesis, the specific aims or the research questions, nor were they involved in developing plans for design or implementation of the study. No participants were involved in the interpretation of study results or write up of the manuscript. There are no plans to disseminate the results of the research to study participants.
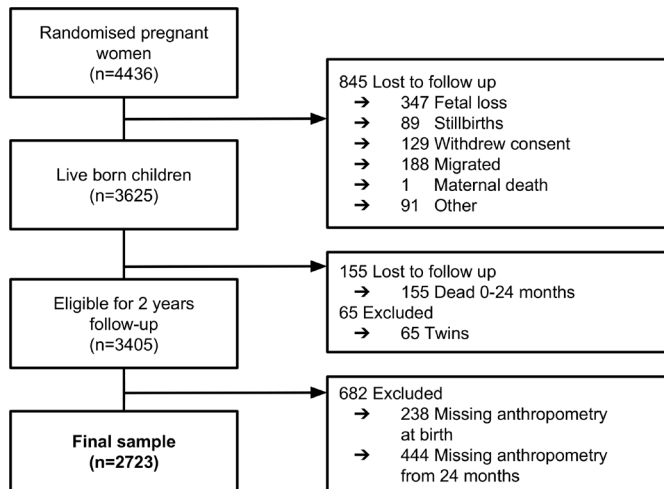
Figure 2 Flow chart of pregnant women and their children included in the data mining analyses of the Maternal and Infant Nutrition Interventions in Matlab cohort from conception to 2 years of age.

RESULTS

There were 4436 women enrolled into the MINIMat trial, of whom 845 were lost to follow-up before delivery, mainly due to fetal death, outmigration, or because they withdrew their consent. Of the 3625 live-born children, including twins and triplets, 155 died between birth and 2 years and 682 were excluded because of missing anthropometry, at birth or at 2 years, resulting in 2723 children available for analysis (figure 2). In the non-analysed group, there was a slightly higher percentage of mothers with more than 5 years of education, younger than 20 years, and belonging to the lowest socioeconomic tertile, and preterm births of children.

The characteristics of the households, mothers, fathers at 8 weeks of gestation and children at birth are given in table 1. The participating mothers had an average age of 26 years (SD 5·6), a mean height of 150 cm (SD 5·3) and a mean weight of 45 kg (SD 6·8) at recruitment. One-third of the women were underweight, with a body

Table 1 Baseline characteristics, prevalence of stunting at 24 months and mean Δ HAZ (change in height-for-age z-score) 0–24 months in the Maternal and Infant Nutrition Interventions in Matlab cohort, Bangladesh.

| Characteristics	n/n (%)	Stunted at 24 months n/n (%)	Mean Δ HAZ 0–24 months
Mother's age (years)			
<20	395/2723 (14.5)	199/395 (50.4)	−0.74
20–29	1556/2723 (57.1)	753/1556 (48.4)	−1.05
>30	772/2723 (28.4)	417/772 (54.0)	−1.28
Mother's education			
No education	913/2723 (33.5)	556/913 (60.9)	−1.27
Enrolled in primary school (1-5y)	624/2723 (22.9)	364/624 (58.3)	−1.24
Completed primary school (>5 y)	1186/2723 (43.6)	449/1186 (37.9)	−0.83
Father's education			
No education	867/2723 (31.8)	532/867 (61.4)	−1.29
Enrolled in primary school (1-5y)	670/2723 (24.6)	369/670 (55.1)	−1.12
Completed primary school (>5 y)	1186/2723 (43.6)	468/1186 (39.5)	−0.89
Parity			
First child	791/2723 (29.0)	348/791 (44.0)	−0.76
Second child	774/2723 (28.4)	385/774 (49.7)	−1.09
Third or more child	1158/2723 (42.5)	636/1158 (54.9)	−1.28
Number of saris mother owns			
<5	1078/2723 (39.6)	665/1078 (61.5)	−1.26
5–8	865/2723 (31.8)	427/865 (49.4)	−1.03
>8	780/2723 (28.6)	277/780 (35.5)	−0.87
Child at birth			
Small for gestational age	1606/2723 (59.0)	972/1606 (60.5)	−1.26
Appropriate for gestational age	1117/2723 (41.0)	397/1117 (35.5)	−0.94
Low birth weight	797/2723 (29.3)	546/797 (68.5)	−0.56
Normal birth weight	1926/2723 (70.7)	823/1926 (42.7)	−1.29
Preterm (<37 weeks of gestation)	190/2723 (7.0)	117/190 (61.6)	0.02
Term	2533/2723 (93)	1252/2533 (49.4)	−1.15
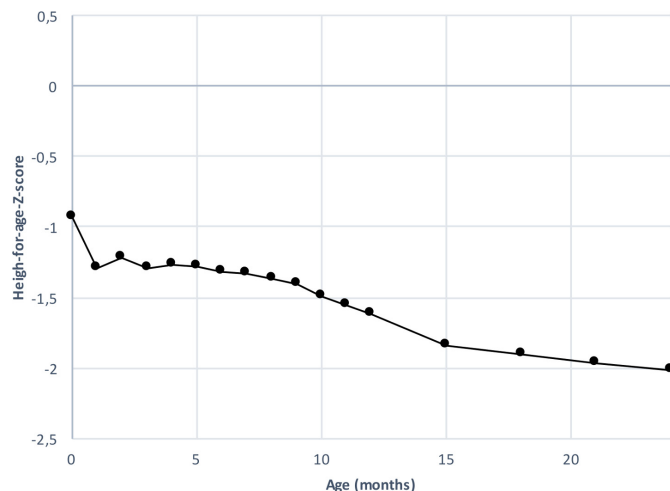
Figure 3 Height-for-age z-scores from birth to 24 months in the Maternal and Infant Nutrition Interventions in Matlab cohort in rural Bangladesh.

mass index below 18·5 at pregnancy week 8. The average number of years of education was similar for mothers and fathers (5 years). The sample of children comprised an equal proportion of girls and boys, and the average birth length was 47·8 cm (SD 2·2), and of birth weight, 2676 grams (SD 410·5). At birth, HAZ was low (mean = −0·94), and declined further at up to 2 years of age with a mean change of −1 HAZ, resulting in a mean HAZ at 2 years of −2·0 (figure 3) and 50% being stunted (girls 51·1%, boys 48·5%).

Relative importance of predictors for stunting at 24 months and change in height scores from birth to 24 months

The relative importance of predictors with respect to their ability to explain the probability of stunting at 24 months and the change in HAZ from birth to 24 months is presented in figures 4 and 5. HAZ and weight-for-age z-scores (WAZ) at birth were the most important
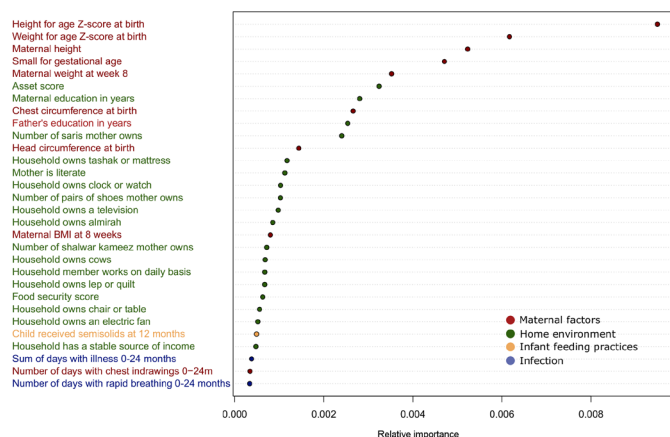

Figure 4 Conditional random forest plot ranking the relative importance of 30 predictors with regard to their ability to explain the presence of stunting at 24 months of age. The Maternal and Infant Nutrition Interventions in Matlab cohort in rural Bangladesh. Colour coding according to WHO conceptual framework on causes of stunting.
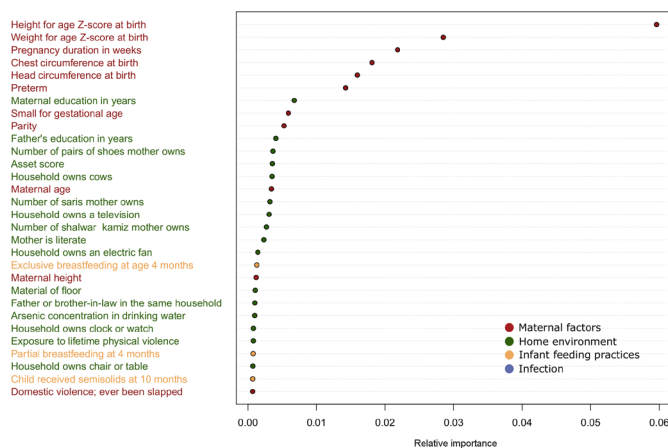

Figure 5 Conditional random forest plot ranking the relative importance of 30 predictors with regard to their ability to explain the variation in change in height-for-age z-scores (HAZ) (Δ HAZ) from birth to 24 months of age. The Maternal and Infant Nutrition Interventions in Matlab cohort in rural Bangladesh. Colour coding according to WHO conceptual framework on causes of stunting.

predictors of stunting at 24 months, followed by maternal height, small for gestational age (SGA), maternal weight at 8 weeks of gestation, household asset score and parental education. The most important factors for Δ HAZ were HAZ and WAZ at birth, pregnancy duration, head and chest circumference at birth and maternal education.

Subgroups with different levels of stunting at 24 months and levels of change in height scores from birth to 24 months

The CIT presented in figures 6 and 7 display subgroups with different probability of stunting at 24 months and levels of Δ HAZ 0–24 months due to distinctive combinations of levels of certain predictors. The CIT for stunting and ΔHAZ were composed of subgroups defined by the same predictors, specifically, HAZ at birth, maternal height, father's educational level and the number of saris owned by the mother. The probability of stunting ranged from 14% to 84%. Children with a HAZ at birth below −1·19, born to mothers with a height below 151.4 cm, who owned less than five saris, had the highest probability of stunting at 24 months, at 84%. Children of a father with more than 7 years of education, who had HAZ at birth above −0·2, had the lowest probability of stunting at 24 months, at 14% (figure 6). The difference in Δ HAZ between the identified subgroups of children with the most negative change and the subgroup with the most positive change was 2·22 HAZ. Children who already had a low HAZ at birth (≤−2·33) had the most positive change in HAZ from birth up to 24 months (+0·18 HAZ), while children who were born with a HAZ above 0.19 had the most negative Δ HAZ (−2·04 HAZ) (figure 7).

DISCUSSION

In our analysis of 309 predictors characterising household, environmental, biological and socioeconomic
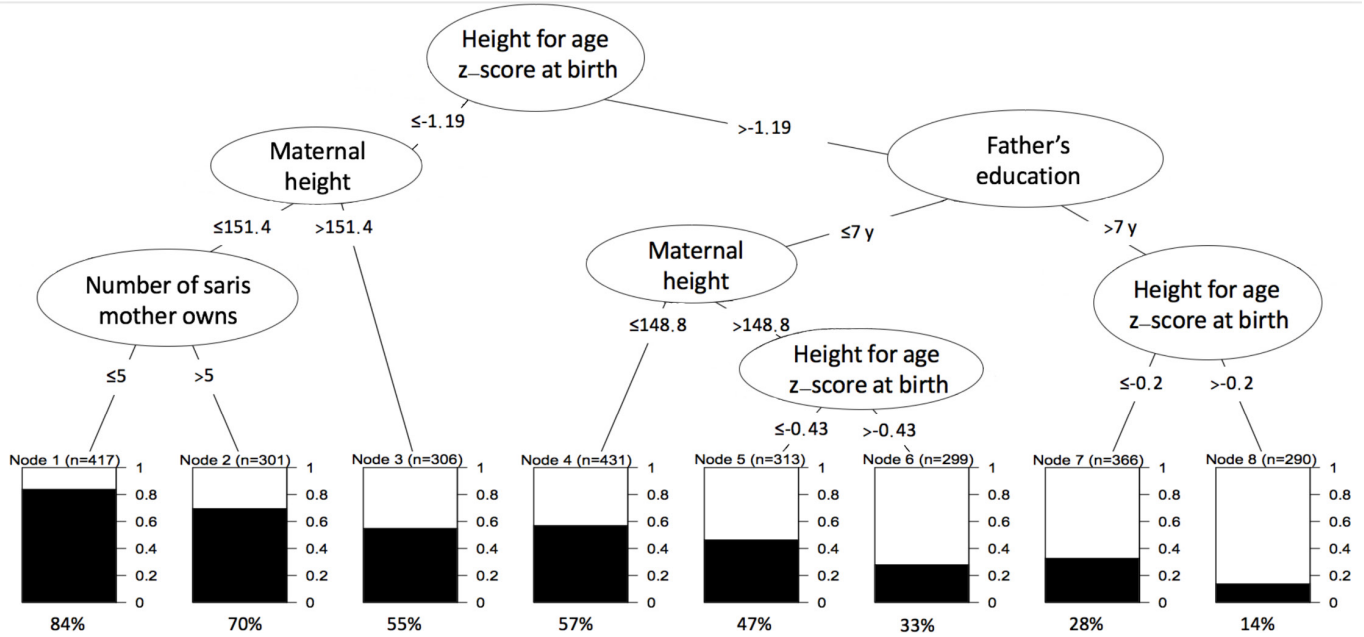
Figure 6 Conditional inference tree identifying subgroups with different probabilities of stunting at 24 months. The Maternal and Infant Nutrition Interventions in Matlab cohort in rural Bangladesh.

factors, we found birth size, maternal anthropometry and parental education to be the most influential for linear growth up to and stunting at 24 months. Conditions after birth, such as feeding practices and morbidity, were less important for linear growth trajectories and stunting at 2 years. The difference between the identified subgroups of children with the highest and lowest probabilities of stunting was high.

The most important predictors of stunting at 24 months were different indicators of size at birth, maternal height, asset score and maternal education. These findings are in line with a multicountry longitudinal study that found birth or enrolment weight of the infant and maternal height to have the highest cumulative ORs for linear growth deficit up to 2 years of age.[11] These results add to the growing evidence that a large part of linear growth
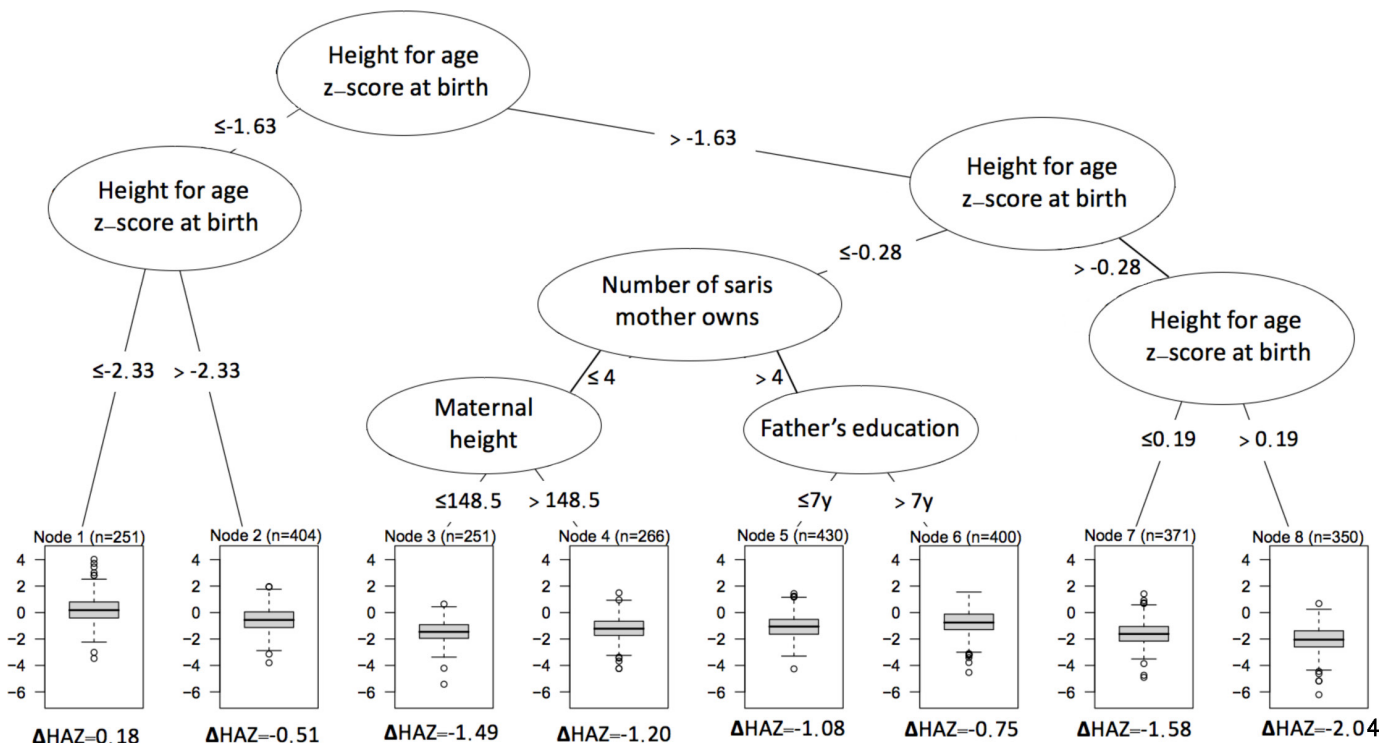
Figure 7 Conditional inference tree identifying subgroups with different mean change in height-for-age z-scores (HAZ) (Δ HAZ=HAZ_{24}–HAZ_0) 0–24 months within the Maternal and Infant Nutrition Interventions in Matlab cohort in rural Bangladesh.

faltering already originates in fetal life.[11 17 43] In a pooled analysis of 19 birth cohorts with longitudinal follow-up, 20% of stunting was attributable to SGA weight at birth.[17] That study did not include any postnatal factors in the analysis. In a study in Indonesia, neonatal length and weight were the strongest predictors of nutritional status and increases in weight and length during infancy.[43] Our study included both prenatal and postnatal factors and, in contrast to most other studies, assessed the relative importance of different potential predictors and the public health importance of each element.

In a study with pooled data from five demographic and health surveys in South Asia, maternal height and underweight, household wealth, maternal education and minimum dietary diversity were found to be the most important factors among children aged 6–23 months.[16] Similar results were reported from a study in India.[44] These studies were, however, cross sectional, without access to birth characteristics.

Maternal height is a strong determinant of fetal growth[45] that indirectly reflect the epigenetic heredity. Maternal height is directly associated with the uterine volume,[46] cephalopelvic disproportion and subsequent infant and childhood stunting, and child mortality.[47 48] In a previous analysis of the MINIMat cohort, a short maternal height was strongly associated with stunting all the way up to 10 years of age.[48] Thus, factors that well precede pregnancy generate a vicious intergenerational cycle, where small mothers give birth to small children of whom a high proportion become and remain stunted. In the CIT for stunting at 24 months, children who were born with a higher HAZ but who had shorter mothers were as likely to be stunted as children with lower HAZ at birth but with a taller mother. This finding suggests that intergenerational improvements in height are achievable and that interventions with a particular focus on adolescents and women of reproductive health are needed to break the vicious intergenerational cycle.

A strong relationship between stunting and poverty has been reported from many low–middle income settings.[49] Asset score and other socioeconomic markers, such as the number of shoes and saris the mother owned, were highly ranked in the random forest analysis and categorised subgroups with a higher probability of stunting and undesirable linear growth trajectories. Poverty is associated with unfavourable food and sanitation practices that can lead to poor nutrition and an increased occurrence of infections during pregnancy, infancy and childhood. Poverty increases the risk of maternal stress, depression[50] and weak mother-to-child interaction and stimulation.

The number of shoes and saris the mother owns might also be markers of the woman's status in the household. During the last few decades, the importance of women's position in household and society for child nutrition has been emphasised.[51] Maternal status is associated with food allocation to mother and child, and a higher level of maternal autonomy has been associated with better child weight and lower levels of stunting.[52] The subordinate position of women in South Asia has been suggested to be a contributor to the high prevalence of child undernutrition in the region, compared with other areas with equivalent levels of economic growth and food security.[51]

An acknowledged way of increasing women's position is through improved education. The remarkable health achievements in Bangladesh over the past two decades can partly be attributed to the progress in access to education, especially at primary level and for girls.[53] However, there is a considerable risk of not completing primary school for both girls and boys.[54] In 2013, the continuation to the last grade of primary school (5 years) was 75%[55] and, in our study, less than 50%. In the conditional decision trees models for stunting and change in HAZ, the cut-off values for paternal and maternal education in the groups with a lower prevalence of stunting and a more positive change in HAZ from birth to 24 months ranged from 6 to 8 years, furthering the importance of girls and boys enrolling in and continuing at school.

It may seem contradictory that children who were born with a very short length had the smallest change in HAZ. This finding most likely reflects a situation where linear growth had already been severely restricted in fetal life.

A multicountry pooled analysis of cohort studies showed that a higher cumulative burden of diarrhoea increased the risk of stunting.[56] In situations, where measles still occurred, its impact on growth and mortality risks were repeatedly documented.[57] One explanation to the discrepancy between our results and previous findings could be Bangladesh's remarkable success in achieving the globally highest coverage of oral rehydration therapy in diarrhoea,[58] which may have reduced the impact on linear growth. Another factor is the almost universal immunisation coverage[53 59] that has reduced or partly eliminated immunization-preventable morbidity and the subsequent effect on growth. Our previous publications on the MINIMat prenatal nutrition interventions' effects on child growth and mortality were not mediated through morbidity,[23 60] further supporting the modest impact of child morbidity on linear growth in our sample.[34] In other settings with lower coverage of diarrhoea treatment and immunisation, the relative importance of these factors may be greater.

Suboptimal infant and early childhood feeding practices have, in earlier studies, been reported as significant risk factors for stunting.[61] A systematic review and meta-analysis of 17 trials showed an average effect of 0.5 cm in height when children 6–24 months had been randomised to appropriate complementary foods.[62] The infant feeding variables included in our analysis ranked low in the random forest analysis and did not show up in any of the CIT. In spite of the relatively few documented effects of complementary feeding programme on stunting, these interventions are often the priority in efforts to combat stunting.

The nutrition interventions from preconception to 2 years of age currently recommended by WHO include efforts to ensure exclusive breast feeding, adequate

complementary feeding, appropriate nutritional care of sick and malnourished children and proper intake of vitamin A, iron and iodine for women and children.[19] All of these, except micronutrient supplementation to pregnant women, are focused on the postnatal period from birth up to 2 years. Our results strengthen the evidence that the process of becoming stunted already begins in utero, as well as the importance of intergenerational effects. Although worthwhile, the present focus on postnatal interventions results in missed opportunities to intervene before or during the first 9 months when the process of stunting is established.

So, what possibilities do we have to improve the postnatal linear growth trajectories prenatally? Attained height is mainly dependent on one's genetic potential for linear growth, in turn determined by DNA sequence polymorphism[63 64] and epigenetic heredity,[65] and to some extent the environment. The modulation of non-DNA sequence epigenetic heredity has been proposed to be one of the leading factors explaining variations in height and height changes over generations,[65] especially in more deprived populations.[66] Postnatal interventions can influence factors in the environment that constrain the ability to increase linear growth, while prenatal interventions also have the potential to modulate the actual growth potential through an epigenetic modification that results from changes to gene expression in response to the fetal environment.

Established prenatal nutritional interventions include balanced energy-protein supplementation, multiple micronutrient supplements and nutritional counselling and education. Unfortunately, most studies evaluating these interventions report only birth weight, not length, which is why evidence to directly assess the effect on fetal linear growth is limited. Meta-analyses and randomised trials evaluating these interventions report their positive impact on birth weight and a reduced risk of LBW.[67–74] Effect sizes vary from increases in birth weight of 20–200 g, with the smallest effects seen in studies of multiple micronutrients and bigger effects seen by balanced energy-protein and lipid-based nutrient supplements. Considerable heterogeneity in growth response is common, and is related to the mother's nutritional status when entering pregnancy and possibly also to the genetic potential to benefit. In the MINIMat food and micronutrient interventions, all women received food supplementation, but they were randomised to an early invitation to supplementation (week 9) or the usual programme start of supplementation (week 20). Children of mothers who participated in food supplementation from early pregnancy (vs the usual start) had a 13% reduction in stunting up to 5 years.[60]

There is increasing evidence that preconception interventions may be even more appropriate.[75] A few trials examining the effect of interventions initiated before pregnancy are underway, but few results have so far been published.[76] Preconception interventions have the potential to bring about epigenetic modulation and improved growth in present and future generations. Thus, the launch and evaluation of interventions targeting adolescent and women of reproductive age that focus on adequate health, education and nutrition before and during pregnancy are needed, especially in South Asia with its high burden of maternal undernutrition and young age at first pregnancy.[77] Targeting high-risk subgroups, in this setting characterised by short, poor, women with low education, can be another strategy to address the intractable problem of stunting.

Strengths and limitations

The extensive database that was available for our analysis covered a wide range of household, family and environmental factors, child characteristics at birth, feeding and morbidity. Infant and young child growth was carefully assessed from birth up to 2 years. The MINIMat cohort was implemented in an excellent research infrastructure that fulfils the prerequisites for obtaining high-quality longitudinal data. Experienced field workers and study nurses collected data on the 309 variables during pregnancy and the following 2 years. They received repeated training, including standardisation exercises, and were supervised by senior medical doctors.

Some potential determinants were not present in the database. Household water, sanitation and hygiene characteristics were limited to information on arsenic contamination of the drinking water, but diarrhoea and other morbidity information were included in our analyses. Further, the cohort did not include the collection of stools for the study of enteropathogens in the child, which may be associated with the risk of stunting.[11] Paternal height, which may be related to fetal growth, was not available.[78] The mothers' smoking habits were not represented in the data, as smoking was extremely rare among women in the study area.

There were slight differences in basic characteristics of the analysed and non-analysed groups. These differences had most likely no influence on the primary outcomes of this study. There were no or few missing values of the critical variables that ranked high in the random forest and defined the subgroups in the CIT. A substudy was carried out to ensure the most accurate method to impute missing data. Thus, it is also highly unlikely that missing data influenced the main findings.

Some of the included variables like 'household asset score' are composite variables, which depend on individual variables like TV ownership, number of cows, and so on. Presence of both composite and individual variables creates computational problems for traditional models like linear regression and for some machine learning models due to a possible high correlation between the individual and the composite variables. However, CIT methods perform automatic variable selection by choosing the most relevant variable (with the strongest association to the response) at each decision tree split step.[39] Accordingly, these methods automatically choose either a composite variable or an individual variable at

each split step based on the relevance of this variable to the response.

Traditional methods like linear regression often have lower predictive power than DM methods. In some cases, the traditional methods are not even possible to compute due to a high number of predictor variables and complex interactions. The method used in this work, CIT, belongs to the class of interpretable machine learning models and display precise information on the priority, size and direction of the association of the predictors with the outcome. In addition, the risk group identification, including the prioritisation and relevant cut-offs of risk factors, can be of high public health relevance for the design and targeting of appropriate interventions with the most significant benefit. Thus, we believe that the CIT framework has a large potential in public health and medical applications.

It can be noted that the CRF and the CIT models are not fully comparable. This can be explained by two factors. First, many predictors that were important in the CRF model are relatively highly correlated and thus have a similar relationship to the response. Once one of these variables is selected by the decision tree in a split, there is a high chance that the remaining correlated variables (although also important according to the CRF) will not be picked up as the next splitting variable. Second, the CRF models and the CIT models cannot be matched directly. The CRF is a combination of many trees and is thus a more flexible model than a CIT. However, CRFs are nearly black-box models: the only interpretable information that these models deliver is the variable importance measure. On the contrary, CITs are 'transparent' and interpretable models but have a smaller predictive power. This is another reason why these models are not generally capable of efficiently embedding all the variables that are important in the CRFs.

Another potential limitation is that decision trees do not deliver p values or CIs. The cross-validation method, however, ensures that the selected tree is optimal. This validation method was chosen superior to other model validation methods, for example, the training-test approach, as it uses the potential of the data to a greater extent at the cost of a greater computational burden.

The study setting was a low socioeconomic area in rural Bangladesh, where maternal and child undernutrition in early life still is widespread. The growth trajectories of our cohort were consistent with established growth trajectories in South Asia, where children are born below WHO growth reference and falter dramatically up to 24 months of age.[5] The subcontinents of South Asia and sub-Saharan Africa share similar proportions of stunted children and faltering patterns. The sub-Saharan African children are, however, on average born slightly bigger than children in South Asia,[5] which makes our results mainly relevant for the South Asian context.

CONCLUSION

This cohort study of determinants of young child stunting in a rural Bangladeshi setting included a wide range of high-quality prenatal and postnatal data, household and family information, environmental factors, child characteristics at birth, infant feeding and morbidity. Prenatal factors including birth size, the mother's anthropometry and parental education were the most critical factors for stunting at 24 months. These results should be seen in contrast to present practice and recommendations that mainly are limited to child interventions. The findings emphasise the benefit of interventions before conception and during pregnancy to reach a substantial reduction in stunting.

Author affiliations
[1]Women's and Children's Health, Uppsala Universitet, Uppsala, Sweden
[2]Center for Epidemiology and Community Medicine, Stockholm, Sweden
[3]Department of Computer and Information Sciences, Linkopings universitet, Linkoping, Sweden
[4]Department of Disease Control, London School of Hygiene and Tropical Medicine, London, UK
[5]Maternal and Child Health Division, International Centre for Diarrhoeal Disease Research, Dhaka, Bangladesh
[6]Health System and Population Studies Division, International Centre for Diarrhoeal Disease Research, Dhaka, Bangladesh
[7]Infectious Diseases Division, International Centre for Diarrhoeal Disease Research, Dhaka, Bangladesh

REFERENCES

1. Pradhan M, Sahn DE, Younger SD. Decomposing World health inequality. *J Health Econ* 2003;22:271–93.

2. Victora CG, Adair L, Fall C, *et al*. Maternal and child undernutrition: consequences for adult health and human capital. *The Lancet* 2008;371:340–57.

3. Black RE, Victora CG, Walker SP, *et al*. Maternal and child undernutrition and overweight in low-income and middle-income countries. *The Lancet* 2013;382:427–51.

4. General Assembly U. *Transforming our world: the 2030 agenda for sustainable development*. United Nations, 2015: 1–35.

5. Victora CG, de Onis M, Hallal PC, *et al*. Worldwide timing of growth faltering: revisiting implications for interventions. *Pediatrics* 2010;125:e473–80.

6. Black RE, Allen LH, Bhutta ZA, *et al*. Maternal and child undernutrition: global and regional exposures and health consequences. *The Lancet* 2008;371:243–60.

7. Bayer R, Galea S. Public Health in the Precision-Medicine Era. - PubMed - NCBI. *N Engl J Med* 2015;373:499–501.

8. Martorell R, Ramakrishnan U, Schroeder DG, *et al*. Intrauterine growth retardation, body size, body composition and physical performance in adolescence. *Eur J Clin Nutr* 1998;52 Suppl 1(Suppl 1):S43–52.

9. Frongillo EA Jr. Symposium: causes and etiology of stunting. Introduction. *J Nutr* 1999;129(2S Suppl):529S–30.

10. Cole TJ, Parkin JM. Infection and its effect on the growth of young children: a comparison of the Gambia and Uganda. *Trans R Soc Trop Med Hyg* 1977;71:196–8.

11. Investigators M-EN. Childhood stunting in relation to the pre- and postnatal environment during the first 2 years of life: the MAL-ED longitudinal birth cohort study. *PLoS Medicine* 2017:1–21.

12. Espo M, Kulmala T, Maleta K, *et al*. Determinants of linear growth and predictors of severe stunting during infancy in rural Malawi. *Acta Paediatr* 2002;91:1364–70.

13. Rahman A, Chowdhury S. Determinants of chronic malnutrition among preschool children in Bangladesh. *J Biosoc Sci* 2007;39:161–73.

14. Amin MR, Murimi MW, Moyeda Carabaza AF. Factors associated with stunting among 0-23 Months-Old children in rural Bangladesh. *J Nutr Educ Behav* 2017;49:S21.

15. Hong R, Banta JE, Betancourt JA. Relationship between household wealth inequality and chronic childhood under-nutrition in Bangladesh. *Int J Equity Health* 2006;5:1–10.

16. Kim R, Mejía-Guevara I, Corsi DJ, *et al*. Relative importance of 13 correlates of child stunting in South Asia: insights from nationally representative data from Afghanistan, Bangladesh, India, Nepal, and Pakistan. *Soc Sci Med* 2017;187:144–54.

17. Christian P, Lee SE, Donahue Angel M, *et al*. Risk of childhood undernutrition related to small-for-gestational age and preterm birth in low- and middle-income countries. *Int J Epidemiol* 2013;42:1340–55.

18. World Health Organization. Who Guidline: daily iron and folic acid supplementation in pregnant women. *WHO* 2012:1–32.

19. World Health Organization. Essential nutrition actions: improving maternal, newborn, infant and young child health and nutrition. *WHO* 2013:1–144.

20. Bhutta ZA, Das JK, Rizvi A, *et al*. Evidence-Based interventions for improvement of maternal and child nutrition: what can be done and at what cost? *The Lancet* 2013;382:452–77.

21. Stewart CP, Iannotti L, Dewey KG, *et al*. Contextualising complementary feeding in a broader framework for stunting prevention. *Matern Child Nutr* 2013;9 Suppl 2:27–45.

22. Breiman L, Friedman J, Stone CJ, *et al*. *Classification and regression trees*. Chapman and Hall/CRC, 1984.

23. Persson Lars Åke, Arifeen S, Ekström E-C, *et al*. Effects of prenatal micronutrient and early food supplementation on maternal hemoglobin, birth weight, and infant mortality among children in Bangladesh: the MINIMat randomized trial. *JAMA* 2012;307:2050–9.

24. Khan AI, Hawkesworth S, Ekström E-C, *et al*. Effects of exclusive breastfeeding intervention on child growth and body composition: the MINIMat trial, Bangladesh. *Acta Paediatr* 2013;102:815–23.

25. Gwatkin DR, Rustein S, Johnson K, *et al*. *Socio-economic differences in health nutrition and population in Bangladesh. Washington*. DC: HNP/Poverty Thematic Group of the World Bank, 2000.

26. Frongillo EA, Chowdhury N, Ekström E-C, *et al*. Understanding the experience of household food insecurity in rural Bangladesh leads to a measure different from that used in other countries. *J Nutr* 2003;133:4158–62.

27. Straus MA, Douglas EM. A short form of the revised conflict tactics scales, and typologies for severity and mutuality. *Violence Vict* 2004;19:507–20.

28. WHO. WHO Multi-country study on women's health and domestic violence against women. *WHO* 2005:1–38.

29. Asling-Monemi K, Naved RT, Persson LA. Violence against women and the risk of fetal and early childhood growth impairment: a cohort study in rural Bangladesh. *Arch Dis Child* 2009;94:775–9.

30. Rahman A, Vahter M, Smith AH, *et al*. Arsenic exposure during pregnancy and size at birth: a prospective cohort study in Bangladesh. *Am J Epidemiol* 2009;169:304–12.

31. Arifeen SE, Black RE, Caulfield LE, *et al*. Infant growth patterns in the slums of Dhaka in relation to birth weight, intrauterine growth retardation, and prematurity. *Am J Clin Nutr* 2000;72:1010–7.

32. World Health Organization. Indicators for assessing breastfeeding practices. *WHO* 1991:1–16.

33. Moore SE, Prentice AM, Coward WA, *et al*. Use of stable-isotope techniques to validate infant feeding practices reported by Bangladeshi women receiving breastfeeding counseling. *Am J Clin Nutr* 2007;85:1075–82.

34. Kallioinen M, Ekström E-C, Khan AI, *et al*. Prenatal early food and multiple micronutrient supplementation trial reduced infant mortality in Bangladesh, but did not influence morbidity. *Acta Paediatr* 2017;106:1979–86.

35. Arnold BF, Galiani S, Ram PK, *et al*. Optimal recall period for caregiver-reported illness in risk factor and intervention studies: a multicountry study. *Am J Epidemiol* 2013;177:361–70.

36. de Onis M, Onyango AW, Borghi E, *et al*. Development of a who growth reference for school-aged children and adolescents. *Bull World Health Organ* 2007;85:660–7.

37. Jönsson P, Wohlin C. An evaluation of k-nearest neighbour imputation using Likert data. *Proceedings of th InternationalSymposium onSoftware Metrics* 2004:108–18.

38. Stekhoven DJ, Bühlmann P. MissForest-non-parametric missing value imputation for mixed-type data. *Bioinformatics* 2012;28:112–8.

39. Torsten Hothorn KHAZ. ctree: conditional inference trees, 2016. https://www.r-project.org/. 1–21. page 6551–21. Available: https://cran.r-project.org/web/packages/partykit/vignettes/ctree.pdf [Accessed 24 Mar 2017].

40. Hastie T, Tibshirani R, Friedman J. *The Elements of Statistical Learning*. New York, NY: Springer Science & Business Media, 2013.

41. R Core Team. R: a language and environment for stastical computing, 2017. Available: httpa://www.R-project.org/

42. Achim Zeileis T. Party: a laboratory for recursive Partytioning 2017:1–18.

43. Schmidt MK, Muslimatun S, West CE, *et al*. Nutritional status and linear growth of Indonesian infants in West Java are determined more by prenatal environment than by postnatal factors. *J Nutr* 2002;132:2202–7.

44. Corsi DJ, Mejía-Guevara I, Subramanian SV. Risk factors for chronic undernutrition among children in India: estimating relative importance, population attributable risk and fractions. *Soc Sci Med* 2016;157:165–85.

45. Kramer MS. The epidemiology of adverse pregnancy outcomes: an overview. *J Nutr* 2003;133(5 Suppl 2):1592S–6.

46. A who Collaborative study of maternal Anhropomehtry and pregnancy outcomes. *International Journal of Gynecology and Obstetrics* 1996;57:1–15.

47. Özaltin E, Hill K, Subramanian SV. Association of maternal stature with offspring mortality, underweight, and stunting in low- to middle-income countries. *JAMA* 2010;303:1507–21.

48. Svefors P, Rahman A, Ekström E-C, *et al*. Stunted at 10 years. linear growth trajectories and stunting from birth to Pre-Adolescence in a rural Bangladeshi cohort. *PLoS One* 2016;11:e0149700–18.

49. Van de Poel E, Hosseinpoor AR, Speybroeck N, *et al*. Socioeconomic inequality in malnutrition in developing countries. *Bull World Health Organ* 2008;86:282–91.

50. Bradley RH, Corwyn RF. Socioeconomic status and child development. *Annu Rev Psychol* 2002;53:371–99.

51. Ramalingaswami V, U J, J R. *The Asian enigma*, 1996.

52. Shroff M, Griffiths P, Adair L, *et al*. Maternal autonomy is inversely related to child stunting in Andhra Pradesh, India. *Matern Child Nutr* 2009;5:64–74.

53. Chowdhury AMR, Bhuiya A, Chowdhury ME, *et al*. The Bangladesh paradox: exceptional health achievement despite economic poverty. *The Lancet* 2013;382:1734–45.

54. Asadullah MN, Chaudhury N. Girls school enrolment figures are good in Bangladesh but drop out rates are worrying. *Development and Cooperation* 2017;59:447–72.

55. UNESCO. Efa 2015 national review: Bangladesh, 2014. unesdoc.unesco.org. Available: http://unesdoc.unesco.org/images/0023/002305/230507e.pdf [Accessed 13 Feb 2018].

56. Checkley W, Buckley G, Gilman RH, *et al*. Multi-country analysis of the effects of diarrhoea on childhood stunting. *Int J Epidemiol* 2008;37:816–30.

57. Koster FT, Curlin GC, Aziz KMA, et al. Synergistic impact of measels and diarrhoea on nutrtion and mortality in Bangladesh. *Bulletin of the World Health Organization* 1981:901–8.
58. El Arifeen S, Christou A, Reichenbach L, et al. Community-Based approaches and partnerships: innovations in health-service delivery in Bangladesh. *The Lancet* 2013;382:2012–26.
59. WHO, UNICEF. *Bangladesh: who and UNICEF estimates of immunization coverage: 2016 revision*, 2016: 1–30.
60. Khan AI, Kabir I, Ekström E-C, et al. Effects of prenatal food and micronutrient supplementation on child growth from birth to 54 months of age: a randomized trial in Bangladesh. *Nutr J* 2011;10:134.
61. Martorell R. Physical growth and development of the malnourished child: contributions from 50 years of research at INCAP. *Food Nutr Bull* 2010;31:68–82.
62. Imdad A, Yakoob MY, Bhutta ZA. Impact of maternal education about complementary feeding and provision of complementary foods on child growth in developing countries. *BMC Public Health* 2011;11 Suppl 3(Suppl 3):S25.
63. Weedon MN, Frayling TM. Reaching new heights: insights into the genetics of human stature. *Trends Genet* 2008;24:595–603.
64. Silventoinen K, Sammalisto S, Perola M, et al. Heritability of adult body height: a comparative study of twin cohorts in eight countries. *Twin Res* 2003;6:399–408.
65. Simeone P, Alberti S. Epigenetic heredity of human height. *Physiol Rep* 2014;2:e12047–7.
66. Silventoinen K, Kaprio J, Lahelma E, et al. Relative effect of genetic and environmental factors on body height: differences across birth cohorts among Finnish men and women. *Am J Public Health* 2000;90:627–30.
67. Imdad A, Bhutta ZA. Effect of balanced protein energy supplementation during pregnancy on birth outcomes. *BMC Public Health* 2011;11 Suppl 3(Suppl 3):S17.
68. Fall CHD, Fisher DJ, Osmond C, et al. Multiple micronutrient supplementation during pregnancy in low-income countries: a meta-analysis of effects on birth size and length of gestation. *Food Nutr Bull* 2009;30(4 Suppl):S533–S546.
69. Ramakrishnan U, Grant FK, Goldenberg T, et al. Effect of multiple micronutrient supplementation on pregnancy and infant outcomes: a systematic review. *Paediatr Perinat Epidemiol* 2012;26 Suppl 1:153–67.
70. West KP, Shamim AA, Mehra S, et al. Effect of maternal multiple micronutrient vs iron-folic acid supplementation on infant mortality and adverse birth outcomes in rural Bangladesh: the JiVitA-3 randomized trial. *JAMA* 2014;312:2649–58.
71. Johnson W, Darboe MK, Sosseh F, et al. Association of prenatal lipid-based nutritional supplementation with fetal growth in rural Gambia. *Matern Child Nutr* 2017;13:e12367.
72. Ashorn P, Alho L, Ashorn U, et al. Supplementation of maternal diets during pregnancy and for 6 months postpartum and infant diets Thereafter with Small-Quantity lipid-based nutrient supplements does not promote child growth by 18 months of age in rural Malawi: a randomized controlled trial. *J Nutr* 2015;145:1345–53.
73. Adu-Afarwuah S, Lartey A, Okronipa H, et al. Lipid-Based nutrient supplement increases the birth size of infants of primiparous women in Ghana. *Am J Clin Nutr* 2015;101:ajcn.:091546–846.
74. Huybregts L, Roberfroid D, Lanou H, et al. Prenatal food supplementation fortified with multiple micronutrients increases birth length: a randomized controlled trial in rural Burkina Faso. *Am J Clin Nutr* 2009;90:1593–600.
75. Gluckman PD, Hanson MA, Cooper C, et al. Effect of in utero and early-life conditions on adult health and disease 2009;359:61–73.
76. Potdar RD, Sahariah SA, Gandhi M, et al. Improving women's diet quality preconceptionally and during gestation: effects on birth weight and prevalence of low birth weight--a randomized controlled efficacy trial in India (Mumbai Maternal Nutrition Project). *Am J Clin Nutr* 2014;100:1257–68.
77. The World Bank. *World development indicators: reproductive health*, 2016: 1–6.
78. Wills AK, Chinchwadkar MC, Joglekar CV, et al. Maternal and paternal height and BMI and patterns of fetal growth: the Pune maternal nutrition study. *Early Hum Dev* 2010;86:535–40.