BMJ Open

Relative importance of pre- and postnatal determinants of stunting. Data mining approaches to the Maternal and Infant Nutrition Interventions in Matlab (MINIMat) cohort, Bangladesh

Journal:	*BMJ Open*
Manuscript ID	bmjopen-2018-025154
Article Type:	Research
Date Submitted by the Author:	01-Aug-2018
Complete List of Authors:	Svefors, Pernilla; Uppsala Universitet, Women's and children's health Sysoev, Oleg; Linkopings universitet, Department of Computer and Information Sciences Ekstrom, Eva- Charlotte; Uppsala Universitet, Women's and children's health Persson, Lars Ake; London School of Hygiene and Tropical Medicine, Department of Disease control Arifeen, Shams E.; International Centre for Diarrhoeal Disease Research, Maternal and Child Health Division Naved, Ruchira; International Centre for Diarrhoeal Disease Research Rahman, Anisur ; International Centre for Diarrhoeal Disease Research, Maternal and Child Health Division Khan, Ashraful Islam; International Centre for Diarrhoeal Disease Research Selling, Katarina; Uppsala Universitet, Women's and children's health
Keywords:	Nutrition < TROPICAL MEDICINE, PUBLIC HEALTH, EPIDEMIOLOGY, PAEDIATRICS

Relative importance of pre- and postnatal determinants of stunting. Data mining approaches to the Maternal and Infant Nutrition Interventions in Matlab (MINIMat) cohort, Bangladesh

Pernilla Svefors[1], Oleg Sysoev[2], Eva-Charlotte Ekström[1], Lars-Åke Persson[3], Shams El Arifeen[4], Ruchira Naved[4], Anisur Rahman[4], Ashraful Islam Khan[4], Katarina Ekholm Selling[1]

[1] International Maternal and Child Health, Department of Women's and Children's Health, Uppsala University, Uppsala, Sweden. [2] Department of Computer and Information Sciences, Linköping University, Linköping, Sweden, [3] Department of Disease Control, London School of Hygiene and Tropical Medicine, London, UK, [4] International Centre for Diarrhoeal Disease Research, Bangladesh (iccdr,b), Dhaka, Bangladesh.

Abstract

Introduction The 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 pre- and postnatal determinants of linear growth 0–24 months and the risk factors for stunting at two years, and to identify subgroups with different growth trajectories and levels of stunting at two years.

Methods Conditional inference-tree-based methods were applied to the extensive Maternal and Infant Nutrition Interventions in Matlab (MINIMat) trial database with 309 variables of 2,723 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 two 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 two years.

Conclusion The results of this study, together with findings from recent reviews, motivate a change in policy and practice, emphasizing the benefit of interventions before conception and during pregnancy to reach a substantial reduction in stunting.

Strengths and limitations of this study

- Assesses the relative public health importance of pre- and post-natal risk factors.
- The extensive database with over 300 variables available for the analysis covers a wide range of pre and postnatal household, family, and environmental factors, child characteristics at birth, infant feeding, and morbidity. However, some potential determinants were not present in the database.
- Includes high-quality longitudinal data with low rates of missing data.
- Employes 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. They do not however, deliver *p*-values or confidence intervals to the results.

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 adult chronic diseases [3]. In 2012, the 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 two years of age, i.e., 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 emphasized the personalized perspective under the universal coverage of health care. Identifying and targeting high-risk subgroups have thus been highlighted as one of the strategies to reach this goal.

Previous studies employing classical statistical methods have identified a wide range of pre- and post-natal factors associated with impaired growth [7-12]. Low birth weight, maternal height, maternal education, poverty and inadequate complementary feeding practices have been recognized as important risk factors [13-15]. Some analyses emphasize the importance of fetal growth restriction for later stunted growth, but rarely is the relative importance of pre- and post-natal factors assessed [16]. Despite these findings, policy documents and recommendations emphasize interventions especially after birth, and pre-natal recommendations are usually limited to routine micronutrient supplementation for pregnant women [17-19].

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 visualized in the WHO's conceptual framework on childhood stunting [20], 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 combinations of predictors. Decision trees [21] 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. 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 randomized 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 pre- and post-natal anthropometric assessments, socioeconomic and biological characteristics of the mother and father, information on maternal exposure to violence, household food security, breast- 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 0–24 months and risk factors for

stunting at two years, and to identify risk groups with negative growth trajectories and high prevalence of stunting at two years.

Methods

Study setting, participants and study design

The MINIMat trial (Maternal and Infant Nutrition Interventions in Matlab, 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 randomized trial primarily evaluating the effect of an early invitation to prenatal food supplementation (versus usual timing) combined with multiple micronutrient supplementation (versus usual program iron-folate) to pregnant women on maternal hemoglobin, birth weight, gestational age at birth, and infant mortality [22]. Further, the participating women were randomly assigned to either counselling for exclusive breastfeeding or a different health education message of equivalent intensity [23]. 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 two 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 the WHO conceptual framework of stunting [20]. 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 [24]. A validated household food security scale was created from eleven 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 [25]. The participating women were also asked whether they had suffered any of thirty morbidity symptoms from twelve different categories, including airway, urinary tract, fever, circulation, bowel, or pain symptoms during the last month. A sum score ranging from zero to twelve was created based on absence of symptoms or those not recorded for each category.

Home visits were followed by clinic visits at local health sub-centers. Maternal height and weight were measured at around eight weeks of gestation using a stadiometer to the nearest 0.1 cm and an electronic scale (Uniscale; SECA) 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 the WHO collaborative study

questionnaire was used [26,27], based on the conflict tactic scale covering physical, sexual and emotional violence and controlling behavior [28]. Household drinking water was analyzed for arsenic concentration [29].

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 [30]. At birth, data on sex, birth weight, length, and breastfeeding practices were collected. During the subsequent two-year study period, the mother-and-child pairs were visited monthly in their homes during the first year, and every three months during the second year. On these occasions, data on infant feeding practices, child morbidity and anthropometry were collected. The mothers were interviewed about breastfeeding and complementary feeding practices. Breastfeeding practices were categorized into exclusive, predominant, partial, or any breastfeeding for each month from one to twelve months. The total time for exclusive, predominant, and any breastfeeding was calculated. The WHO recommendations guided the breastfeeding assessment [31] and results were validated with a stable-isotope technique. The classification of exclusive breastfeeding was found to suffer from limited misclassification in both directions and to be accurate at the group level [32]. The food given to the infant was categorized into semi-solids and solids each month from one to twelve 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 had any of the following symptoms during the last week; fever, cough, difficult breathing, chest in-drawing, rapid breathing, diarrhea, bloody diarrhea and the duration of these symptoms [33]. Categories were created based on whether the child had suffered from fever, respiratory symptoms,

suspected pneumonia, or diarrhea, 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 seven days for child morbidity. One week has been found to be optimal for this kind of morbidity recall assessment [34].

Children's weight was measured by SECA beam and electronic scales (UNICEF Uniscale; SECA Gmbh & Co, Hamburg, Germany) 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 two years, height was measured to the nearest 0.1 cm, using a freestanding stadiometer. Head and chest circumference were 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 standardization of anthropometric measurements, was conducted periodically.

Outcomes

Height-for-age z-scores (HAZ) were calculated from the measured length and height data using the program WHOAnthro, based on the WHO growth reference for children [35]. Children with a HAZ below minus two SD-scores were classified as stunted. Two outcomes were analyzed: stunting at 24 months and the change in HAZ from birth to 24 months, referred to as Δ HAZ.

Statistical analysis

A database was created with 309 variables characterizing 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 sub-set of records that had height measurements at birth and 24 months was selected (n=2 723). The average percent of missing values among all the predictors were

4 %. The highest percent missing were 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 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 neighbor imputation [36], and random forest imputation [37]. The design of the study followed a procedure similar to the strategy described in Jonsson et al. [36], see S 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 [21] are data mining 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 (CART) have been shown to be biased [38]. This motivated us to select the Conditional Inference Trees (CIT) framework, a method that embeds a statistical hypothesis-testing framework into a recursive partitioning algorithm used for model building [38]. Conditional inference trees were used in order to identify sub-groups characterized 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 data set 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[39]. 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 conditional inference trees is created by means of drawing subsamples from the original data and fitting a unique randomized conditional inference tree 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 [39]. 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 exact parameters of the reported trees are shown in STable 1. The programming language R version 3.2.4 [40] and the 'party' package [41] 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.

Results

There were 4436 women enrolled into the MINIMat trial, of whom 845 were lost to follow-up before delivery, mainly due to fetal loss, outmigration, or because they withdrew their consent. Of the 3625 live born children, 155 died between birth and two years and 682 were excluded because of missing anthropometry, at birth or at two years, resulting in 2723 children available for analysis (Figure 2). In the non-analyzed group there was a slightly higher percentage of mothers with more than five years of education, younger than 20 years, and belonging to the lowest socioeconomic tertile, and preterm births of children (data not shown).

The characteristics of the households, mothers, fathers at eight 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 BMI below 18·5 at pregnancy week eight. 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 two years of age with a mean change of -1 HAZ, resulting in a mean HAZ at two years of -2·0 (Figure 3) and 50% being stunted (girls 51·1%, boys 48·5%)

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 MINIMat cohort, Bangladesh.

Characteristics	*n*/*n* (%)	Stunted at 24 months *n*/*n* (%)	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 (>5y)	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 (>5y)	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 (SGA)	1606/2723 (59.0)	972/1606 (60.5)	-1.26
Appropriate for Gestational Age (AGA)	1117/2723 (41.0)	397/1117 (35.5)	-0.94
Low Birth Weight (LBW)	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

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 are presented in Figure 4 and 5. HAZ and weight-for-age Z-scores (WAZ) at birth were the most important predictors of stunting at 24 months, followed by maternal height, Small for Gestational Age (SGA), maternal weight at eight 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 conditional inference trees presented in Figure 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 conditional inference trees for stunting and **Δ**HAZ 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 seven years of education, who had HAZ at birth above -0·2, had the lowest probability of stunting at 24 months, at 14%. 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).

Discussion

In our analysis of 309 predictors characterizing household, environmental, biological, and socioeconomic 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 two years. The difference between the identified subgroups of children with the highest and lowest probabilities of stunting was as high. The probability of stunting at two years for a child born small of a short mother with limited resources (few

saris), was 84%, while a child of better birth length with an educated father only had a probability of 14% to be stunted at two years.

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 two years. The MINIMat cohort was implemented in an excellent research infrastructure that fulfills 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 two years. They received repeated training, including standardization exercises, and were supervised by senior medical doctors.

Some potential determinants were not present in the database. Household water, sanitation, and hygiene (WASH) characteristics have been shown to be associated with the risk of growth restriction by increasing the risk of infections, primarily diarrheal diseases [42]. WASH data in the MINIMat database were limited to information on arsenic contamination of the drinking water, but diarrhea 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 [10]. Paternal height, which may be related to fetal growth, was not available [43]. 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 analyzed and non-analyzed 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 sub-groups in the conditional inference trees. A sub-study 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.

Decision-tree-based methods permit the inclusion of a large number of predictor variables of different types. Complex dependencies between predictor and response variables may be modeled without any need to specify the form of dependence or consider issues regarding multicollinearity. Also, the methods automatically identify interactions and include these in the models. In classical regression models, the inclusion of this large number of predictor variables and their interactions is not computationally possible. A benefit of applying random forest modelling compared to using conventional models with relative risks or odds ratios is that it ranks the predictors according to how important these are for the explaining the outcome. The random forest analysis does not provide information on whether the predictors have a positive or negative relation to the outcome. The conditional inference trees, on the other hand, display precise information on the priority, size, and direction of the association of the predictors with the outcome. The risk group identification, including the prioritization and relevant cut-offs of risk factors, is of high public health relevance for the design and targeting of appropriate interventions with the most significant benefit.

If the data contain two essential and highly correlated predictors, the conditional inference tree method may select only one of them in the analysis, although the other predictor might be as important. Further, decision trees do not deliver *p*-values or confidence intervals to the results. The cross-validation method, however, ensures that the selected tree is optimal. This validation method was chosen superior to other model validation methods, e.g., 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 the WHO growth reference and falter dramatically up to 24 months of age [5]. In South Asia, 39% or 64 million children under five years are reportedly stunted, which accounts for 40% of the global burden. Sub-Saharan Africa is the region with the second highest frequency of stunting. Although these sub-continents share a similar proportion of stunted children and faltering patterns from 3 to 24 months, the sub-Saharan African children are on average born slightly bigger than children in South Asia [5]. This dissimilarity in growth patterns across the continents makes our results mainly relevant for the South Asian context.

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 multi-country longitudinal study that found birth or enrollment weight of the infant and maternal height to have the highest cumulative odds ratios for linear growth deficit up to two years of age [10]. These results add to the growing evidence that a large part of linear growth faltering already originates in fetal life [10,44,45]. In a pooled analysis of 19 birth cohorts with longitudinal follow-up, 20% of stunting was attributable to small-for-gestational-age weight at birth [16]. That study did not include any post-natal 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 [45]. Our study included both pre- and post-natal factors and, in contrast to most other studies, assessed not only the relative importance of different potential predictors, but also 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 [15]. Similar results were reported from a study in India [46]. These studies were, however, cross-sectional, without access to birth characteristics.

Maternal height is a strong determinant of fetal growth [47] that indirectly reflect the epigenetic heredity. Maternal height is directly associated with the uterine volume [48], cephalo-pelvic disproportion and subsequent infant and childhood stunting, and child mortality [49,50]. 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 [50]. 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 conditional inference trees 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 [51]. 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 categorized subgroups with a higher probability of stunting and undesirable linear growth trajectories. Poverty is associated with unfavorable 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 [52] 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 emphasized [53]. 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 [54]. 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 to other areas with equivalent levels of economic growth and food security [53].

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 [55] However, there is a considerable risk of not completing primary school for both girls and boys [56]. In 2013, the continuation to the last grade of primary school (5 years) was 75% [57] 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 not only enrolling in but also 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 multi-country pooled analysis of cohort studies showed that a higher cumulative burden of diarrhea increased the risk of stunting [58]. In situations, where measles still occurred, its impact on growth and mortality risks were repeatedly documented [59]. 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 diarrhea [60], which may have reduced the impact on linear growth. Another factor is the almost universal immunization coverage [61,62] 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 [22,63], further supporting the modest impact of child morbidity on linear growth in our sample [33]. In other settings with lower coverage of diarrhea treatment and immunization, 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 [64]. 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 randomized to appropriate complementary foods [65]. The infant feeding variables included in our analysis ranked low in the random forest analysis and did not show up in any of the conditional inference trees. In spite of the relatively few documented effects of complementary feeding programs on stunting, these interventions are often the priority in efforts to combat stunting.

The nutrition interventions from pre-conception to two years of age currently recommended by the WHO include efforts to ensure exclusive breastfeeding, adequate complementary feeding, appropriate nutritional care of sick and malnourished children and proper intake of vitamin A, iron and iodine for women and children [18]. All of these, except micronutrient supplementation to pregnant women, are focused on the postnatal period from birth up to two 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 nine 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 [66,67] and epigenetic heredity [68], 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[68], especially in more deprived populations [69]. 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 counseling 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 randomized trials evaluating these interventions report their positive impact on birth weight and a reduced risk of LBW [70-77]. Effect sizes vary from increases in birth weight of 20–200g, 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 randomized to an early invitation to supplementation (week 9) or the usual program start of supplementation (week 20). Children of mothers who participated in food supplementation from early pregnancy (versus the usual start) had a 13% reduction in stunting up to five years [63].

There is increasing evidence that preconception interventions may be even more appropriate[78]. A few trials examining the effect of interventions initiated before pregnancy are underway, but few results have so far been published [79]. 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 is needed, especially in South Asia with its high burden of maternal undernutrition and young age at first pregnancy [80]. Targeting high-risk subgroups, in this setting characterized by short, poor, women with low education, can be another strategy to address the intractable problem of stunting.

Contributors

PS contributed to study design, data analysis and interpretation of the results and had the main responsibility of writing the paper. LÅP and SEA were principal investigators of the MINIMat project. ECN, LÅP and KES contributed to the study design. ECE, RN, AR and AIK took part in and supervised data collection. PS, OS, and KES analysed the data. All authors contributed to the preparation of the database, interpretation of the results and reviewed and approved the final version of the manuscript.

Funding

The MINIMat research study was funded by the icddr,b, United Nations Children's Emergency Fund, Swedish International Development Cooperation Agency (Sida), UK Medical Research Council, Swedish Research Council, Department for International Development (DFID), Japan Society for the Promotion of Science, Child Health and Nutrition Research Initiative, Uppsala University, and US Agency for International Development. The funding agencies had no role in the design and conduct of the study, in

the collection, analysis, and interpretation of the data, or in the preparation, review, or approval of the manuscript.

Acknowledgements

We thank the participants and their families in Matlab for their continuing involvement in the MINIMat trial, and the field-team members and data management staff for their excellent work.

Legend to Figures

Figure 1. Factors, variables and outcomes included in the analysis of data from the MINIMat cohort, Bangladesh. Grouping according to the WHO conceptual framework on childhood stunting [20]

Figure 2. Flow chart of pregnant women and their children included in the data mining analyses of the MINIMat cohort from conception to two years of age.

Figure 3. Height-for-age Z-scores from birth to 24 months in the MINIMat cohort in rural Bangladesh.

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 MINIMat cohort in rural Bangladesh. Colour coding according to the 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 HAZ (**Δ** HAZ) from birth to 24 months of age. The MINIMat cohort in rural Bangladesh. Colour coding according to the WHO conceptual framework on causes of stunting.

Figure 6. Conditional inference tree identifying sub-groups with different probabilities of stunting at 24 months. The MINIMat cohort in rural Bangladesh.

Figure 7. Conditional inference tree identifying sub-groups with different mean change in HAZ (**Δ** HAZ=HAZ_{24}–HAZ_0) 0–24 months within the MINIMat cohort in rural Bangladesh.

References

1 Pradhan M, Sahn DE, Younger SD. Decomposing world health inequality. *Journal of Health Economics* 2003;**22**:271–93. doi:10.1016/S0167-6296(02)00123-6

2 Victora CG, Adair L, Fall C, *et al.* Maternal and child undernutrition: consequences for adult health and human capital. *Lancet* 2008;**371**:340–57. doi:10.1016/S0140-6736(07)61692-4

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.

4 General Assembly, UNICEF. 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. doi:10.1542/peds.2009-1519

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. doi:10.1016/S0140-6736(07)61690-0

7 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**:S43–52–discussionS52–3.

8 Frongillo EA Jr. *Symposium: Causes and Etiology of Stunting. J Nutr* 1999;**129**:529S–530S. doi:10.1016/S0140-6736(94)91337-4

9 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. doi:10.1016/0035-9203(77)90005-0

10 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. doi:10.1371/journal.pmed.1002408

11 Espo M, Kulmala T, Maleta K, *et al.* Determinants of linear growth and predictors of severe stunting during infancy in rural Malawi. *Acta Paediatrica* 2002;**91**:1364–70. doi:10.1111/j.1651-2227.2002.tb02835.x

12 Rahman A, Chowdhury S. Determinants of Chronic Malnutrtion Among Preschool Children In Bangladesh. *J Biosoc Sci* 2006;**39**:161. doi:10.1017/S0021932006001295

13 MS MRAM, RN MWMPL, BS AFMC. Factors Associated with Stunting Among 0-23 Months-Old Children in Rural Bangladesh. *Journal of Nutrition Education and Behavior* 2017;**49**:S21. doi:10.1016/j.jneb.2017.05.061

14 Hong R, Banta JE, Betancourt JA. Relationship between household wealth inequality and chronic childhood under-nutrition in Bangladesh. *International Journal for Equity in Health* 2006;**5**:1–10. doi:10.1186/1475-9276-5-15

15 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. *Social Science & Medicine* 2017;**187**:144–54. doi:10.1016/j.socscimed.2017.06.017

16 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. doi:10.1093/ije/dyt109

17 World Health Organization. WHO Guidline: Daily iron and folic acid supplementation in pregnant women. *WHO* 2012;:1–32.

18 World Health Organization. Essential Nutrition Actions: improving maternal, newborn, infant and young child health and nutrition. *WHO* 2013;:1–144.

19 Bhutta ZA, Das JK, Arjumand R, *et al.* Maternal and Child Nutrition 2 Evidence-based interventions for improvement of maternaland child nutrition: what can be done and at what cost? *The Lancet* 2013;**382**:452–77. doi:10.1016/S0140-6736(13)60996-4

20 Stewart CP, Iannotti L, Dewey KG, *et al.* Contextualising complementary feeding in a broader framework for stunting prevention. *Matern Child Nutr* 2013;**9**:27–45. doi:10.1111/mcn.12088

21 Breiman L, Friedman J, Stone CJ, *et al. Classification and Regression Trees*. Chapman and Hall/CRC 1984.

22 Persson LÅ, 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. doi:10.1001/jama.2012.4061

23 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 Paediatrica* 2013;**102**:815–23. doi:10.1111/apa.12282

24 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.

25 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;:4158–62.

26 Straus MA, Douglas EM. A Short Form of the Revised Conflict Tactics Scales, and Typologies for Severity and Mutuality. *Violence and Victims* 2004;**19**:1–15.

27 WHO. WHO Multi-country Study on Women's Health and Domestic Violence against Women. *WHO* 2005;:1–38.

28 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. doi:10.1136/adc.2008.144444

29 Rahman A, Vahter M, Smith AH, *et al.* Arsenic Exposure During Pregnancy and Size at Birth: A Prospective Cohort Study in Bangladesh. *American Journal of Epidemiology* 2008;**169**:304–12. doi:10.1093/aje/kwn332

30 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.

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

32 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;:1075–82.

33 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 Paediatrica* 2017;**382**:427–8. doi:10.1111/apa.14009

34 Arnold BF, Galiani S, Ram PK, *et al.* Optimal Recall Period for Caregiver-reported Illness in Risk Factor and Intervention Studies: A Multicountry Study. *American Journal of Epidemiology* 2013;**177**:361–70. doi:10.1093/aje/kws281

35 Onis M de, Onyango AW, Borghi E, *et al.* Development of a WHO growth reference for school-aged children and adolescents. *Bulletin of the World Health Organization* 2007;**85**:660–7. doi:10.1590/S0042-96862007000900010

36 Jönsson P, Wohlin C. An Evaluation of K-Nearest Neighbour Imputation Using Likert Data. *Proceedings of th International Symposium on Software Metrics* 2004;:108–18.

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

38 Torsten Hothorn KHAZ. ctree: Conditional Inference Trees. www.cran.r-project.org. 2016;:1–21.https://cran.r-project.org/web/packages/partykit/vignettes/ctree.pdf (accessed 24 Mar2017).

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

40 R Core Team. R: A language and environment for stastical computing. 2017. httpa://www.R-project.org/

41 Achim Zeileis THKH. party: A Laboratory for Recursive Partytioning. 2017;:1–18.

42 Cumming O, Cairncross S. Can water, sanitation and hygiene help eliminate stunting? Current evidence and policy implications. *Matern Child Nutr* 2016;**12**:91–105. doi:10.1111/mcn.12258

43 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 Human Development* 2010;**86**:535–40. doi:10.1016/j.earlhumdev.2010.07.002

44 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. doi:10.1093/ije/dyt109

45 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;:2202–7.

46 Corsi DJ, Mejía-Guevara I, Subramanian SV. Risk factors for chronic undernutrition among children in India: Estimating relative importance, population attributable risk and fractions. *Social Science & Medicine* 2016;**157**:165–85. doi:10.1016/j.socscimed.2015.11.014

47 Kramer MS. The Epidemiology of Adverse Pregnancy Outcomes: An Overview. *J Nutr* 2003;:1592S–1596S.

48 A WHO Collaborative Study of Maternal Anhropomehtry and Pregnancy Outcomes. *International Journal of Gynecology and Obstetrics* 1996;**57**:1–15.

49 Özaltin E. Association of Maternal Stature With Offspring Mortality, Underweight, and Stunting in Low- to Middle-Income Countries. *JAMA* 2010;**303**:1507–21. doi:10.1001/jama.2010.450

50 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. doi:10.1371/journal.pone.0149700

51 Van de Poel E, Hosseinpoor AR, Speybroeck N, *et al.* Socioeconomic inequality in malnutrition in developing countries. *Bulletin of the World Health Organization* 2008;**86**:282–91. doi:10.1590/S0042-96862008000400013

52 Bradley RH, Corwyn RF. Socioeconomic Status and Child Development. *Annual Review of Phsycology* 2002;**53**:371–99.

53 Ramalingaswami V, U J, J R. The Asian enigma. 1996.

54 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. doi:10.1111/j.1740-8709.2008.00161.x

55 Chowdhury AMR, Bhuiya A, Chowdhury ME, *et al.* The Bangladesh paradox: exceptional health achievement despite economic poverty. *The Lancet* 2013;**382**:1734–45. doi:10.1016/S0140-6736(13)62148-0

56 Asadullah MN, Chaudhury N. Girls school enrolement figures are good in Bangladesh but drop out rates are worrying. *Development and Cooperation* 2017;**59**:447–72. doi:10.1086/681929

57 UNESCO. EFA 2015 national review: Bangladesh; 2014. unesdoc.unesco.org. http://unesdoc.unesco.org/images/0023/002305/230507e.pdf (accessed 13 Feb2018).

58 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. doi:10.1093/ije/dyn099

59 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.

60 Arifeen DrPH El S, MPH AC, ScD LR, *et al.* Community-based approaches and partnerships: innovations in health-service delivery in Bangladesh. *The Lancet* 2013;**382**:2012–26. doi:10.1016/S0140-6736(13)62149-2

61 WHO, UNICEF. Bangladesh: WHO and UNICEF estimates of immunization coverage: 2016 revision. 2016;:1–30.

62 Chowdhury AMR, Bhuiya A, Chowdhury ME, *et al.* The Bangladesh paradox: exceptional health achievement despite economic poverty. *The Lancet* 2013;**382**:1734–45. doi:10.1016/S0140-6736(13)62148-0

63 Khan AI, Kabir I, Ekstrom EC, *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.

64 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.

65 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**:S25. doi:10.1186/1471-2458-11-S3-S25

66 Weedon MN, Frayling TM. Reaching new heights: insights into the genetics of human stature. *Trends in Genetics* 2008;**24**:595–603. doi:10.1016/j.tig.2008.09.006

67 Silventoinen K, Sammalisto S, Perola M, *et al.* Heritability of Adult Body Height: A Comparative Study of Twin Cohorts in Eight Countries. *Twin Research* 2003;**6**:399–408. doi:10.1375/twin.6.5.399

68 Simeone P, Alberti S. Epigenetic heredity of human height. *Physiological Reports* 2014;**2**:e12047–7. doi:10.14814/phy2.12047

69 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. *American Journal of Public Health* 2000;**90**:627–30.

70 Imdad A, Bhutta ZA. Effect of balanced protein energy supplementation during pregnancy on birth outcomes. *BMC Public Health* 2011;**11**:S17. doi:10.1186/1471-2458-11-S3-S17

71 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**:S533–46. doi:10.1177/15648265090304S408

72 Ramakrishnan U, Grant FK, Goldenberg T, *et al.* Effect of Multiple Micronutrient Supplementation on Pregnancy and Infant Outcomes: A Systematic Review. *Paediatric and Perinatal Epidemiology* 2012;**26**:153–67. doi:10.1111/j.1365-3016.2012.01276.x

73 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. doi:10.1001/jama.2014.16819

74 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. doi:10.1111/mcn.12367

75 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. doi:10.3945/jn.114.207225

76 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. doi:10.3945/ajcn.114.091546

77 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. *American Journal of Clinical Nutrition* 2009;**90**:1593–600. doi:10.3945/ajcn.2009.28253

78 Gluckman PD, Hanson MA, Cooper C, *et al.* Effect of In Utero and Early-Life Conditions on Adult Health and Disease. *http://dxdoiorg/101056/NEJMra0708473* 2009;**359**:61–73. doi:10.1056/NEJMra0708473

79 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). *American Journal of Clinical Nutrition* 2014;**100**:1257–68. doi:10.3945/ajcn.114.084921

80 The World Bank. World Development Indicators: Reproductive health. 2016;:1–6.

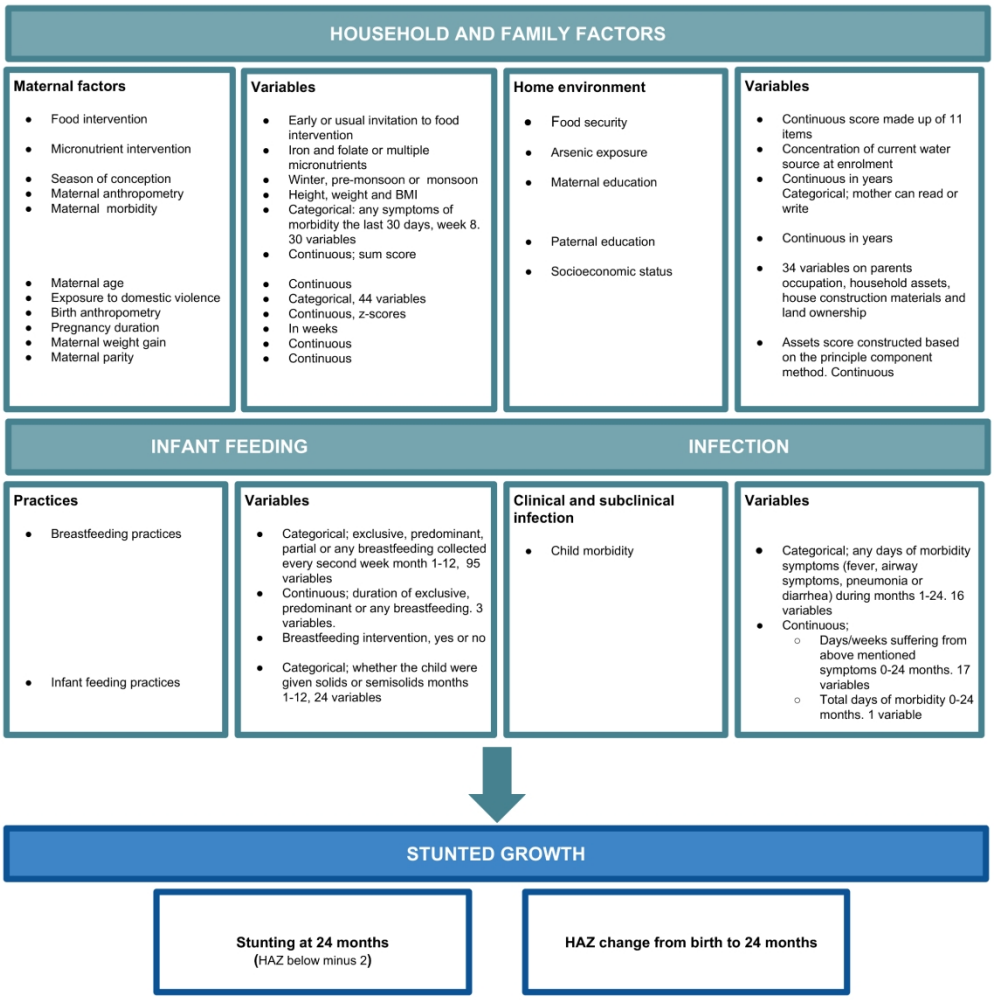

Figure 1. Factors, variables and outcomes included in the analysis of data from the MINIMat cohort, Bangladesh. Grouping according to the WHO conceptual framework on childhood stunting [20]

190x190mm (300 x 300 DPI)

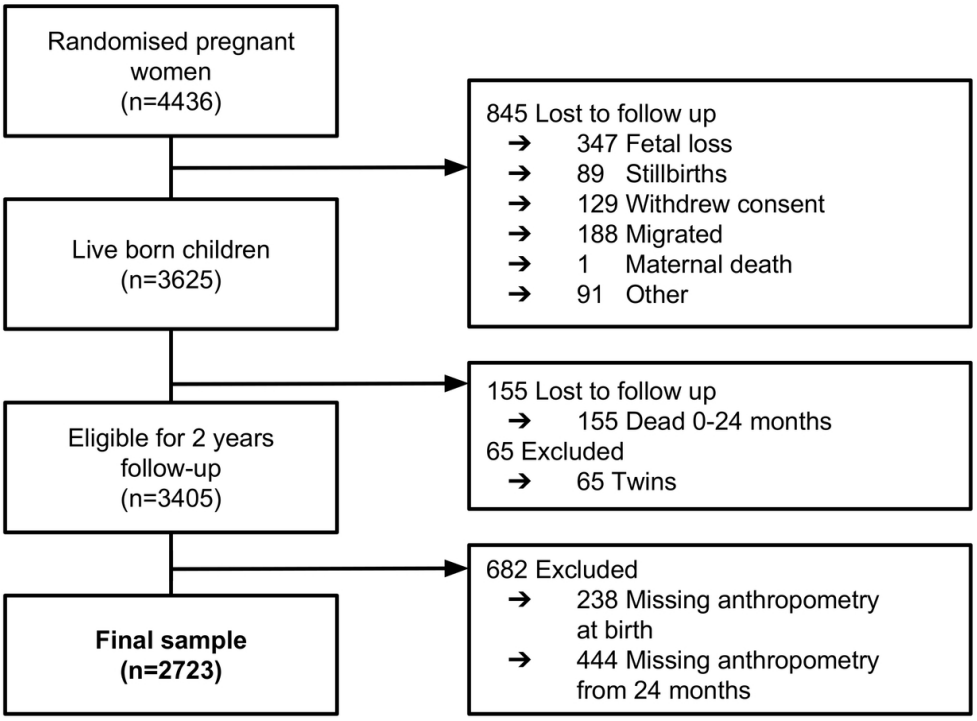

Figure 2. Flow chart of pregnant women and their children included in the data mining analyses of the MINIMat cohort from conception to two years of age.

106x80mm (300 x 300 DPI)

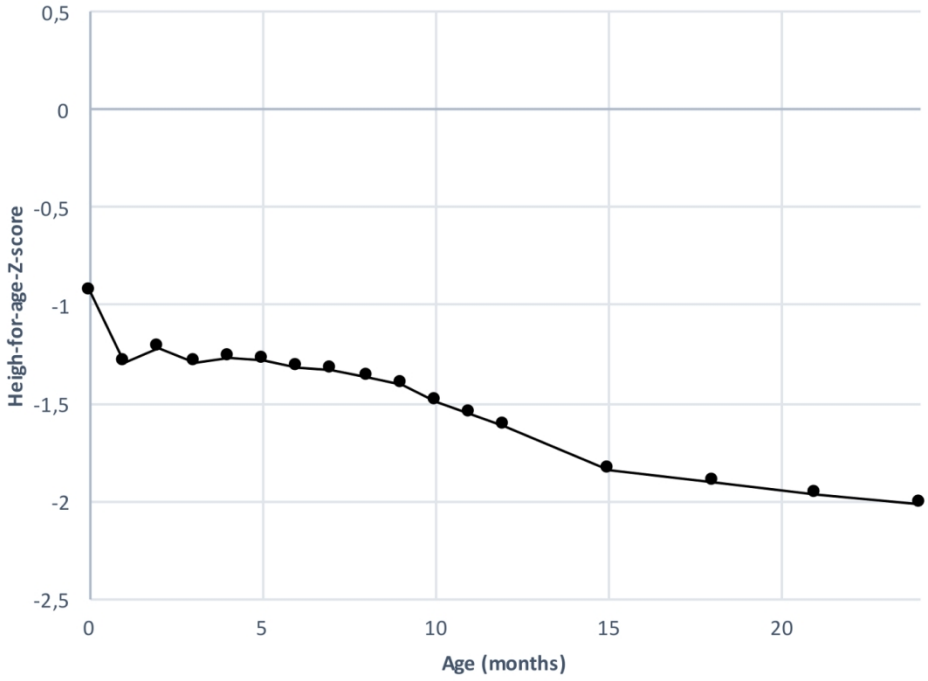

Figure 3. Height-for-age Z-scores from birth to 24 months in the MINIMat cohort in rural Bangladesh.

131x96mm (300 x 300 DPI)

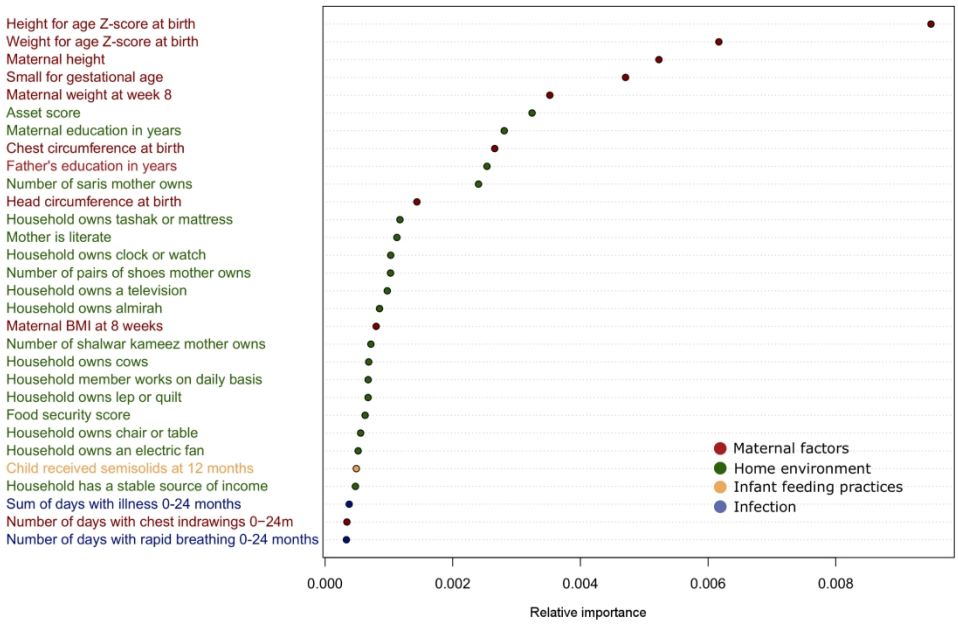

Figure 4. Conditional random forest plot ranking the relative importance of 30 predictors with regard to their ability to explain the presence of stunting at 24 months of age. The MINIMat cohort in rural Bangladesh. Colour coding according to the WHO conceptual framework on causes of stunting.

190x134mm (300 x 300 DPI)

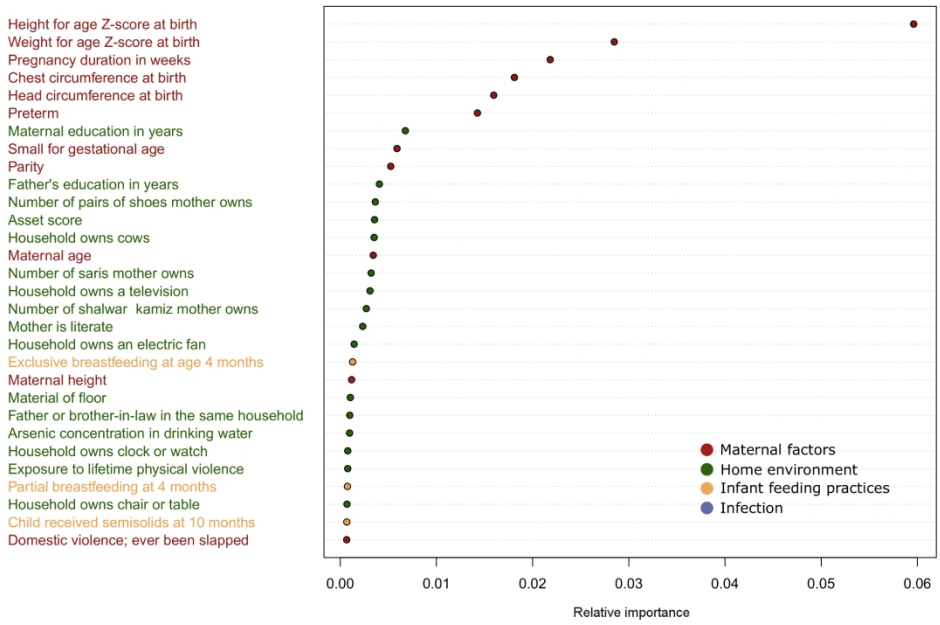

Figure 5. Conditional random forest plot ranking the relative importance of 30 predictors with regard to their ability to explain the variation in change in HAZ (Δ HAZ) from birth to 24 months of age. The MINIMat cohort in rural Bangladesh. Colour coding according to the WHO conceptual framework on causes of stunting.

190x134mm (300 x 300 DPI)

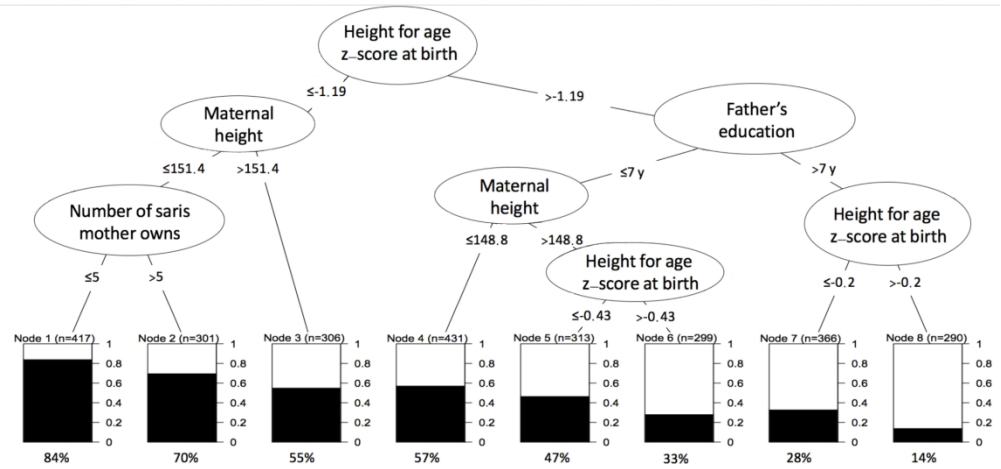

Figure 6. Conditional inference tree identifying sub-groups with different probabilities of stunting at 24 months. The MINIMat cohort in rural Bangladesh.

190x88mm (300 x 300 DPI)

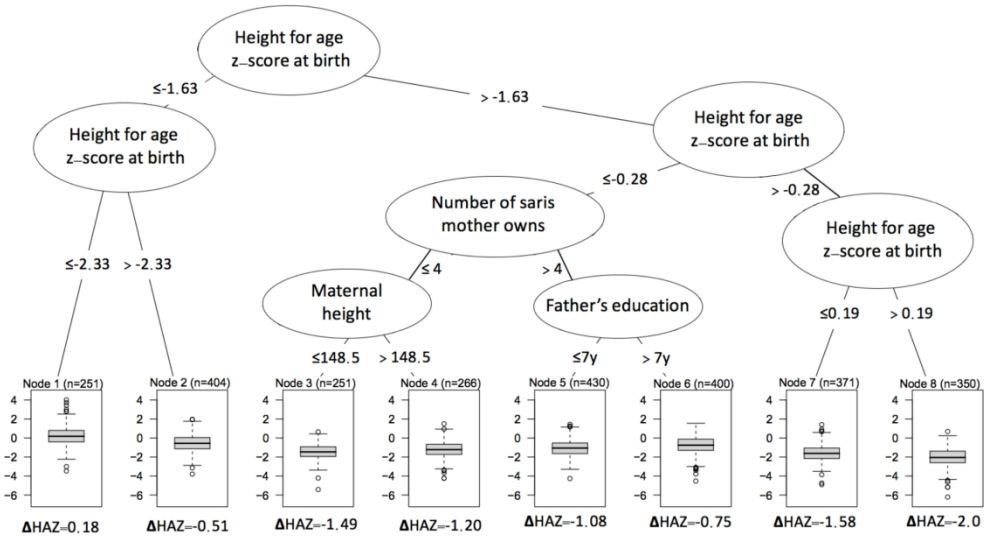

Figure 7. Conditional inference tree identifying sub-groups with different mean change in HAZ (Δ HAZ=HAZ24–HAZ0) 0–24 months within the MINIMat cohort in rural Bangladesh.

190x101mm (300 x 300 DPI)

Supplementation appendix

Simulation study of the predictive performance of three different imputation methods

The following strategy was used to study the imputation accuracy of various methods for the input variables in our analyses. First, we standardized numerical variables in the data and took a sample of the entire data (α) and deleted a proportion (β) of the non-missing values in each variable. Secondly, we employed three different imputation methods to make predictions of the missing values in the data. Lastly, we compared the predictions with the values of the deleted entries, the computed mean-square error (MSE) for the numerical variables, and the percent of the incorrect predictions, misclassification rate (MR), for the categorical ones. The computation of the MSE and MR values was repeated several times for different samples of the original data. The summary results of these computations are presented in Tables 1-4. It can be concluded that random forests[1] provided a statistically significantly better imputation than the variable mean and K-nearest neighbor imputation methods. The design of the study followed a procedure similar to the strategy described in Jonsson et al [2].

Table 1: Means and Standard errors of the MR[2] and the MSE[3] for different imputation methods, computed from m=100 samples, $\alpha = 0.05$, $\beta = 0.05$

	Variable mean	KNN[1]	Random forest
Mean (MR[2])	0.17755631	0.187499573	0.131724506
Standard Error (MR[2])	0.00360524	0.003795385	0.003759032
Mean (MSE[3])	1.01903348	0.901518114	0.541867921
Standard error (MSE[3])	0.01640172	0.016414433	0.015157205

[1] K-nearest neighbour
[2]Missclassification rate
[3]Mean square error
α = proportion of the non-missing values deleted
β = proportion of the original data sampled

Table 2: Means and Standard errors of the MR[2] and the MSE[3] for different imputation methods, computed from m=100 samples, $\alpha = 0.05$, $\beta = 0.15$

	Variable mean	KNN[1]	Random forest
Mean (MR[2])	0.175774830	0.187158897	0.131724506
Standard Error (MR[2])	0.003075253	0.003317242	0.003302446
Mean (MSE[3])	1.00474998	0.922010327	0.556762189
Standard error (MSE[3])	0.01012910	0.009595471	0.008949707

[1] K-nearest neighbour
[2]Missclassification rate
[3]Mean square error
α = proportion of the non-missing values deleted
β = proportion of the original data sampled

Table 3: Means and Standard errors of the MR[2] and the MSE[3] for different imputation methods, computed from m=100 samples, $\alpha = 0.2$, $\beta = 0.05$

	Variable mean	KNN[1]	Random forest
Mean (MR[2])	0.1625007370	0.1608280983	0.094319580
Standard Error (MR[2])	0.0005210379	0.0005181798	0.000367369
Mean (MSE[3])	1.0023969039	0.7975006166	0.450253626
Standard error (MSE[3])	0.0068209597	0.0066997794	0.006069386

[1] K-nearest neighbour
[2]Missclassification rate
[3]Mean square error
α = proportion of the non-missing values deleted
β = proportion of the original data sampled

Table 4: Means and Standard errors of discrete and continuous variables for different imputation methods. Computed from m=100 samples, $\alpha = 0.2$, $\beta = 0.15$

	Variable mean	KNN[1]	Random forest
Mean, discrete	0.1626095174	0.1617267853	0.1017561946
Standard error, Discrete	0.0003670347	0.0003618961	0.0002612874
Mean, continuous	0.9984641615	0.8195273545	0.4593241548
Standard error, continuous	0.0040175223	0.0040319899	0.0034449935

[1] K-nearest neighbour
[2]Missclassification rate
[3]Mean square error
α = proportion of the non-missing values deleted
β = proportion of the original data sampled

References

1. Stekhoven DJ, Buhlmann P. MissForest—non-parametric missing value imputation for mixed-type data. Bioinformatics. 2012;: 112–118.

2. Jönsson P, Wohlin C. An Evaluation of K-Nearest Neighbour Imputation Using Likert Data. Proceedings of th International Symposium on Software Metrics. 2004;: 108–118.

STROBE Statement—checklist of items that should be included in reports of observational studies

	Item No	Recommendation	
Title and abstract	1	(*a*) Indicate the study's design with a commonly used term in the title or the abstract	1
		(*b*) Provide in the abstract an informative and balanced summary of what was done and what was found	2
Introduction			
Background/rationale	2	Explain the scientific background and rationale for the investigation being reported	4-5
Objectives	3	State specific objectives, including any prespecified hypotheses	6
Methods			
Study design	4	Present key elements of study design early in the paper	5
Setting	5	Describe the setting, locations, and relevant dates, including periods of recruitment, exposure, follow-up, and data collection	6
Participants	6	(*a*) *Cohort study*—Give the eligibility criteria, and the sources and methods of selection of participants. Describe methods of follow-up *Case-control study*—Give the eligibility criteria, and the sources and methods of case ascertainment and control selection. Give the rationale for the choice of cases and controls *Cross-sectional study*—Give the eligibility criteria, and the sources and methods of selection of participants	6
		(*b*) *Cohort study*—For matched studies, give matching criteria and number of exposed and unexposed *Case-control study*—For matched studies, give matching criteria and the number of controls per case	
Variables	7	Clearly define all outcomes, exposures, predictors, potential confounders, and effect modifiers. Give diagnostic criteria, if applicable	9 Figure 1
Data sources/ measurement	8*	For each variable of interest, give sources of data and details of methods of assessment (measurement). Describe comparability of assessment methods if there is more than one group	*7-9*
Bias	9	Describe any efforts to address potential sources of bias	7-9
Study size	10	Explain how the study size was arrived at	Not applicable
Quantitative variables	11	Explain how quantitative variables were handled in the analyses. If applicable, describe which groupings were chosen and why	7-9
Statistical methods	12	(*a*) Describe all statistical methods, including those used to control for confounding	10-11
		(*b*) Describe any methods used to examine subgroups and interactions	10-11
		(*c*) Explain how missing data were addressed	10-11
		(*d*) *Cohort study*—If applicable, explain how loss to follow-up was addressed *Case-control study*—If applicable, explain how matching of cases and controls was addressed	10

Cross-sectional study—If applicable, describe analytical methods taking account of sampling strategy	
(*e*) Describe any sensitivity analyses	10

Continued on next page10

Results			
Participants	13*	(a) Report numbers of individuals at each stage of study—eg numbers potentially eligible, examined for eligibility, confirmed eligible, included in the study, completing follow-up, and analysed	Fig 2 12
		(b) Give reasons for non-participation at each stage Fig 2	Fig 2
		(c) Consider use of a flow diagram Figure 2	Fig 2
Descriptive data	14*	(a) Give characteristics of study participants (eg demographic, clinical, social) and information on exposures and potential confounders	Table 1, 13
		(b) Indicate number of participants with missing data for each variable of interest	13
		(c) *Cohort study*—Summarise follow-up time (eg, average and total amount)	13
Outcome data	15*	*Cohort study*—Report numbers of outcome events or summary measures over time	*13*
		Case-control study—Report numbers in each exposure category, or summary measures of exposure	
		Cross-sectional study—Report numbers of outcome events or summary measures	
Main results	16	(*a*) Give unadjusted estimates and, if applicable, confounder-adjusted estimates and their precision (eg, 95% confidence interval). Make clear which confounders were adjusted for and why they were included	14
		(*b*) Report category boundaries when continuous variables were categorized	14
		(*c*) If relevant, consider translating estimates of relative risk into absolute risk for a meaningful time period	
Other analyses	17	Report other analyses done—eg analyses of subgroups and interactions, and sensitivity analyses	14
Discussion			
Key results	18	Summarise key results with reference to study objectives	15,18
Limitations	19	Discuss limitations of the study, taking into account sources of potential bias or imprecision. Discuss both direction and magnitude of any potential bias	15,16,17
Interpretation	20	Give a cautious overall interpretation of results considering objectives, limitations, multiplicity of analyses, results from similar studies, and other relevant evidence	18-22
Generalisability	21	Discuss the generalisability (external validity) of the study results	17
Other information			
Funding	22	Give the source of funding and the role of the funders for the present study and, if applicable, for the original study on which the present article is based	23

*Give information separately for cases and controls in case-control studies and, if applicable, for exposed and unexposed groups in cohort and cross-sectional studies.

Note: An Explanation and Elaboration article discusses each checklist item and gives methodological background and published examples of transparent reporting. The STROBE checklist is best used in conjunction with this article (freely available on the Web sites of PLoS Medicine at http://www.plosmedicine.org/, Annals of Internal Medicine at http://www.annals.org/, and Epidemiology at http://www.epidem.com/). Information on the STROBE Initiative is available at www.strobe-statement.org.

BMJ Open

Relative importance of pre- and postnatal determinants of stunting; data mining approaches to the MINIMat cohort, Bangladesh

Journal:	*BMJ Open*
Manuscript ID	bmjopen-2018-025154.R1
Article Type:	Research
Date Submitted by the Author:	22-Mar-2019
Complete List of Authors:	Svefors, Pernilla; Uppsala Universitet, Women's and children's health; Center for epidemiology and community medicine, Region of Stockholm Sysoev, Oleg; Linkopings universitet, Department of Computer and Information Sciences Ekstrom, Eva- Charlotte; Uppsala Universitet, Women's and children's health Persson, Lars Ake; London School of Hygiene and Tropical Medicine, Department of Disease control Arifeen, Shams E.; International Centre for Diarrhoeal Disease Research, Maternal and Child Health Division Naved, Ruchira; International Centre for Diarrhoeal Disease Research Rahman, Anisur ; International Centre for Diarrhoeal Disease Research, Maternal and Child Health Division Khan, Ashraful Islam; International Centre for Diarrhoeal Disease Research Selling, Katarina; Uppsala Universitet, Women's and children's health
Primary Subject Heading:	Global health
Secondary Subject Heading:	Paediatrics, Nutrition and metabolism
Keywords:	Nutrition < TROPICAL MEDICINE, PUBLIC HEALTH, EPIDEMIOLOGY, PAEDIATRICS

Relative importance of pre- and postnatal determinants of stunting; data mining approaches to the MINIMat cohort, Bangladesh

Pernilla Svefors[1*], Oleg Sysoev[2], Eva-Charlotte Ekström[1], Lars-Åke Persson[3], Shams El Arifeen[4], Ruchira Naved[4], Anisur Rahman[4], Ashraful Islam Khan[4], Katarina Ekholm Selling[1]

[1] International Maternal and Child Health, Department of Women's and Children's Health, Uppsala University, Uppsala, Sweden. [2] Department of Computer and Information Sciences, Linköping University, Linköping, Sweden, [3] Department of Disease Control, London School of Hygiene and Tropical Medicine, London, UK, [4] International Centre for Diarrhoeal Disease Research, Bangladesh (icddr,b), Dhaka, Bangladesh.

*Corresponding author pernilla.svefors@kbh.uu.se

Abstract

Introduction The 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 pre- and postnatal determinants of linear growth 0–24 months and the risk factors for stunting at two years, and to identify subgroups with different growth trajectories and levels of stunting at two years.

Methods Conditional inference-tree-based methods were applied to the extensive Maternal and Infant Nutrition Interventions in Matlab (MINIMat) trial database with 309 variables of 2,723 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 two 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 two years.

Conclusion The results of this study, together with findings from recent reviews, motivate a change in policy and practice, emphasizing the benefit of interventions before conception and during pregnancy to reach a substantial reduction in stunting.

Strengths and limitations of this study

- Assesses the relative public health importance of pre- and post-natal risk factors.
- Includes high-quality longitudinal data with low rates of missing data on child growth and a wide range of pre and postnatal household, family, and environmental factors, child characteristics at birth, infant feeding, and morbidity.
- Some potential important determinants of linear growth were not present in the database.
- 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.

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 adult chronic diseases [3]. In 2012, the 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 two years of age, i.e., 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 emphasized the personalized perspective under the universal coverage of health care. Identifying and targeting high-risk subgroups have thus been highlighted as one of the strategies to reach this goal.

Previous studies employing classical statistical methods have identified a wide range of pre- and post-natal factors associated with impaired growth [7-12]. Low birth weight, maternal height, maternal education, poverty and inadequate complementary feeding practices have been recognized as important risk factors [13-15]. Some analyses emphasize the importance of fetal growth restriction for later stunted growth, but rarely is the relative importance of pre- and post-natal factors assessed [16]. Despite these findings, policy documents and recommendations emphasize interventions especially after birth, and pre-natal recommendations are usually limited to routine micronutrient supplementation for pregnant women [17-19].

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 visualized in the WHO's conceptual framework on childhood stunting [20], 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 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 [21]. 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 randomized 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 pre- and post-natal anthropometric assessments, socioeconomic and biological characteristics of the mother and father, information on maternal exposure to violence, household food security, breast- 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 0–24 months and risk factors for stunting at two years, and to identify risk groups with negative growth trajectories and high prevalence of stunting at two years.

Methods

Study setting, participants and study design

The MINIMat trial (Maternal and Infant Nutrition Interventions in Matlab, 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 randomized trial primarily evaluating the effect of an early invitation to prenatal food supplementation (versus usual timing) combined with multiple micronutrient supplementation (versus usual program iron-folate) to pregnant women on maternal hemoglobin, birth weight, gestational age at birth, and infant mortality [22]. Further, the participating women were randomly assigned to either counselling for exclusive breastfeeding or a different health education message of equivalent intensity [23]. 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 two 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 the WHO conceptual framework of stunting [20].

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 [24]. A validated household food security scale was created from eleven 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 [25]. The participating women were also asked whether they had suffered any of thirty morbidity symptoms from twelve different categories, including airway, urinary tract, fever, circulation, bowel, or pain symptoms during the last month. A sum score ranging from zero to twelve was created based on absence of symptoms or those not recorded for each category.

Home visits were followed by clinic visits at local health sub-centers. Maternal height and weight were measured at around eight weeks of gestation using a stadiometer to the nearest 0.1 cm and an electronic scale (Uniscale; SECA) 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 the WHO collaborative study questionnaire was used [26,27], based on the conflict tactic scale covering physical, sexual and emotional violence and controlling behavior [28]. Household drinking water was analyzed for arsenic concentration [29].

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 [30]. At birth, data on sex, birth weight, length, and breastfeeding practices were collected. During the subsequent two-year study period, the mother-and-child pairs were visited monthly in their homes during the first year, and every three months during the second year. On these occasions, data on infant feeding practices, child morbidity and anthropometry were collected. The mothers were interviewed about breastfeeding and complementary feeding practices. Breastfeeding practices were categorized into exclusive, predominant, partial, or any breastfeeding for each month from one to twelve months. The total time for exclusive, predominant, and any breastfeeding was calculated. The WHO recommendations guided the breastfeeding assessment [31] and results were validated with a stable-isotope technique. The classification of exclusive breastfeeding was found to suffer from limited misclassification in both directions and to be accurate at the group level [32]. The food given to the infant was categorized into

semi-solids and solids each month from one to twelve 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 had any of the following symptoms during the last week; fever, cough, difficult breathing, chest in-drawing, rapid breathing, diarrhea, bloody diarrhea and the duration of these symptoms [33]. Categories were created based on whether the child had suffered from fever, respiratory symptoms, suspected pneumonia, or diarrhea, 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 seven days for child morbidity. One week has been found to be optimal for this kind of morbidity recall assessment [34].

Children's weight was measured by SECA beam and electronic scales (UNICEF Uniscale; SECA Gmbh & Co, Hamburg, Germany) 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 two years, height was measured to the nearest 0.1 cm, using a freestanding stadiometer. Head and chest circumference were 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 standardization of anthropometric measurements, was conducted periodically.

Outcomes

Height-for-age z-scores (HAZ) were calculated from the measured length and height data using the program WHOAnthro, based on the WHO growth reference for children [35]. Children with a HAZ below minus two SD-scores were classified as stunted. Two outcomes were analyzed: stunting at 24 months and the change in HAZ from birth to 24 months, referred to as Δ HAZ.

Statistical analysis

A database was created with 309 variables characterizing 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 sub-set of records that had height measurements at birth and 24 months was selected (n=2 723). The average percent of missing values among all the predictors were 4 %. The highest percent missing were 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 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 neighbor imputation [36], and random forest imputation [37]. The design of the study followed a procedure similar to the strategy described in Jonsson et al. [36], see S 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 [21] are data mining 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 (CART) have been shown to be biased [38]. This motivated us to select the Conditional Inference Trees (CIT) framework, a method that embeds a statistical hypothesis-testing framework into a recursive partitioning algorithm used for model building [38]. Conditional inference trees were used in order to identify sub-groups characterized 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 data set 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[39]. 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 conditional inference trees is created by means of drawing subsamples from the original data and fitting a unique randomized conditional inference tree 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 [39]. 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 version 3.2.4 [40] and the 'party' package [41] 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.

Results

There were 4436 women enrolled into the MINIMat trial, of whom 845 were lost to follow-up before delivery, mainly due to fetal loss, outmigration, or because they withdrew their consent. Of the 3625 live born children, 155 died between birth and two years and 682 were excluded because of missing anthropometry, at birth or at two years, resulting in 2723 children available for analysis (Figure 2). In the non-analyzed group there was a slightly higher percentage of mothers with more than five years of education, younger than 20 years, and belonging to the lowest socioeconomic tertile, and preterm births of children (data not shown).

The characteristics of the households, mothers, fathers at eight 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 BMI below 18·5 at pregnancy week eight. 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 two years of age with a mean change of -1 HAZ, resulting in a mean HAZ at two years of -2·0 (Figure 3) and 50% being stunted (girls 51·1%, boys 48·5%)

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 MINIMat 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 (>5y)	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 (>5y)	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 (SGA)	1606/2723 (59.0)	972/1606 (60.5)	-1.26
Appropriate for Gestational Age (AGA)	1117/2723 (41.0)	397/1117 (35.5)	-0.94
Low Birth Weight (LBW)	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

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 are presented in Figure 4 and 5. HAZ and weight-for-age Z-scores (WAZ) at birth were the most important predictors of stunting at 24 months, followed by maternal height, Small for Gestational Age (SGA), maternal weight at eight 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 conditional inference trees presented in Figure 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 conditional inference trees 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 seven years of education, who had HAZ at birth above -0·2, had the lowest probability of stunting at 24 months, at 14%. 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).

Discussion

In our analysis of 309 predictors characterizing household, environmental, biological, and socioeconomic 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

two years. The difference between the identified subgroups of children with the highest and lowest probabilities of stunting was as high.

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 two years. The MINIMat cohort was implemented in an excellent research infrastructure that fulfills 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 two years. They received repeated training, including standardization exercises, and were supervised by senior medical doctors.

Some potential determinants were not present in the database. Household water, sanitation, and hygiene (WASH) characteristics were limited to information on arsenic contamination of the drinking water, but diarrhea 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 [10]. Paternal height, which may be related to fetal growth, was not available [42]. 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 analyzed and non-analyzed 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 sub-groups in the conditional inference trees. A sub-study 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.

A benefit of applying random forest modelling compared to using conventional models with relative risks or odds ratios is that it ranks the predictors according to how important these are for the explaining the outcome. The random forest analysis, however, does not provide information on whether the predictors have a positive or negative relation to the outcome. The conditional inference trees, on the other hand, display precise information on the priority, size, and direction of the association of the predictors with the outcome. The risk group identification, including the prioritization 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.

A potential limitation of the conditional inference tree method is that if the data contain two essential and highly correlated predictors, the conditional inference tree method may select only one of them in the analysis, although the other predictor might be as important. Further, decision trees do not deliver *p*-values or confidence intervals to the results. The cross-validation method, however, ensures that the selected tree is optimal. This validation method was chosen superior to other model validation methods, e.g., 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 the WHO growth reference and falter dramatically up to 24 months of age [5]. The sub-continents 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.

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 multi-country longitudinal study that found birth or enrollment weight of the infant and maternal height to have the highest cumulative odds ratios for linear growth deficit up to two years of age [10]. These results add to the growing evidence that a large part of linear growth faltering already originates in fetal life [10,43,44]. In a pooled analysis of 19 birth cohorts with longitudinal follow-up, 20% of stunting was attributable to small-for-gestational-age weight at birth [16]. That study did not include any post-natal 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 [44]. Our study included both pre- and post-natal factors and, in contrast to most other studies, assessed not only the relative importance of different potential predictors, but also 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 [15]. Similar results were reported from a study in India [45]. These studies were, however, cross-sectional, without access to birth characteristics.

Maternal height is a strong determinant of fetal growth [46] that indirectly reflect the epigenetic heredity. Maternal height is directly associated with the uterine volume [47], cephalo-pelvic disproportion and subsequent infant and childhood stunting, and child mortality [48,49]. 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 [49]. 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 conditional inference trees 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 [50]. 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 categorized subgroups with a higher probability of stunting and undesirable linear growth trajectories. Poverty is associated with unfavorable 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 [51] 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 emphasized [52].

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 [53]. 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 to other areas with equivalent levels of economic growth and food security [52].

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 [54]. However, there is a considerable risk of not completing primary school for both girls and boys [55]. In 2013, the continuation to the last grade of primary school (5 years) was 75% [56] 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 not only enrolling in but also 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 multi-country pooled analysis of cohort studies showed that a higher cumulative burden of diarrhea increased the risk of stunting [57]. In situations, where measles still occurred, its impact on growth and mortality risks were repeatedly documented [58]. 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 diarrhea [59], which may have reduced the impact on linear

growth. Another factor is the almost universal immunization coverage [60,61] 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 [22,62], further supporting the modest impact of child morbidity on linear growth in our sample [33]. In other settings with lower coverage of diarrhea treatment and immunization, 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 [63]. 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 randomized to appropriate complementary foods [64]. The infant feeding variables included in our analysis ranked low in the random forest analysis and did not show up in any of the conditional inference trees. In spite of the relatively few documented effects of complementary feeding programs on stunting, these interventions are often the priority in efforts to combat stunting.

The nutrition interventions from pre-conception to two years of age currently recommended by the WHO include efforts to ensure exclusive breastfeeding, adequate complementary feeding, appropriate nutritional care of sick and malnourished children and proper intake of vitamin A, iron and iodine for women and children [18]. All of these, except micronutrient supplementation to pregnant women, are focused on the postnatal period from birth up to two 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 nine 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 [65,66] and epigenetic heredity [67], 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[67], especially in more deprived populations [68]. 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 counseling 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 randomized trials evaluating these interventions report their positive impact on birth weight and a reduced risk of LBW [69-76]. Effect sizes vary from increases in birth weight of 20–200g, 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

randomized to an early invitation to supplementation (week 9) or the usual program start of supplementation (week 20). Children of mothers who participated in food supplementation from early pregnancy (versus the usual start) had a 13% reduction in stunting up to five years [62].

There is increasing evidence that preconception interventions may be even more appropriate[77]. A few trials examining the effect of interventions initiated before pregnancy are underway, but few results have so far been published [78]. 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 is needed, especially in South Asia with its high burden of maternal undernutrition and young age at first pregnancy [79]. Targeting high-risk subgroups, in this setting characterized by short, poor, women with low education, can be another strategy to address the intractable problem of stunting.

Contributors

PS contributed to study design, data analysis and interpretation of the results and had the main responsibility of writing the paper. LÅP and SEA were principal investigators of the MINIMat project. ECE, LÅP and KES contributed to the study design. ECE, RN, AR and AIK took part in and supervised data collection. PS, OS, and KES analysed the data. All authors contributed to the preparation of the database, interpretation of the results and reviewed and approved the final version of the manuscript.

Competing interests

The authors declare that they have no competing interests

Data sharing statment

Data are available from the authors upon reasonable request and with permission from the principal investigator of the MINIMat study.

Funding

The MINIMat research study was funded by the icddr,b, United Nations Children's Emergency Fund, Swedish International Development Cooperation Agency (Sida), UK Medical Research Council, Swedish Research Council, Department for International Development (DFID), Japan Society for the Promotion of Science, Child Health and Nutrition Research Initiative, Uppsala University, and US Agency for International Development. The funding agencies had no role in the design and conduct of the study, in the collection, analysis, and interpretation of the data, or in the preparation, review, or approval of the manuscript.

Acknowledgements

We thank the participants and their families in Matlab for their continuing involvement in the MINIMat trial, and the field-team members and data management staff for their excellent work.

381

382

Legend to Figures

Figure 1. Factors, variables and outcomes included in the analysis of data from the MINIMat cohort, Bangladesh. Grouping according to the WHO conceptual framework on childhood stunting [20]

Figure 2. Flow chart of pregnant women and their children included in the data mining analyses of the MINIMat cohort from conception to two years of age.

Figure 3. Height-for-age Z-scores from birth to 24 months in the MINIMat cohort in rural Bangladesh.

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 MINIMat cohort in rural Bangladesh. Colour coding according to the 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 HAZ (Δ HAZ) from birth to 24 months of age. The MINIMat cohort in rural Bangladesh. Colour coding according to the WHO conceptual framework on causes of stunting.

Figure 6. Conditional inference tree identifying sub-groups with different probabilities of stunting at 24 months. The MINIMat cohort in rural Bangladesh.

Figure 7. Conditional inference tree identifying sub-groups with different mean change in HAZ (Δ HAZ=HAZ_{24}–HAZ_{0}) 0–24 months within the MINIMat cohort in rural Bangladesh.

References

1 Pradhan M, Sahn DE, Younger SD. Decomposing world health inequality. *Journal of Health Economics* 2003;**22**:271–93. doi:10.1016/S0167-6296(02)00123-6

2 Victora CG, Adair L, Fall C, *et al.* Maternal and child undernutrition: consequences for adult health and human capital. *Lancet* 2008;**371**:340–57. doi:10.1016/S0140-6736(07)61692-4

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.

4 General Assembly, UNICEF. 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. doi:10.1542/peds.2009-1519

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. doi:10.1016/S0140-6736(07)61690-0

7 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**:S43–52–discussionS52–3.

8 Frongillo EA Jr. *Symposium: Causes and Etiology of Stunting. J Nutr* 1999;**129**:529S–530S. doi:10.1016/S0140-6736(94)91337-4

9 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. doi:10.1016/0035-9203(77)90005-0

10 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. doi:10.1371/journal.pmed.1002408

11 Espo M, Kulmala T, Maleta K, *et al.* Determinants of linear growth and predictors of severe stunting during infancy in rural Malawi. *Acta Paediatrica* 2002;**91**:1364–70. doi:10.1111/j.1651-2227.2002.tb02835.x

12 Rahman A, Chowdhury S. Determinants of Chronic Malnutrtion Among Preschool Children In Bangladesh. *J Biosoc Sci* 2006;**39**:161. doi:10.1017/S0021932006001295

13 MS MRAM, RN MWMPL, BS AFMC. Factors Associated with Stunting Among 0-23 Months-Old Children in Rural Bangladesh. *Journal of Nutrition Education and Behavior* 2017;**49**:S21. doi:10.1016/j.jneb.2017.05.061

14 Hong R, Banta JE, Betancourt JA. Relationship between household wealth inequality and chronic childhood under-nutrition in Bangladesh. *International Journal for Equity in Health* 2006;**5**:1–10. doi:10.1186/1475-9276-5-15

15 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. *Social Science & Medicine* 2017;**187**:144–54. doi:10.1016/j.socscimed.2017.06.017

16 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. doi:10.1093/ije/dyt109

17 World Health Organization. WHO Guidline: Daily iron and folic acid supplementation in pregnant women. *WHO* 2012;:1–32.

18 World Health Organization. Essential Nutrition Actions: improving maternal, newborn, infant and young child health and nutrition. *WHO* 2013;:1–144.

19 Bhutta ZA, Das JK, Arjumand R, *et al.* Maternal and Child Nutrition 2 Evidence-based interventions for improvement of maternaland child nutrition: what can be done and at what cost? *The Lancet* 2013;**382**:452–77. doi:10.1016/S0140-6736(13)60996-4

20 Stewart CP, Iannotti L, Dewey KG, *et al.* Contextualising complementary feeding in a broader framework for stunting prevention. *Matern Child Nutr* 2013;**9**:27–45. doi:10.1111/mcn.12088

21 Breiman L, Friedman J, Stone CJ, *et al. Classification and Regression Trees*. Chapman and Hall/CRC 1984.

22 Persson LÅ, 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. doi:10.1001/jama.2012.4061

23 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 Paediatrica* 2013;**102**:815–23. doi:10.1111/apa.12282

24 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.

25 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;:4158–62.

26 Straus MA, Douglas EM. A Short Form of the Revised Conflict Tactics Scales, and Typologies for Severity and Mutuality. *Violence and Victims* 2004;**19**:1–15.

27 WHO. WHO Multi-country Study on Women's Health and Domestic Violence against Women. *WHO* 2005;:1–38.

28 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. doi:10.1136/adc.2008.144444

29 Rahman A, Vahter M, Smith AH, *et al.* Arsenic Exposure During Pregnancy and Size at Birth: A Prospective Cohort Study in Bangladesh. *American Journal of Epidemiology* 2008;**169**:304–12. doi:10.1093/aje/kwn332

30 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.

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

32 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;:1075–82.

33 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 Paediatrica* 2017;**382**:427–8. doi:10.1111/apa.14009

34 Arnold BF, Galiani S, Ram PK, *et al.* Optimal Recall Period for Caregiver-reported Illness in Risk Factor and Intervention Studies: A Multicountry Study. *American Journal of Epidemiology* 2013;**177**:361–70. doi:10.1093/aje/kws281

35 Onis M de, Onyango AW, Borghi E, *et al.* Development of a WHO growth reference for school-aged children and adolescents. *Bulletin of the World Health Organization* 2007;**85**:660–7. doi:10.1590/S0042-96862007000900010

36 Jönsson P, Wohlin C. An Evaluation of K-Nearest Neighbour Imputation Using Likert Data. *Proceedings of th International Symposium on Software Metrics* 2004;:108–18.

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

38 Torsten Hothorn KHAZ. ctree: Conditional Inference Trees. www.cran.r-project.org. 2016;:1–21.https://cran.r-project.org/web/packages/partykit/vignettes/ctree.pdf (accessed 24 Mar2017).

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

40 R Core Team. R: A language and environment for stastical computing. 2017. httpa://www.R-project.org/

41 Achim Zeileis THKH. party: A Laboratory for Recursive Partytioning. 2017;:1–18.

42 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 Human Development* 2010;**86**:535–40. doi:10.1016/j.earlhumdev.2010.07.002

43 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. doi:10.1093/ije/dyt109

44 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;:2202–7.

45 Corsi DJ, Mejía-Guevara I, Subramanian SV. Risk factors for chronic undernutrition among children in India: Estimating relative importance, population attributable risk and fractions. *Social Science & Medicine* 2016;**157**:165–85. doi:10.1016/j.socscimed.2015.11.014

46 Kramer MS. The Epidemiology of Adverse Pregnancy Outcomes: An Overview. *J Nutr* 2003;:1592S–1596S.

47 A WHO Collaborative Study of Maternal Anhropomehtry and Pregnancy Outcomes. *International Journal of Gynecology and Obstetrics* 1996;**57**:1–15.

48 Özaltin E. Association of Maternal Stature With Offspring Mortality, Underweight, and Stunting in Low- to Middle-Income Countries. *JAMA* 2010;**303**:1507–21. doi:10.1001/jama.2010.450

49 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. doi:10.1371/journal.pone.0149700

50 Van de Poel E, Hosseinpoor AR, Speybroeck N, *et al.* Socioeconomic inequality in malnutrition in developing countries. *Bulletin of the World Health Organization* 2008;**86**:282–91. doi:10.1590/S0042-96862008000400013

51 Bradley RH, Corwyn RF. Socioeconomic Status and Child Development. *Annual Review of Phsycology* 2002;**53**:371–99.

52 Ramalingaswami V, U J, J R. The Asian enigma. 1996.

53 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. doi:10.1111/j.1740-8709.2008.00161.x

54 Chowdhury AMR, Bhuiya A, Chowdhury ME, *et al.* The Bangladesh paradox: exceptional health achievement despite economic poverty. *The Lancet* 2013;**382**:1734–45. doi:10.1016/S0140-6736(13)62148-0

55 Asadullah MN, Chaudhury N. Girls school enrolement figures are good in Bangladesh but drop out rates are worrying. *Development and Cooperation* 2017;**59**:447–72. doi:10.1086/681929

56 UNESCO. EFA 2015 national review: Bangladesh; 2014. unesdoc.unesco.org. http://unesdoc.unesco.org/images/0023/002305/230507e.pdf (accessed 13 Feb2018).

57 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. doi:10.1093/ije/dyn099

58 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.

59 Arifeen DrPH El S, MPH AC, ScD LR, *et al.* Community-based approaches and partnerships: innovations in health-service delivery in Bangladesh. *The Lancet* 2013;**382**:2012–26. doi:10.1016/S0140-6736(13)62149-2

60 WHO, UNICEF. Bangladesh: WHO and UNICEF estimates of immunization coverage: 2016 revision. 2016;:1–30.

61 Chowdhury AMR, Bhuiya A, Chowdhury ME, *et al.* The Bangladesh paradox: exceptional health achievement despite economic poverty. *The Lancet* 2013;**382**:1734–45. doi:10.1016/S0140-6736(13)62148-0

62 Khan AI, Kabir I, Ekstrom EC, *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.

63 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.

64 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**:S25. doi:10.1186/1471-2458-11-S3-S25

65 Weedon MN, Frayling TM. Reaching new heights: insights into the genetics of human stature. *Trends in Genetics* 2008;**24**:595–603. doi:10.1016/j.tig.2008.09.006

66 Silventoinen K, Sammalisto S, Perola M, *et al.* Heritability of Adult Body Height: A Comparative Study of Twin Cohorts in Eight Countries. *Twin Research* 2003;**6**:399–408. doi:10.1375/twin.6.5.399

67 Simeone P, Alberti S. Epigenetic heredity of human height. *Physiological Reports* 2014;**2**:e12047–7. doi:10.14814/phy2.12047

68 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. *American Journal of Public Health* 2000;**90**:627–30.

69 Imdad A, Bhutta ZA. Effect of balanced protein energy supplementation during pregnancy on birth outcomes. *BMC Public Health* 2011;**11**:S17. doi:10.1186/1471-2458-11-S3-S17

70 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**:S533–46. doi:10.1177/15648265090304S408

71 Ramakrishnan U, Grant FK, Goldenberg T, *et al.* Effect of Multiple Micronutrient Supplementation on Pregnancy and Infant Outcomes: A Systematic Review. *Paediatric and Perinatal Epidemiology* 2012;**26**:153–67. doi:10.1111/j.1365-3016.2012.01276.x

72 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. doi:10.1001/jama.2014.16819

73 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. doi:10.1111/mcn.12367

74 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. doi:10.3945/jn.114.207225

75 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. doi:10.3945/ajcn.114.091546

76 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. *American Journal of Clinical Nutrition* 2009;**90**:1593–600. doi:10.3945/ajcn.2009.28253

77 Gluckman PD, Hanson MA, Cooper C, *et al.* Effect of In Utero and Early-Life Conditions on Adult Health and Disease. *http://dxdoiorg/101056/NEJMra0708473* 2009;**359**:61–73. doi:10.1056/NEJMra0708473

78 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). *American Journal of Clinical Nutrition* 2014;**100**:1257–68. doi:10.3945/ajcn.114.084921

79 The World Bank. World Development Indicators: Reproductive health. 2016;:1–6.

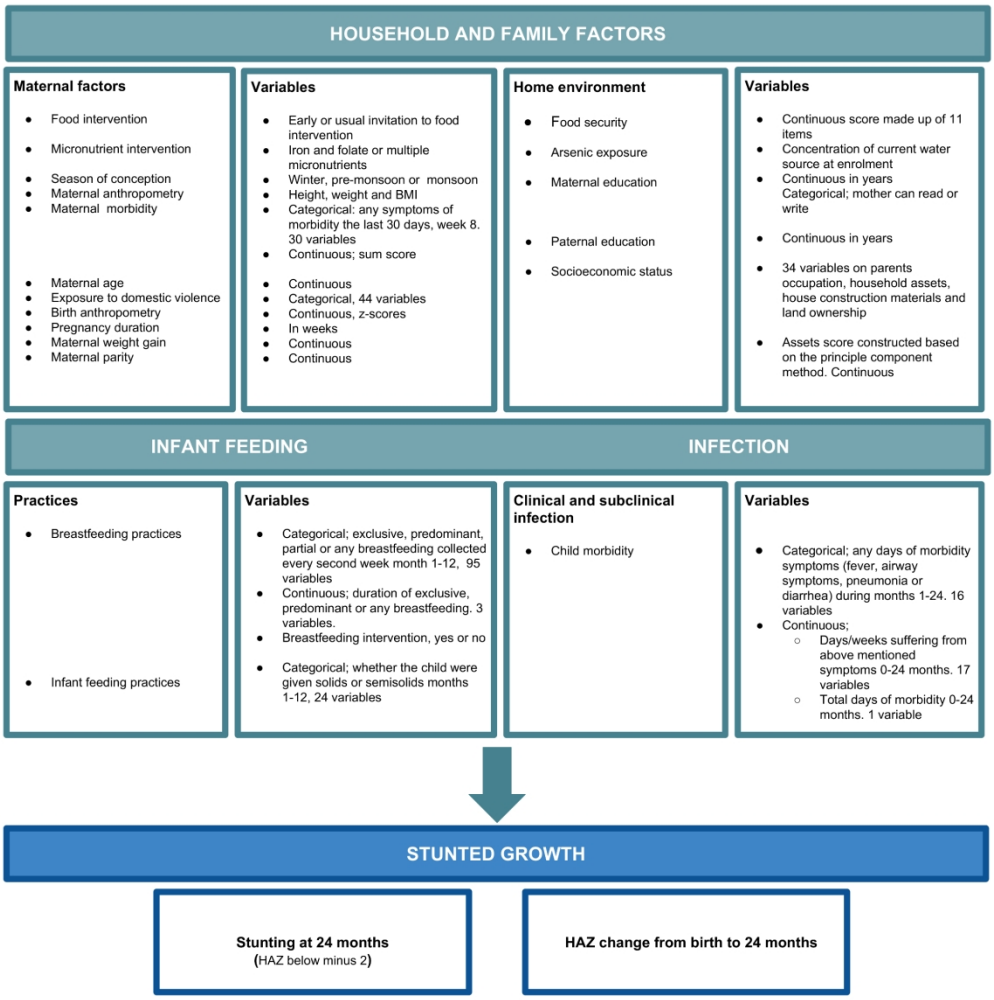

Figure 1. Factors, variables and outcomes included in the analysis of data from the MINIMat cohort, Bangladesh. Grouping according to the WHO conceptual framework on childhood stunting [20]

190x190mm (300 x 300 DPI)

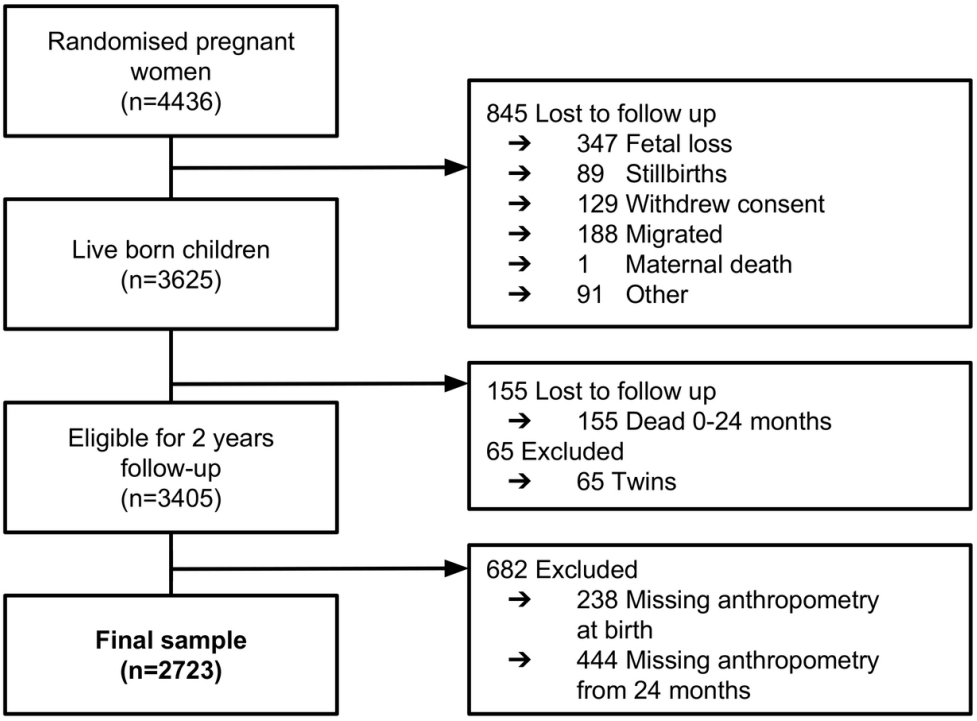

Figure 2. Flow chart of pregnant women and their children included in the data mining analyses of the MINIMat cohort from conception to two years of age.

106x80mm (300 x 300 DPI)

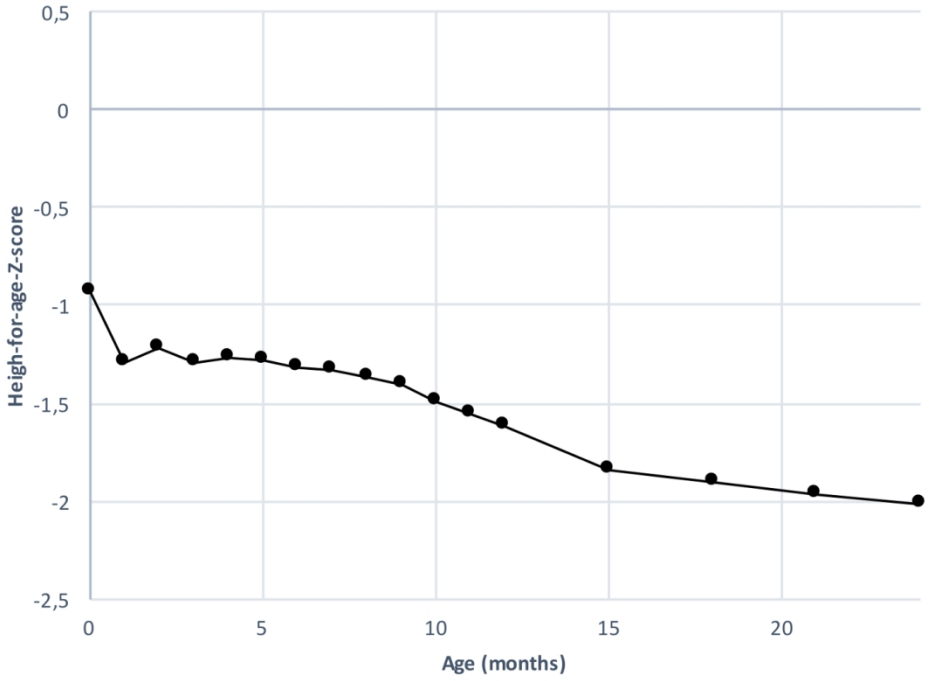

Figure 3. Height-for-age Z-scores from birth to 24 months in the MINIMat cohort in rural Bangladesh.

131x96mm (300 x 300 DPI)

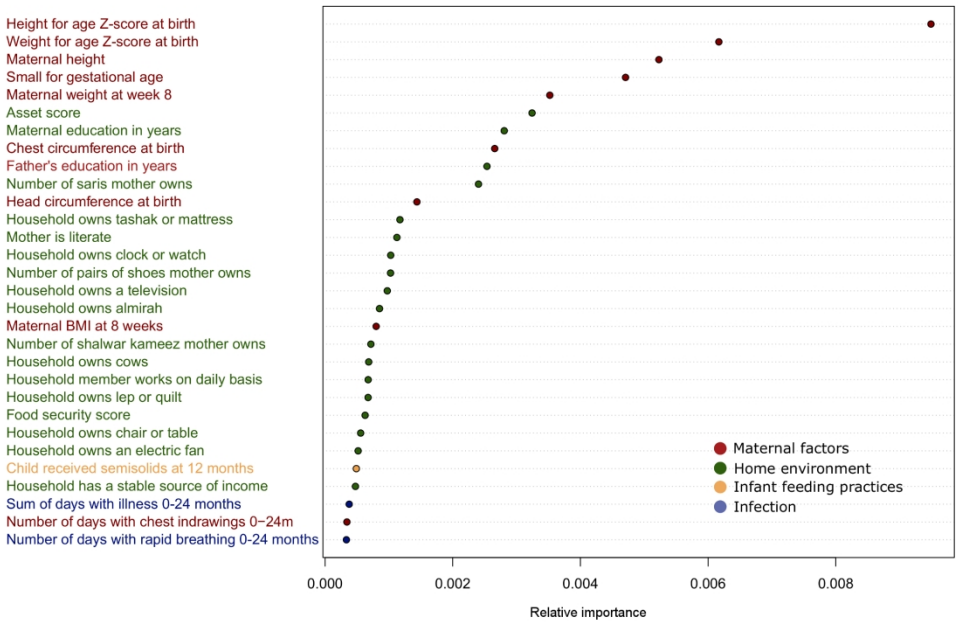

Figure 4. Conditional random forest plot ranking the relative importance of 30 predictors with regard to their ability to explain the presence of stunting at 24 months of age. The MINIMat cohort in rural Bangladesh. Colour coding according to the WHO conceptual framework on causes of stunting.

190x134mm (300 x 300 DPI)

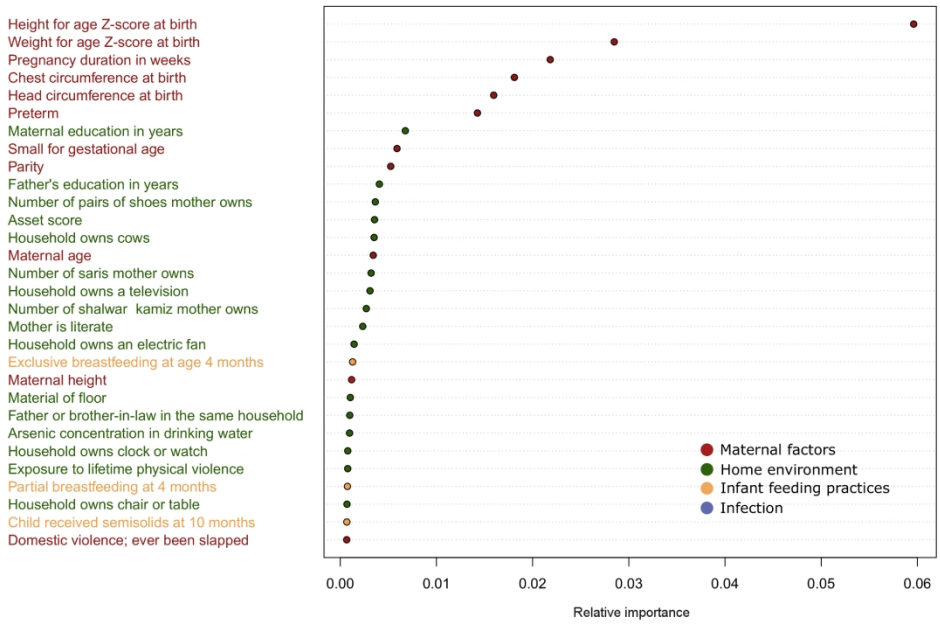

Figure 5. Conditional random forest plot ranking the relative importance of 30 predictors with regard to their ability to explain the variation in change in HAZ (Δ HAZ) from birth to 24 months of age. The MINIMat cohort in rural Bangladesh. Colour coding according to the WHO conceptual framework on causes of stunting.

190x134mm (300 x 300 DPI)

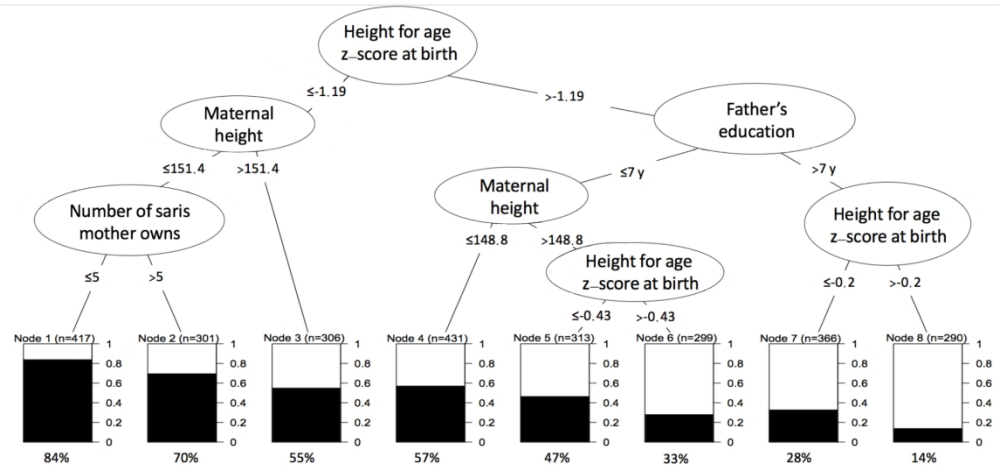

Figure 6. Conditional inference tree identifying sub-groups with different probabilities of stunting at 24 months. The MINIMat cohort in rural Bangladesh.

190x88mm (300 x 300 DPI)

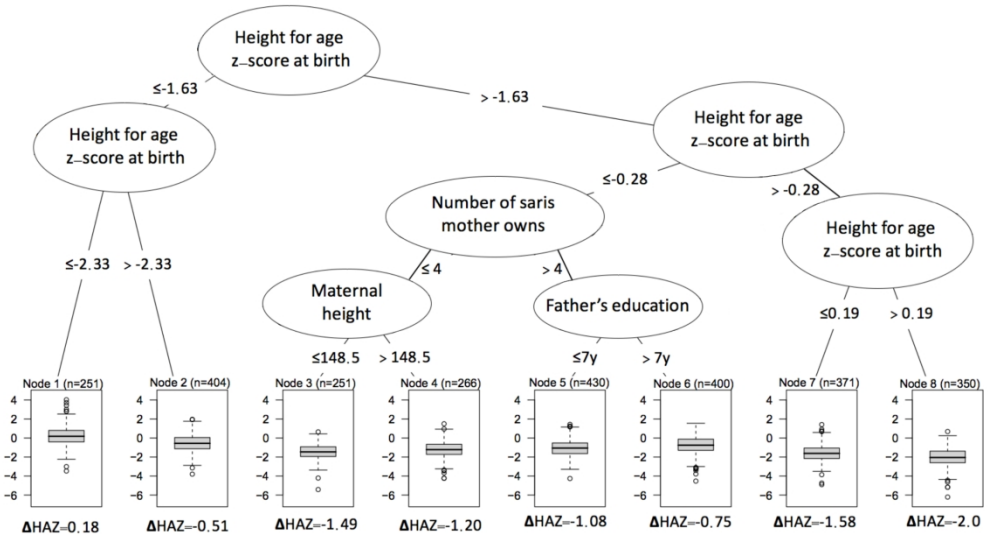

Figure 7. Conditional inference tree identifying sub-groups with different mean change in HAZ (Δ HAZ=HAZ24–HAZ0) 0–24 months within the MINIMat cohort in rural Bangladesh.

190x101mm (300 x 300 DPI)

Supplementation appendix

Simulation study of the predictive performance of three different imputation methods

The following strategy was used to study the imputation accuracy of various methods for the input variables in our analyses. First, we standardized numerical variables in the data and took a sample of the entire data (α) and deleted a proportion (β) of the non-missing values in each variable. Secondly, we employed three different imputation methods to make predictions of the missing values in the data. Lastly, we compared the predictions with the values of the deleted entries, the computed mean-square error (MSE) for the numerical variables, and the percent of the incorrect predictions, misclassification rate (MR), for the categorical ones. The computation of the MSE and MR values was repeated several times for different samples of the original data. The summary results of these computations are presented in Tables 1-4. It can be concluded that random forests[1] provided a statistically significantly better imputation than the variable mean and K-nearest neighbor imputation methods. The design of the study followed a procedure similar to the strategy described in Jonsson et al [2].

Table 1: Means and Standard errors of the MR[2] and the MSE[3] for different imputation methods, computed from m=100 samples, $\alpha = 0.05$, $\beta = 0.05$

	Variable mean	KNN[1]	Random forest
Mean (MR[2])	0.17755631	0.187499573	0.131724506
Standard Error (MR[2])	0.00360524	0.003795385	0.003759032
Mean (MSE[3])	1.01903348	0.901518114	0.541867921
Standard error (MSE[3])	0.01640172	0.016414433	0.015157205

[1] K-nearest neighbour
[2]Missclassification rate
[3]Mean square error
α = proportion of the non-missing values deleted
β = proportion of the original data sampled

Table 2: Means and Standard errors of the MR[2] and the MSE[3] for different imputation methods, computed from m=100 samples, $\alpha = 0.05$, $\beta = 0.15$

	Variable mean	KNN[1]	Random forest
Mean (MR[2])	0.175774830	0.187158897	0.131724506
Standard Error (MR[2])	0.003075253	0.003317242	0.003302446
Mean (MSE[3])	1.00474998	0.922010327	0.556762189
Standard error (MSE[3])	0.01012910	0.009595471	0.008949707

[1] K-nearest neighbour
[2]Missclassification rate
[3]Mean square error
α = proportion of the non-missing values deleted
β = proportion of the original data sampled

Table 3: Means and Standard errors of the MR[2] and the MSE[3] for different imputation methods, computed from m=100 samples, $\alpha = 0.2$, $\beta = 0.05$

	Variable mean	KNN[1]	Random forest
Mean (MR[2])	0.1625007370	0.1608280983	0.094319580
Standard Error (MR[2])	0.0005210379	0.0005181798	0.000367369
Mean (MSE[3])	1.0023969039	0.7975006166	0.450253626
Standard error (MSE[3])	0.0068209597	0.0066997794	0.006069386

[1] K-nearest neighbour
[2]Missclassification rate
[3]Mean square error
α = proportion of the non-missing values deleted
β = proportion of the original data sampled

Table 4: Means and Standard errors of discrete and continuous variables for different imputation methods. Computed from m=100 samples, $\alpha = 0.2$, $\beta = 0.15$

	Variable mean	KNN[1]	Random forest
Mean, discrete	0.1626095174	0.1617267853	0.1017561946
Standard error, Discrete	0.0003670347	0.0003618961	0.0002612874
Mean, continuous	0.9984641615	0.8195273545	0.4593241548
Standard error, continuous	0.0040175223	0.0040319899	0.0034449935

[1] K-nearest neighbour
[2]Missclassification rate
[3]Mean square error
α = proportion of the non-missing values deleted
β = proportion of the original data sampled

References

1. Stekhoven DJ, Buhlmann P. MissForest—non-parametric missing value imputation for mixed-type data. Bioinformatics. 2012;: 112–118.

2. Jönsson P, Wohlin C. An Evaluation of K-Nearest Neighbour Imputation Using Likert Data. Proceedings of th International Symposium on Software Metrics. 2004;: 108–118.

STROBE Statement—checklist of items that should be included in reports of observational studies

	Item No	Recommendation	
Title and abstract	1	(*a*) Indicate the study's design with a commonly used term in the title or the abstract	1
		(*b*) Provide in the abstract an informative and balanced summary of what was done and what was found	2
Introduction			
Background/rationale	2	Explain the scientific background and rationale for the investigation being reported	4-5
Objectives	3	State specific objectives, including any prespecified hypotheses	6
Methods			
Study design	4	Present key elements of study design early in the paper	5
Setting	5	Describe the setting, locations, and relevant dates, including periods of recruitment, exposure, follow-up, and data collection	6
Participants	6	(*a*) *Cohort study*—Give the eligibility criteria, and the sources and methods of selection of participants. Describe methods of follow-up *Case-control study*—Give the eligibility criteria, and the sources and methods of case ascertainment and control selection. Give the rationale for the choice of cases and controls *Cross-sectional study*—Give the eligibility criteria, and the sources and methods of selection of participants	6
		(*b*) *Cohort study*—For matched studies, give matching criteria and number of exposed and unexposed *Case-control study*—For matched studies, give matching criteria and the number of controls per case	
Variables	7	Clearly define all outcomes, exposures, predictors, potential confounders, and effect modifiers. Give diagnostic criteria, if applicable	9 Figure 1
Data sources/ measurement	8*	For each variable of interest, give sources of data and details of methods of assessment (measurement). Describe comparability of assessment methods if there is more than one group	*7-9*
Bias	9	Describe any efforts to address potential sources of bias	7-9
Study size	10	Explain how the study size was arrived at	Not applicable
Quantitative variables	11	Explain how quantitative variables were handled in the analyses. If applicable, describe which groupings were chosen and why	7-9
Statistical methods	12	(*a*) Describe all statistical methods, including those used to control for confounding	10-11
		(*b*) Describe any methods used to examine subgroups and interactions	10-11
		(*c*) Explain how missing data were addressed	10-11
		(*d*) *Cohort study*—If applicable, explain how loss to follow-up was addressed *Case-control study*—If applicable, explain how matching of cases and controls was addressed	10

Cross-sectional study—If applicable, describe analytical methods taking account of sampling strategy	
(*e*) Describe any sensitivity analyses	10

Continued on next page10

Results			
Participants	13*	(a) Report numbers of individuals at each stage of study—eg numbers potentially eligible, examined for eligibility, confirmed eligible, included in the study, completing follow-up, and analysed	Fig 2 12
		(b) Give reasons for non-participation at each stage Fig 2	Fig 2
		(c) Consider use of a flow diagram Figure 2	Fig 2
Descriptive data	14*	(a) Give characteristics of study participants (eg demographic, clinical, social) and information on exposures and potential confounders	Table 1, 13
		(b) Indicate number of participants with missing data for each variable of interest	13
		(c) *Cohort study*—Summarise follow-up time (eg, average and total amount)	13
Outcome data	15*	*Cohort study*—Report numbers of outcome events or summary measures over time	*13*
		Case-control study—Report numbers in each exposure category, or summary measures of exposure	
		Cross-sectional study—Report numbers of outcome events or summary measures	
Main results	16	(*a*) Give unadjusted estimates and, if applicable, confounder-adjusted estimates and their precision (eg, 95% confidence interval). Make clear which confounders were adjusted for and why they were included	14
		(*b*) Report category boundaries when continuous variables were categorized	14
		(*c*) If relevant, consider translating estimates of relative risk into absolute risk for a meaningful time period	
Other analyses	17	Report other analyses done—eg analyses of subgroups and interactions, and sensitivity analyses	14
Discussion			
Key results	18	Summarise key results with reference to study objectives	15,18
Limitations	19	Discuss limitations of the study, taking into account sources of potential bias or imprecision. Discuss both direction and magnitude of any potential bias	15,16,17
Interpretation	20	Give a cautious overall interpretation of results considering objectives, limitations, multiplicity of analyses, results from similar studies, and other relevant evidence	18-22
Generalisability	21	Discuss the generalisability (external validity) of the study results	17
Other information			
Funding	22	Give the source of funding and the role of the funders for the present study and, if applicable, for the original study on which the present article is based	23

*Give information separately for cases and controls in case-control studies and, if applicable, for exposed and unexposed groups in cohort and cross-sectional studies.

Note: An Explanation and Elaboration article discusses each checklist item and gives methodological background and published examples of transparent reporting. The STROBE checklist is best used in conjunction with this article (freely available on the Web sites of PLoS Medicine at http://www.plosmedicine.org/, Annals of Internal Medicine at http://www.annals.org/, and Epidemiology at http://www.epidem.com/). Information on the STROBE Initiative is available at www.strobe-statement.org.

BMJ Open

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

Journal:	*BMJ Open*
Manuscript ID	bmjopen-2018-025154.R2
Article Type:	Research
Date Submitted by the Author:	07-Jun-2019
Complete List of Authors:	Svefors, Pernilla; Uppsala Universitet, Women's and children's health; Center for epidemiology and community medicine, Region of Stockholm Sysoev, Oleg; Linkopings universitet, Department of Computer and Information Sciences Ekstrom, Eva- Charlotte; Uppsala Universitet, Women's and children's health Persson, Lars Ake; London School of Hygiene and Tropical Medicine, Department of Disease control Arifeen, Shams E.; International Centre for Diarrhoeal Disease Research, Maternal and Child Health Division Naved, Ruchira; International Centre for Diarrhoeal Disease Research Rahman, Anisur ; International Centre for Diarrhoeal Disease Research, Maternal and Child Health Division Khan, Ashraful Islam; International Centre for Diarrhoeal Disease Research Selling, Katarina; Uppsala Universitet, Women's and children's health
Primary Subject Heading:	Global health
Secondary Subject Heading:	Paediatrics, Nutrition and metabolism
Keywords:	Nutrition < TROPICAL MEDICINE, PUBLIC HEALTH, EPIDEMIOLOGY, PAEDIATRICS

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

Pernilla Svefors[1*], Oleg Sysoev[2], Eva-Charlotte Ekström[1], Lars-Åke Persson[3], Shams El Arifeen[4], Ruchira Naved[4], Anisur Rahman[4], Ashraful Islam Khan[4], Katarina Ekholm Selling[1]

[1] International Maternal and Child Health, Department of Women's and Children's Health, Uppsala University, Uppsala, Sweden. [2] Department of Computer and Information Sciences, Linköping University, Linköping, Sweden, [3] Department of Disease Control, London School of Hygiene and Tropical Medicine, London, UK, [4] International Centre for Diarrhoeal Disease Research, Bangladesh (icddr,b), Dhaka, Bangladesh.

*Corresponding author pernilla.svefors@kbh.uu.se

Abstract

Introduction The 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 pre- and postnatal determinants of linear growth 0–24 months and the risk factors for stunting at two years, and to identify subgroups with different growth trajectories and levels of stunting at two years.

Methods Conditional inference-tree-based methods were applied to the extensive Maternal and Infant Nutrition Interventions in Matlab (MINIMat) trial database with 309 variables of 2,723 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 two 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 two years.

Conclusion The results of this study emphasize 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 pre 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

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 adult chronic diseases [3]. In 2012, the 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 two years of age, i.e., 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 emphasized the personalized perspective under the universal coverage of health care. 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 pre- and post-natal factors associated with impaired growth [8-13]. Low birth weight, maternal height, maternal education, poverty and inadequate complementary feeding practices have been recognized as important risk factors [14-16]. Some analyses emphasize the importance of fetal growth restriction for later stunted growth, but rarely is the relative importance of pre- and post-natal factors assessed [17]. Despite these findings, policy documents and recommendations emphasize interventions especially after birth, and pre-natal 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 visualized in the 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 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 randomized 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 pre- and post-natal anthropometric assessments, socioeconomic and biological characteristics of the mother and father, information on maternal exposure to violence, household food security, breast- 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 0–24 months and risk factors for stunting at two years, and to identify risk groups with negative growth trajectories and high prevalence of stunting at two years.

Methods

Study setting, participants and study design

The MINIMat trial (Maternal and Infant Nutrition Interventions in Matlab, 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 randomized trial primarily evaluating the effect of an early invitation to prenatal food supplementation (versus usual timing) combined with multiple micronutrient supplementation (versus usual program iron-folate) to pregnant women on maternal hemoglobin, birth weight, gestational age at birth, and infant mortality [23]. Further, the participating women were randomly assigned to either counselling for exclusive breastfeeding 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 two 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 the 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 eleven 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 thirty morbidity symptoms from twelve different categories, including airway, urinary tract, fever, circulation, bowel, or pain symptoms during the last month. A sum score ranging from zero to twelve was created based on absence of symptoms or those not recorded for each category.

Home visits were followed by clinic visits at local health sub-centers. Maternal height and weight were measured at around eight weeks of gestation using a stadiometer to the nearest 0.1 cm and an electronic scale (Uniscale; SECA) 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 the WHO collaborative study questionnaire was used [27,28], based on the conflict tactic scale covering physical, sexual and emotional violence and controlling behavior [29]. Household drinking water was analyzed 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 two-year study period, the mother-and-child pairs were visited monthly in their homes during the first year, and every three months during the second year. On these occasions, data on infant feeding practices, child morbidity and anthropometry were collected. The mothers were interviewed about breastfeeding and complementary feeding practices. Breastfeeding practices were categorized into exclusive, predominant, partial, or any breastfeeding for each month from one to twelve months. The total time for exclusive, predominant, and any breastfeeding was calculated. The WHO recommendations guided the breastfeeding assessment [32] and results were validated with a stable-isotope technique. The classification of exclusive breastfeeding 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 categorized into semi-solids and solids each month from one to twelve 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 had any of the following symptoms during the last week; fever, cough, difficult breathing, chest in-drawing, rapid breathing, diarrhea, bloody diarrhea and the duration of these symptoms [34]. Categories were created based on whether the child had suffered from fever, respiratory symptoms, suspected pneumonia, or diarrhea, 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 seven 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, Hamburg, Germany) 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 two years, height was measured to the nearest 0.1 cm, using a freestanding stadiometer. Head and chest circumference were 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 standardization of anthropometric measurements, was conducted periodically.

Outcomes

Height-for-age z-scores (HAZ) were calculated from the measured length and height data using the program WHOAnthro, based on the WHO growth reference for children [36]. Children with a HAZ below minus two SD-scores were classified as stunted. Two outcomes were analyzed: 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 i.e. ΔHAZ = HAZ at 24 months - HAZ at birth.

Statistical analysis

A database was created with 309 variables characterizing 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 sub-set of records that had height measurements at birth and 24 months was selected (n=2 723). The average percent of missing values among all the predictors were 4 %. The highest percent missing were 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 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 neighbor imputation [37], and random forest imputation [38]. The design of the study followed a procedure similar to the strategy described in Jonsson et al. [37], see S 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 data mining 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 (CART) have been shown to be biased [39]. This motivated us to select the Conditional Inference Trees (CIT) framework, a method that embeds a statistical hypothesis-testing framework into a recursive partitioning algorithm used for model building [39]. Conditional inference trees were used in order to

identify sub-groups characterized 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 data set 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 conditional inference trees is created by means of drawing subsamples from the original data and fitting a unique randomized conditional inference tree 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 version 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.

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 two years and 682 were excluded because of missing anthropometry, at birth or at two years, resulting in 2723 children available for analysis (Figure 2). In the non-analyzed group, there was a slightly higher percentage of mothers with more than five 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 eight 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 BMI below 18·5 at pregnancy week eight. 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 two years of age with a mean change of -1 HAZ, resulting in a mean HAZ at two years of -2·0 (Figure 3) and 50% being stunted (girls 51·1%, boys 48·5%)

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 MINIMat 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 (>5y)	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 (>5y)	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 (SGA)	1606/2723 (59.0)	972/1606 (60.5)	-1.26
Appropriate for Gestational Age (AGA)	1117/2723 (41.0)	397/1117 (35.5)	-0.94
Low Birth Weight (LBW)	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

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 are presented in Figure 4 and 5. HAZ and weight-for-age Z-scores (WAZ) at birth were the most important predictors of stunting at 24 months, followed by maternal height, Small for Gestational Age (SGA), maternal weight at eight 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 conditional inference trees presented in Figure 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 conditional inference trees 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 seven 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 characterizing household, environmental, biological, and socioeconomic 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 two years. The difference between the identified subgroups of children with the highest and lowest probabilities of stunting was as 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 multi-country longitudinal study that found birth or enrollment weight of the infant and maternal height to have the highest cumulative odds ratios for linear growth deficit up to two years of age [11]. These results add to the growing evidence that a large part of linear growth faltering already originates in fetal life [11,43,44]. In a pooled analysis of 19 birth cohorts with longitudinal follow-up, 20% of stunting was attributable to small-for-gestational-age weight at birth [17]. That study did not include any post-natal 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 [44]. Our study

included both pre- and post-natal factors and, in contrast to most other studies, assessed not only the relative importance of different potential predictors, but also 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 [45]. These studies were, however, cross-sectional, without access to birth characteristics.

Maternal height is a strong determinant of fetal growth [46] that indirectly reflect the epigenetic heredity. Maternal height is directly associated with the uterine volume [47], cephalo-pelvic disproportion and subsequent infant and childhood stunting, and child mortality [48,49]. 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 [49]. 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 conditional inference trees 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 [50]. 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 categorized subgroups with a higher probability of stunting and undesirable linear growth trajectories. Poverty is associated with unfavorable 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 [51] 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 emphasized [52]. 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 [53]. 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 to other areas with equivalent levels of economic growth and food security [52].

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 [54]. However, there is a considerable risk of not completing primary school for both girls and boys [55]. In 2013, the continuation to the last grade of primary school (5 years) was 75% [56] 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 not only enrolling in but also 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 multi-country pooled analysis of cohort studies showed that a higher cumulative burden of diarrhea increased the risk of stunting [57]. In situations, where measles still occurred, its impact on growth and mortality risks were repeatedly documented [58]. 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 diarrhea [59], which may have reduced the impact on linear growth. Another factor is the almost universal immunization coverage [60,61] 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 [23,62], further supporting the modest impact of child morbidity on linear growth in our sample [34]. In other settings with lower coverage of diarrhea treatment and immunization, 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 [63]. 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 randomized to appropriate complementary foods [64]. The infant feeding variables included in our analysis ranked low in the random forest analysis and did not show up in any of the conditional inference trees. In spite of the relatively few documented effects of complementary feeding programs on stunting, these interventions are often the priority in efforts to combat stunting.

The nutrition interventions from pre-conception to two years of age currently recommended by the WHO include efforts to ensure exclusive breastfeeding, 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 two 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 nine 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 [65,66] and epigenetic heredity [67], 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[67], especially in more deprived populations [68]. 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 counseling 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

randomized trials evaluating these interventions report their positive impact on birth weight and a reduced risk of LBW [69-76]. Effect sizes vary from increases in birth weight of 20–200g, 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 randomized to an early invitation to supplementation (week 9) or the usual program start of supplementation (week 20). Children of mothers who participated in food supplementation from early pregnancy (versus the usual start) had a 13% reduction in stunting up to five years [62].

There is increasing evidence that preconception interventions may be even more appropriate[77]. A few trials examining the effect of interventions initiated before pregnancy are underway, but few results have so far been published [78]. 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 is needed, especially in South Asia with its high burden of maternal undernutrition and young age at first pregnancy [79]. Targeting high-risk subgroups, in this setting characterized 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 two years. The MINIMat cohort was implemented in an excellent research infrastructure that fulfills 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 two years. They received repeated training, including standardization exercises, and were supervised by senior medical doctors.

Some potential determinants were not present in the database. Household water, sanitation, and hygiene (WASH) characteristics were limited to information on arsenic contamination of the drinking water, but diarrhea 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 [80]. 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 analyzed and non-analyzed 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 sub-groups in the conditional inference trees. A sub-study 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, etc. 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 data mining 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, Conditional Inference Trees, 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 prioritization 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. Firstly, 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. Secondly, 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 confidence intervals. The cross-validation method, however, ensures that the selected tree is optimal. This validation method was chosen superior to other model validation methods, e.g., 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 the WHO growth reference and falter dramatically up to 24 months of age [5]. The sub-continents 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 pre- 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 emphasize the benefit of interventions before conception and during pregnancy to reach a substantial reduction in stunting.

Contributors

PS contributed to study design, data analysis and interpretation of the results and had the main responsibility of writing the paper. LÅP and SEA were principal investigators of the MINIMat project. ECE, LÅP and KS contributed to the study design. ECE, RN, AR and AIK took part in and supervised data collection. PS, OS, and KS analysed the data. All authors contributed to the preparation of the database, interpretation of the results and reviewed and approved the final version of the manuscript.

Competing interests

The authors declare that they have no competing interests

Data sharing statement

Data are available from the authors upon reasonable request and with permission from the principal investigator of the MINIMat study.

Funding

The MINIMat research study was funded by the icddr,b, United Nations Children's Emergency Fund, Swedish International Development Cooperation Agency (Sida), UK Medical Research Council, Swedish Research Council, Department for International Development (DFID), Japan Society for the Promotion of Science, Child Health and Nutrition Research Initiative, Uppsala University, and US Agency for International Development. The funding agencies had no role in the design and conduct of the study, in the collection, analysis, and interpretation of the data, or in the preparation, review, or approval of the manuscript.

Acknowledgements

We thank the participants and their families in Matlab for their continuing involvement in the MINIMat trial, and the field-team members and data management staff for their excellent work.

410

411

Legend to Figures

Figure 1. Factors, variables and outcomes included in the analysis of data from the MINIMat cohort, Bangladesh. Grouping according to the WHO conceptual framework on childhood stunting [21]

Figure 2. Flow chart of pregnant women and their children included in the data mining analyses of the MINIMat cohort from conception to two years of age.

Figure 3. Height-for-age Z-scores from birth to 24 months in the MINIMat cohort in rural Bangladesh.

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 MINIMat cohort in rural Bangladesh. Colour coding according to the 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 HAZ (Δ HAZ) from birth to 24 months of age. The MINIMat cohort in rural Bangladesh. Colour coding according to the WHO conceptual framework on causes of stunting.

Figure 6. Conditional inference tree identifying sub-groups with different probabilities of stunting at 24 months. The MINIMat cohort in rural Bangladesh.

Figure 7. Conditional inference tree identifying sub-groups with different mean change in HAZ (Δ HAZ=HAZ_{24}–HAZ_{0}) 0–24 months within the MINIMat cohort in rural Bangladesh.

References

1 Pradhan M, Sahn DE, Younger SD. Decomposing world health inequality. *Journal of Health Economics* 2003;**22**:271–93. doi:10.1016/S0167-6296(02)00123-6

2 Victora CG, Adair L, Fall C, *et al.* Maternal and child undernutrition: consequences for adult health and human capital. *Lancet* 2008;**371**:340–57. doi:10.1016/S0140-6736(07)61692-4

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.

4 General Assembly, UNICEF. 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. doi:10.1542/peds.2009-1519

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. doi:10.1016/S0140-6736(07)61690-0

7 Bayer R, Galea S. Public Health in the Precision-Medicine Era. - PubMed - NCBI. *N Engl J Med* 2015;**373**:499–501. doi:10.1056/NEJMp1506241

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**:S43–52–discussionS52–3.

9 Frongillo EA Jr. *Symposium: Causes and Etiology of Stunting. J Nutr* 1999;**129**:529S–530S. doi:10.1016/S0140-6736(94)91337-4

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. doi:10.1016/0035-9203(77)90005-0

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. doi:10.1371/journal.pmed.1002408

12 Espo M, Kulmala T, Maleta K, *et al.* Determinants of linear growth and predictors of severe stunting during infancy in rural Malawi. *Acta Paediatrica* 2002;**91**:1364–70. doi:10.1111/j.1651-2227.2002.tb02835.x

13 Rahman A, Chowdhury S. Determinants of Chronic Malnutrtion Among Preschool Children In Bangladesh. *J Biosoc Sci* 2006;**39**:161. doi:10.1017/S0021932006001295

14 Amin R, Murumi MW, Carabaza AFM. Factors Associated with Stunting Among 0-23 Months-Old Children in Rural Bangladesh. *Journal of Nutrition Education and Behavior* 2017;**49**:S21. doi:10.1016/j.jneb.2017.05.061

15 Hong R, Banta JE, Betancourt JA. Relationship between household wealth inequality and chronic childhood under-nutrition in Bangladesh. *International Journal for Equity in Health* 2006;**5**:1–10. doi:10.1186/1475-9276-5-15

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. *Social Science & Medicine* 2017;**187**:144–54. doi:10.1016/j.socscimed.2017.06.017

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. doi:10.1093/ije/dyt109

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, Arjumand R, *et al.* Maternal and Child Nutrition 2 Evidence-based interventions for improvement of maternaland child nutrition: what can be done and at what cost? *The Lancet* 2013;**382**:452–77. doi:10.1016/S0140-6736(13)60996-4

21 Stewart CP, Iannotti L, Dewey KG, *et al.* Contextualising complementary feeding in a broader framework for stunting prevention. *Matern Child Nutr* 2013;**9**:27–45. doi:10.1111/mcn.12088

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

23 Persson LÅ, 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. doi:10.1001/jama.2012.4061

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 Paediatrica* 2013;**102**:815–23. doi:10.1111/apa.12282

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;:4158–62.

27 Straus MA, Douglas EM. A Short Form of the Revised Conflict Tactics Scales, and Typologies for Severity and Mutuality. *Violence and Victims* 2004;**19**:1–15.

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. doi:10.1136/adc.2008.144444

30 Rahman A, Vahter M, Smith AH, *et al.* Arsenic Exposure During Pregnancy and Size at Birth: A Prospective Cohort Study in Bangladesh. *American Journal of Epidemiology* 2008;**169**:304–12. doi:10.1093/aje/kwn332

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;: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 Paediatrica* 2017;**382**:427–8. doi:10.1111/apa.14009

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. *American Journal of Epidemiology* 2013;**177**:361–70. doi:10.1093/aje/kws281

36 Onis M de, Onyango AW, Borghi E, *et al.* Development of a WHO growth reference for school-aged children and adolescents. *Bulletin of the World Health Organization* 2007;**85**:660–7. doi:10.1590/S0042-96862007000900010

37 Jönsson P, Wohlin C. An Evaluation of K-Nearest Neighbour Imputation Using Likert Data. *Proceedings of th International Symposium on Software Metrics* 2004;:108–18.

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

39 Torsten Hothorn KHAZ. ctree: Conditional Inference Trees. www.cran.r-project.org. 2016;:1–21.https://cran.r-project.org/web/packages/partykit/vignettes/ctree.pdf (accessed 24 Mar2017) page 655.

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. httpa://www.R-project.org/

42 Achim Zeileis THKH. party: A Laboratory for Recursive Partytioning. 2017;:1–18.

43 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. doi:10.1093/ije/dyt109

44 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;:2202–7.

45 Corsi DJ, Mejía-Guevara I, Subramanian SV. Risk factors for chronic undernutrition among children in India: Estimating relative importance, population attributable risk and fractions. *Social Science & Medicine* 2016;**157**:165–85. doi:10.1016/j.socscimed.2015.11.014

46 Kramer MS. The Epidemiology of Adverse Pregnancy Outcomes: An Overview. *J Nutr* 2003;:1592S–1596S.

47 A WHO Collaborative Study of Maternal Anhropomehtry and Pregnancy Outcomes. *International Journal of Gynecology and Obstetrics* 1996;**57**:1–15.

48 Özaltin E. Association of Maternal Stature With Offspring Mortality, Underweight, and Stunting in Low- to Middle-Income Countries. *JAMA* 2010;**303**:1507–21. doi:10.1001/jama.2010.450

49 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. doi:10.1371/journal.pone.0149700

50 Van de Poel E, Hosseinpoor AR, Speybroeck N, *et al.* Socioeconomic inequality in malnutrition in developing countries. *Bulletin of the World Health Organization* 2008;**86**:282–91. doi:10.1590/S0042-96862008000400013

51 Bradley RH, Corwyn RF. Socioeconomic Status and Child Development. *Annual Review of Phsycology* 2002;**53**:371–99.

52 Ramalingaswami V, U J, J R. The Asian enigma. 1996.

53 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. doi:10.1111/j.1740-8709.2008.00161.x

54 Chowdhury AMR, Bhuiya A, Chowdhury ME, *et al.* The Bangladesh paradox: exceptional health achievement despite economic poverty. *The Lancet* 2013;**382**:1734–45. doi:10.1016/S0140-6736(13)62148-0

55 Asadullah MN, Chaudhury N. Girls school enrolement figures are good in Bangladesh but drop out rates are worrying. *Development and Cooperation* 2017;**59**:447–72. doi:10.1086/681929

56 UNESCO. EFA 2015 national review: Bangladesh; 2014. unesdoc.unesco.org. http://unesdoc.unesco.org/images/0023/002305/230507e.pdf (accessed 13 Feb2018).

57 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. doi:10.1093/ije/dyn099

58 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.

59 Arifeen DrPH El S, MPH AC, ScD LR, *et al.* Community-based approaches and partnerships: innovations in health-service delivery in Bangladesh. *The Lancet* 2013;**382**:2012–26. doi:10.1016/S0140-6736(13)62149-2

60 WHO, UNICEF. Bangladesh: WHO and UNICEF estimates of immunization coverage: 2016 revision. 2016;:1–30.

61 Chowdhury AMR, Bhuiya A, Chowdhury ME, *et al.* The Bangladesh paradox: exceptional health achievement despite economic poverty. *The Lancet* 2013;**382**:1734–45. doi:10.1016/S0140-6736(13)62148-0

62 Khan AI, Kabir I, Ekstrom EC, *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.

63 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.

64 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**:S25. doi:10.1186/1471-2458-11-S3-S25

65 Weedon MN, Frayling TM. Reaching new heights: insights into the genetics of human stature. *Trends in Genetics* 2008;**24**:595–603. doi:10.1016/j.tig.2008.09.006

66 Silventoinen K, Sammalisto S, Perola M, *et al.* Heritability of Adult Body Height: A Comparative Study of Twin Cohorts in Eight Countries. *Twin Research* 2003;**6**:399–408. doi:10.1375/twin.6.5.399

67 Simeone P, Alberti S. Epigenetic heredity of human height. *Physiological Reports* 2014;**2**:e12047–7. doi:10.14814/phy2.12047

68 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. *American Journal of Public Health* 2000;**90**:627–30.

69 Imdad A, Bhutta ZA. Effect of balanced protein energy supplementation during pregnancy on birth outcomes. *BMC Public Health* 2011;**11**:S17. doi:10.1186/1471-2458-11-S3-S17

70 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**:S533–46. doi:10.1177/15648265090304S408

71 Ramakrishnan U, Grant FK, Goldenberg T, *et al.* Effect of Multiple Micronutrient Supplementation on Pregnancy and Infant Outcomes: A Systematic Review. *Paediatric and Perinatal Epidemiology* 2012;**26**:153–67. doi:10.1111/j.1365-3016.2012.01276.x

72 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. doi:10.1001/jama.2014.16819

73 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. doi:10.1111/mcn.12367

74 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. doi:10.3945/jn.114.207225

75 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. doi:10.3945/ajcn.114.091546

76 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. *American Journal of Clinical Nutrition* 2009;**90**:1593–600. doi:10.3945/ajcn.2009.28253

77 Gluckman PD, Hanson MA, Cooper C, *et al.* Effect of In Utero and Early-Life Conditions on Adult Health and Disease. *http://dxdoiorg/101056/NEJMra0708473* 2009;**359**:61–73. doi:10.1056/NEJMra0708473

78 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). *American Journal of Clinical Nutrition* 2014;**100**:1257–68. doi:10.3945/ajcn.114.084921

79 The World Bank. World Development Indicators: Reproductive health. 2016;:1–6.

80 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 Human Development* 2010;**86**:535–40. doi:10.1016/j.earlhumdev.2010.07.002

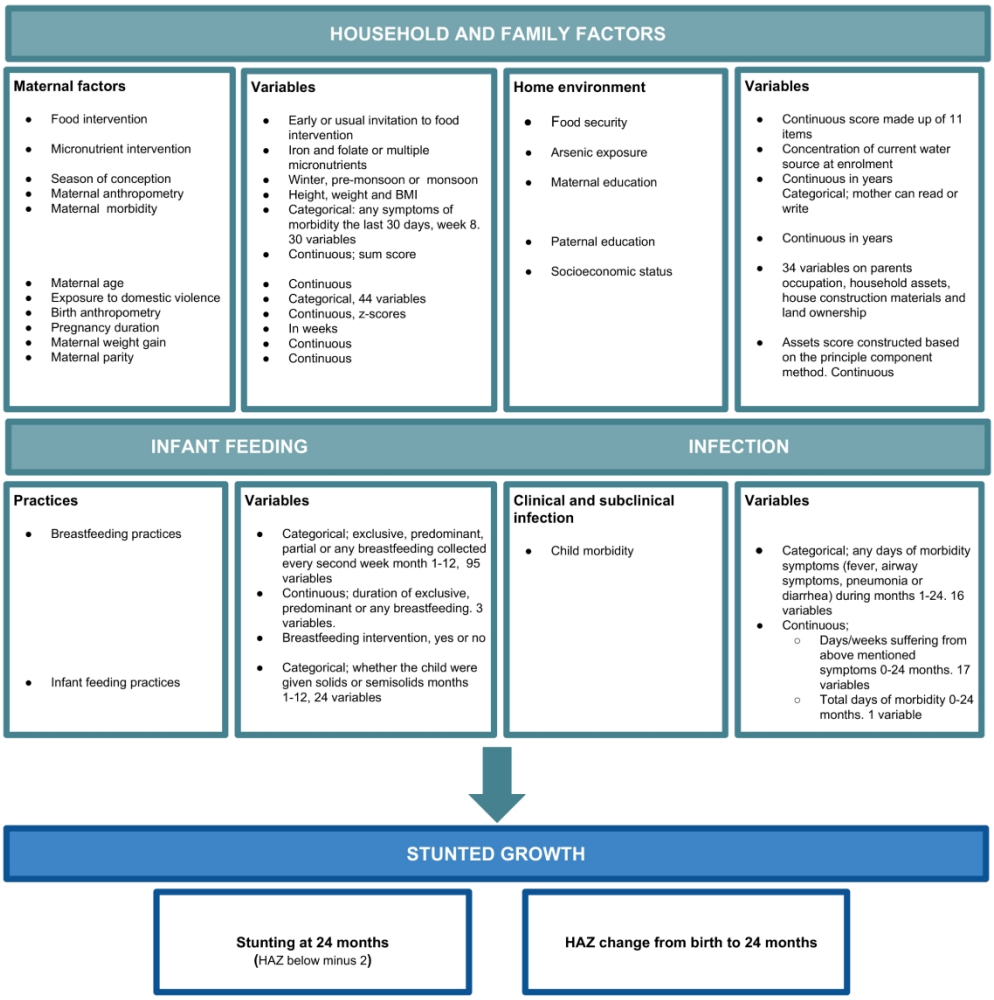

Figure 1. Factors, variables and outcomes included in the analysis of data from the MINIMat cohort, Bangladesh. Grouping according to the WHO conceptual framework on childhood stunting [20]

190x190mm (300 x 300 DPI)

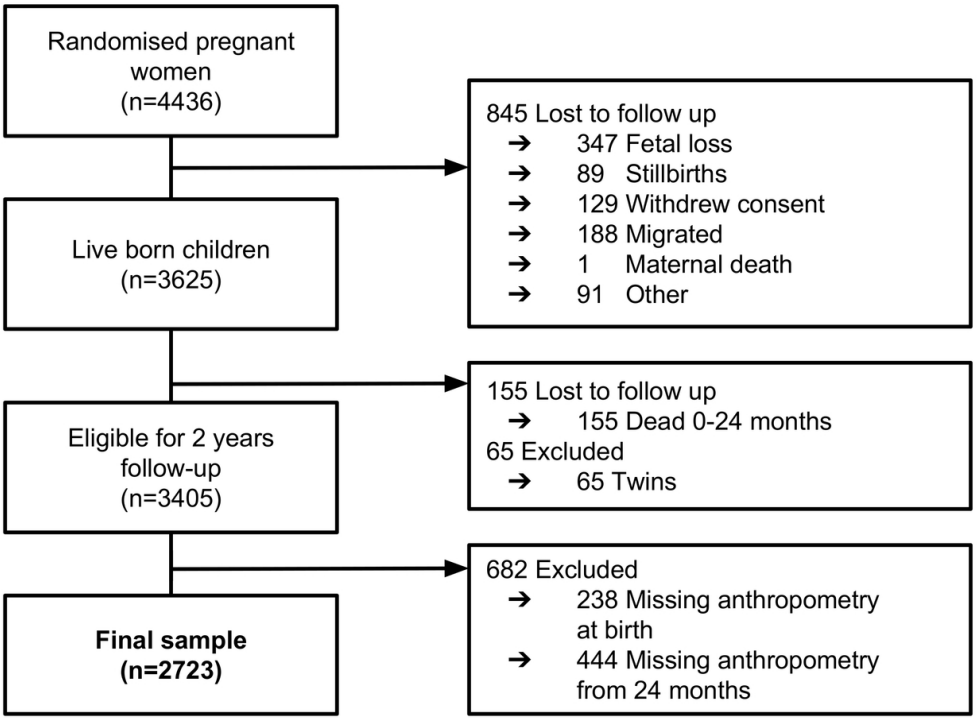

Figure 2. Flow chart of pregnant women and their children included in the data mining analyses of the MINIMat cohort from conception to two years of age.

106x80mm (300 x 300 DPI)

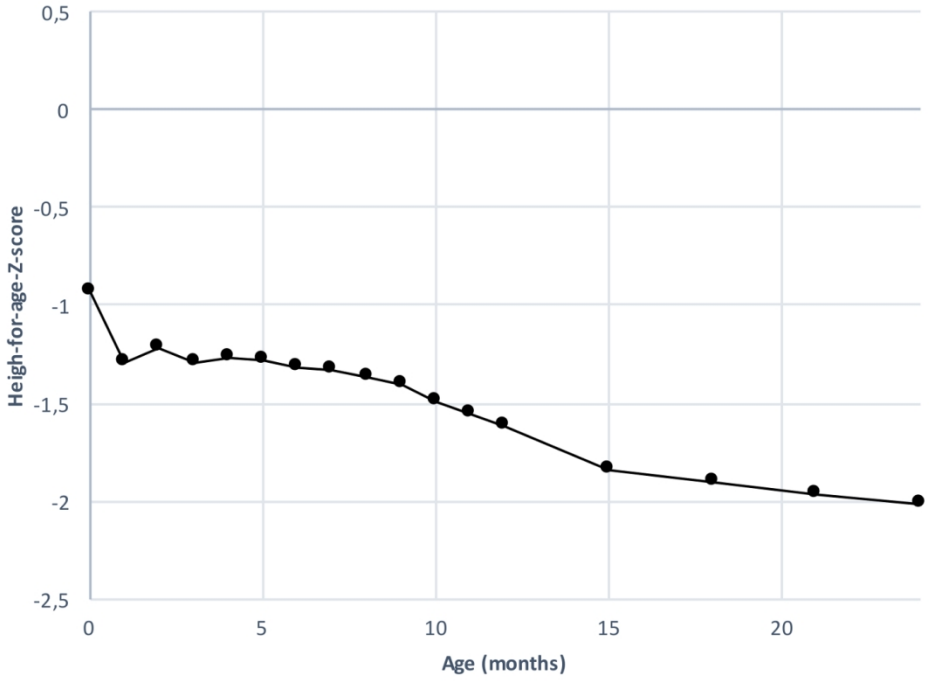

Figure 3. Height-for-age Z-scores from birth to 24 months in the MINIMat cohort in rural Bangladesh.

131x96mm (300 x 300 DPI)

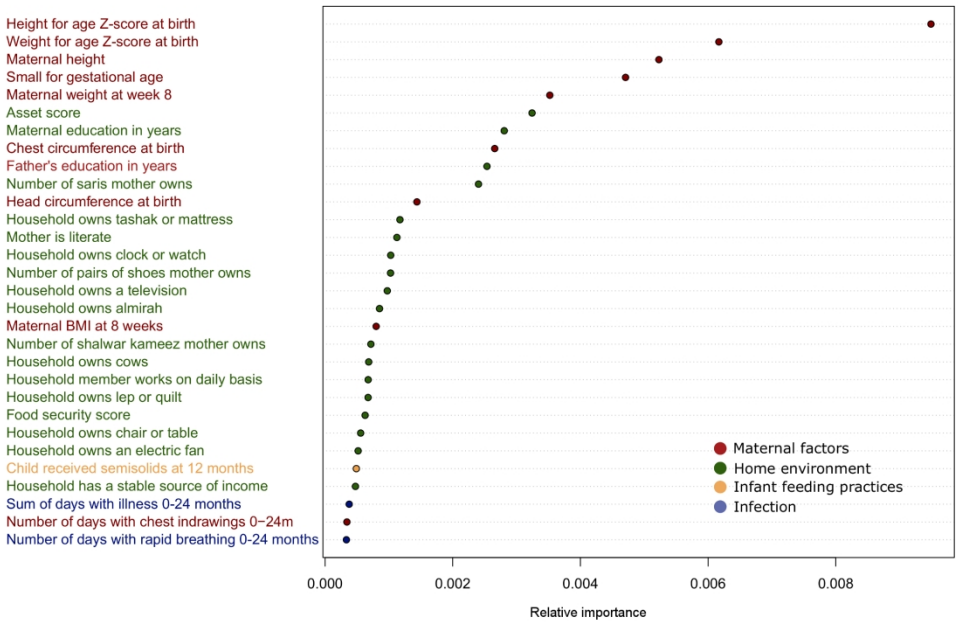

Figure 4. Conditional random forest plot ranking the relative importance of 30 predictors with regard to their ability to explain the presence of stunting at 24 months of age. The MINIMat cohort in rural Bangladesh. Colour coding according to the WHO conceptual framework on causes of stunting.

190x134mm (300 x 300 DPI)

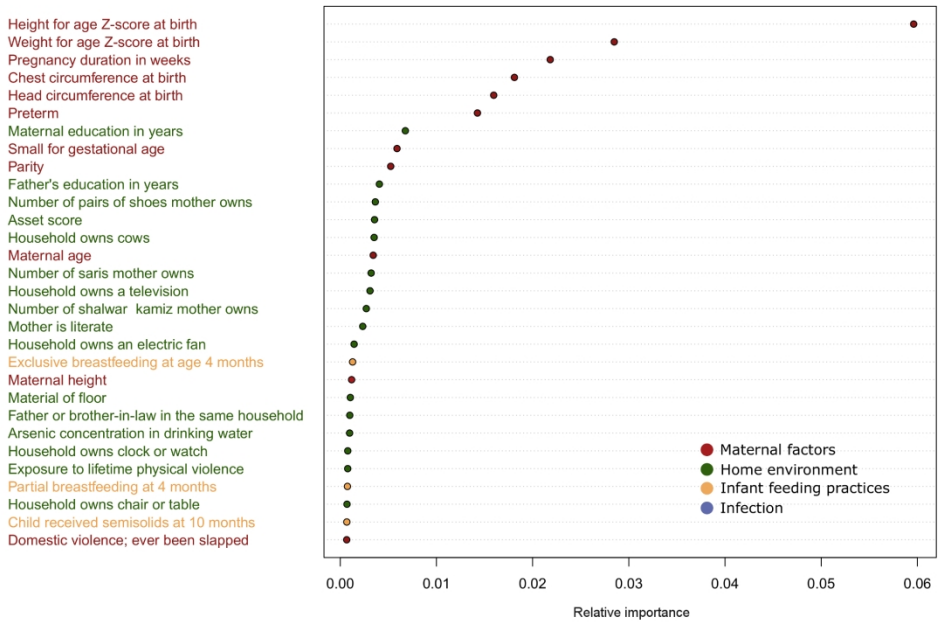

Figure 5. Conditional random forest plot ranking the relative importance of 30 predictors with regard to their ability to explain the variation in change in HAZ (Δ HAZ) from birth to 24 months of age. The MINIMat cohort in rural Bangladesh. Colour coding according to the WHO conceptual framework on causes of stunting.

190x134mm (300 x 300 DPI)

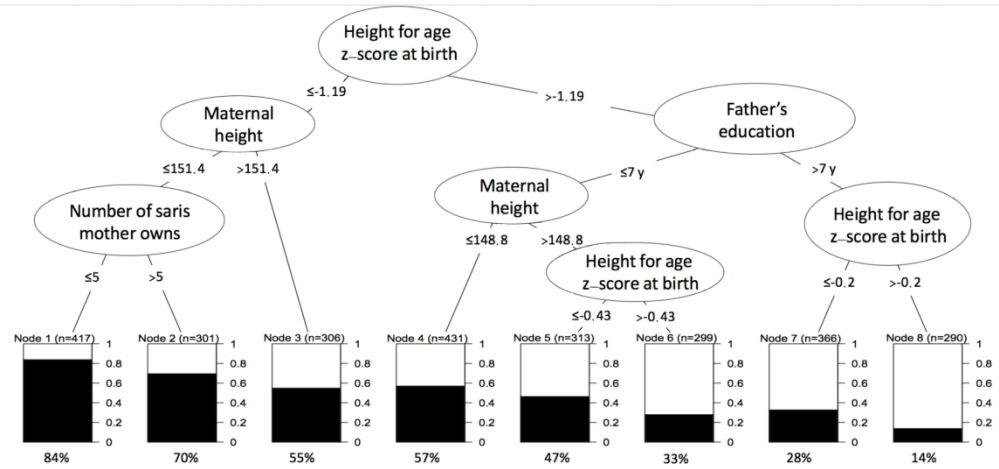

Figure 6. Conditional inference tree identifying sub-groups with different probabilities of stunting at 24 months. The MINIMat cohort in rural Bangladesh.

190x88mm (300 x 300 DPI)

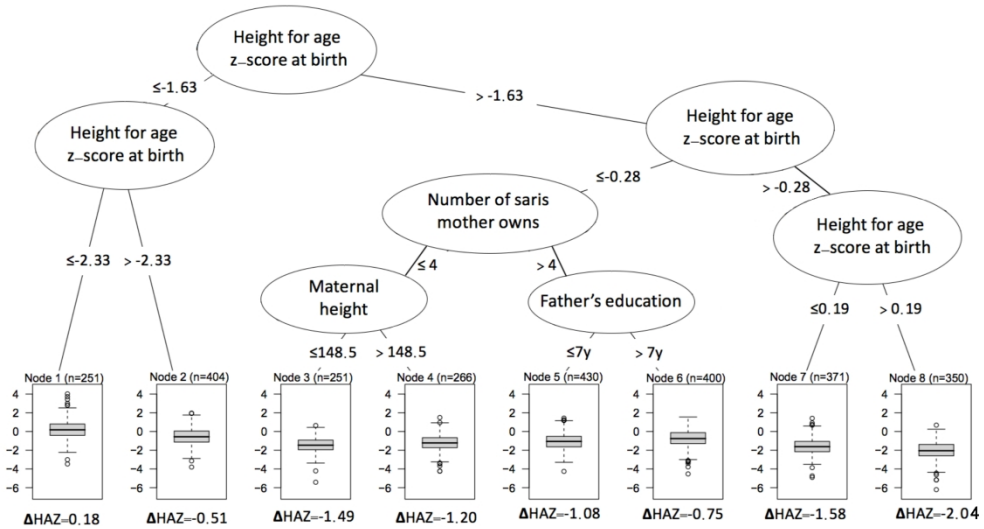

Figure 7. Conditional inference tree identifying sub-groups with different mean change in HAZ (Δ HAZ=HAZ24–HAZ0) 0–24 months within the MINIMat cohort in rural Bangladesh.

193x101mm (300 x 300 DPI)

Supplementation appendix

Simulation study of the predictive performance of three different imputation methods

The following strategy was used to study the imputation accuracy of various methods for the input variables in our analyses. First, we standardized numerical variables in the data and took a sample of the entire data (α) and deleted a proportion (β) of the non-missing values in each variable. Secondly, we employed three different imputation methods to make predictions of the missing values in the data. Lastly, we compared the predictions with the values of the deleted entries, the computed mean-square error (MSE) for the numerical variables, and the percent of the incorrect predictions, misclassification rate (MR), for the categorical ones. The computation of the MSE and MR values was repeated several times for different samples of the original data. The summary results of these computations are presented in Tables 1-4. It can be concluded that random forests[1] provided a statistically significantly better imputation than the variable mean and K-nearest neighbor imputation methods. The design of the study followed a procedure similar to the strategy described in Jonsson et al [2].

Table 1: Means and Standard errors of the MR[2] and the MSE[3] for different imputation methods, computed from m=100 samples, $\alpha = 0.05$, $\beta = 0.05$

	Variable mean	KNN[1]	Random forest
Mean (MR[2])	0.17755631	0.187499573	0.131724506
Standard Error (MR[2])	0.00360524	0.003795385	0.003759032
Mean (MSE[3])	1.01903348	0.901518114	0.541867921
Standard error (MSE[3])	0.01640172	0.016414433	0.015157205

[1] K-nearest neighbour
[2]Missclassification rate
[3]Mean square error
α = proportion of the non-missing values deleted
β = proportion of the original data sampled

Table 2: Means and Standard errors of the MR[2] and the MSE[3] for different imputation methods, computed from m=100 samples, $\alpha = 0.05$, $\beta = 0.15$

	Variable mean	KNN[1]	Random forest
Mean (MR[2])	0.175774830	0.187158897	0.131724506
Standard Error (MR[2])	0.003075253	0.003317242	0.003302446
Mean (MSE[3])	1.00474998	0.922010327	0.556762189
Standard error (MSE[3])	0.01012910	0.009595471	0.008949707

[1] K-nearest neighbour
[2]Missclassification rate
[3]Mean square error
α = proportion of the non-missing values deleted
β = proportion of the original data sampled

Table 3: Means and Standard errors of the MR[2] and the MSE[3] for different imputation methods, computed from m=100 samples, $\alpha = 0.2$, $\beta = 0.05$

	Variable mean	KNN[1]	Random forest
Mean (MR[2])	0.1625007370	0.1608280983	0.094319580
Standard Error (MR[2])	0.0005210379	0.0005181798	0.000367369
Mean (MSE[3])	1.0023969039	0.7975006166	0.450253626
Standard error (MSE[3])	0.0068209597	0.0066997794	0.006069386

[1] K-nearest neighbour
[2]Missclassification rate
[3]Mean square error
α = proportion of the non-missing values deleted
β = proportion of the original data sampled

Table 4: Means and Standard errors of discrete and continuous variables for different imputation methods. Computed from m=100 samples, $\alpha = 0.2$, $\beta = 0.15$

	Variable mean	KNN[1]	Random forest
Mean, discrete	0.1626095174	0.1617267853	0.1017561946
Standard error, Discrete	0.0003670347	0.0003618961	0.0002612874
Mean, continuous	0.9984641615	0.8195273545	0.4593241548
Standard error, continuous	0.0040175223	0.0040319899	0.0034449935

[1] K-nearest neighbour
[2]Missclassification rate
[3]Mean square error
α = proportion of the non-missing values deleted
β = proportion of the original data sampled

References

1. Stekhoven DJ, Buhlmann P. MissForest—non-parametric missing value imputation for mixed-type data. Bioinformatics. 2012;: 112–118.

2. Jönsson P, Wohlin C. An Evaluation of K-Nearest Neighbour Imputation Using Likert Data. Proceedings of th International Symposium on Software Metrics. 2004;: 108–118.

STROBE Statement—checklist of items that should be included in reports of observational studies

	Item No	Recommendation	
Title and abstract	1	(*a*) Indicate the study's design with a commonly used term in the title or the abstract	1
		(*b*) Provide in the abstract an informative and balanced summary of what was done and what was found	2
Introduction			
Background/rationale	2	Explain the scientific background and rationale for the investigation being reported	4-5
Objectives	3	State specific objectives, including any prespecified hypotheses	6
Methods			
Study design	4	Present key elements of study design early in the paper	5
Setting	5	Describe the setting, locations, and relevant dates, including periods of recruitment, exposure, follow-up, and data collection	6
Participants	6	(*a*) *Cohort study*—Give the eligibility criteria, and the sources and methods of selection of participants. Describe methods of follow-up *Case-control study*—Give the eligibility criteria, and the sources and methods of case ascertainment and control selection. Give the rationale for the choice of cases and controls *Cross-sectional study*—Give the eligibility criteria, and the sources and methods of selection of participants	6
		(*b*) *Cohort study*—For matched studies, give matching criteria and number of exposed and unexposed *Case-control study*—For matched studies, give matching criteria and the number of controls per case	
Variables	7	Clearly define all outcomes, exposures, predictors, potential confounders, and effect modifiers. Give diagnostic criteria, if applicable	9 Figure 1
Data sources/ measurement	8*	For each variable of interest, give sources of data and details of methods of assessment (measurement). Describe comparability of assessment methods if there is more than one group	*7-9*
Bias	9	Describe any efforts to address potential sources of bias	7-9
Study size	10	Explain how the study size was arrived at	Not applicable
Quantitative variables	11	Explain how quantitative variables were handled in the analyses. If applicable, describe which groupings were chosen and why	7-9
Statistical methods	12	(*a*) Describe all statistical methods, including those used to control for confounding	10-11
		(*b*) Describe any methods used to examine subgroups and interactions	10-11
		(*c*) Explain how missing data were addressed	10-11
		(*d*) *Cohort study*—If applicable, explain how loss to follow-up was addressed *Case-control study*—If applicable, explain how matching of cases and controls was addressed	10

Cross-sectional study—If applicable, describe analytical methods taking account of sampling strategy	
(e) Describe any sensitivity analyses	10

Continued on next page10

Results			
Participants	13*	(a) Report numbers of individuals at each stage of study—eg numbers potentially eligible, examined for eligibility, confirmed eligible, included in the study, completing follow-up, and analysed	Fig 2 12
		(b) Give reasons for non-participation at each stage Fig 2	Fig 2
		(c) Consider use of a flow diagram Figure 2	Fig 2
Descriptive data	14*	(a) Give characteristics of study participants (eg demographic, clinical, social) and information on exposures and potential confounders	Table 1, 13
		(b) Indicate number of participants with missing data for each variable of interest	13
		(c) *Cohort study*—Summarise follow-up time (eg, average and total amount)	13
Outcome data	15*	*Cohort study*—Report numbers of outcome events or summary measures over time	*13*
		Case-control study—Report numbers in each exposure category, or summary measures of exposure	
		Cross-sectional study—Report numbers of outcome events or summary measures	
Main results	16	(*a*) Give unadjusted estimates and, if applicable, confounder-adjusted estimates and their precision (eg, 95% confidence interval). Make clear which confounders were adjusted for and why they were included	14
		(*b*) Report category boundaries when continuous variables were categorized	14
		(*c*) If relevant, consider translating estimates of relative risk into absolute risk for a meaningful time period	
Other analyses	17	Report other analyses done—eg analyses of subgroups and interactions, and sensitivity analyses	14
Discussion			
Key results	18	Summarise key results with reference to study objectives	15,18
Limitations	19	Discuss limitations of the study, taking into account sources of potential bias or imprecision. Discuss both direction and magnitude of any potential bias	15,16,17
Interpretation	20	Give a cautious overall interpretation of results considering objectives, limitations, multiplicity of analyses, results from similar studies, and other relevant evidence	18-22
Generalisability	21	Discuss the generalisability (external validity) of the study results	17
Other information			
Funding	22	Give the source of funding and the role of the funders for the present study and, if applicable, for the original study on which the present article is based	23

*Give information separately for cases and controls in case-control studies and, if applicable, for exposed and unexposed groups in cohort and cross-sectional studies.

Note: An Explanation and Elaboration article discusses each checklist item and gives methodological background and published examples of transparent reporting. The STROBE checklist is best used in conjunction with this article (freely available on the Web sites of PLoS Medicine at http://www.plosmedicine.org/, Annals of Internal Medicine at http://www.annals.org/, and Epidemiology at http://www.epidem.com/). Information on the STROBE Initiative is available at www.strobe-statement.org.