

BMJ Open is committed to open peer review. As part of this commitment we make the peer review history of every article we publish publicly available.

When an article is published we post the peer reviewers' comments and the authors' responses online. We also post the versions of the paper that were used during peer review. These are the versions that the peer review comments apply to.

The versions of the paper that follow are the versions that were submitted during the peer review process. They are not the versions of record or the final published versions. They should not be cited or distributed as the published version of this manuscript.

BMJ Open is an open access journal and the full, final, typeset and author-corrected version of record of the manuscript is available on our site with no access controls, subscription charges or pay-per-view fees (<u>http://bmjopen.bmj.com</u>).

If you have any questions on BMJ Open's open peer review process please email <u>info.bmjopen@bmj.com</u>

**BMJ** Open

# **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

SCHOLARONE<sup>™</sup> Manuscripts

1

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<sup>1</sup>, Oleg Sysoev<sup>2</sup>, Eva-Charlotte Ekström<sup>1</sup>, Lars-Åke Persson<sup>3</sup>, Shams El Arifeen<sup>4</sup>, Ruchira Naved<sup>4</sup>, Anisur Rahman<sup>4</sup>, Ashraful Islam Khan<sup>4</sup>, Katarina Ekholm Selling<sup>1</sup>

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

BMJ Open: first published as 10.1136/bmjopen-2018-025154 on 5 August 2019. Downloaded from http://bmjopen.bmj.com/ on June 12, 2025 at Agence Bibliographique de l Enseignement Superieur (ABES)

Protected by copyright, including for uses related to text and data mining, Al training, and similar technologies

**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

Page 3 of 44	BMJ Open
1	
2	• Assesses the relative public health importance of pre- and post-natal risk factors.
3	
4 5	• The extensive database with over 300 variables available for the analysis covers a wid
6	
7	range of pre and postnatal household, family, and environmental factors, child
8	
9	characteristics at birth, infant feeding, and morbidity. However, some potential
10	
11	determinants were not present in the database.
12	
13	
14 15	<ul> <li>Includes high-quality longitudinal data with low rates of missing data.</li> </ul>
16	
17	
18	• Employes decision-tree-based methods that permit the inclusion of a high number of
19	
20	predictor variables, variables of different types and automatically discover complex

however, deliver *p*-values or confidence intervals to the results.

BMJ Open: first published as 10.1136/bmjopen-2018-025154 on 5 August 2019. Downloaded from http://bmjopen.bmj.com/ on June 12, 2025 at Agence Bibliographique de l Enseignement Superieur (ABES) . Protected by copyright, including for uses related to text and data mining, Al training, and similar technologies.

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 prenatal recommendations are usually limited to routine micronutrient supplementation for pregnant women [17-19].

### **BMJ** Open

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.

For peer review only - http://bmjopen.bmj.com/site/about/guidelines.xhtml

#### **BMJ** Open

### 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,

#### **BMJ** Open

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  $\Delta$  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

### **BMJ** Open

BMJ Open: first published as 10.1136/bmjopen-2018-025154 on 5 August 2019. Downloaded from http://bmjopen.bmj.com/ on June 12, 2025 at Agence Bibliographique de l Enseignement Superieur (ABES) . Protected by copyright, including for uses related to text and data mining, Al training, and similar technologies.

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  $\Delta$  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

### **BMJ** Open

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.

BMJ Open: first published as 10.1136/bmjopen-2018-025154 on 5 August 2019. Downloaded from http://bmjopen.bmj.com/ on June 12, 2025 at Agence Bibliographique de l Enseignement Superieur (ABES) . Protected by copyright, including for uses related to text and data mining, Al training, and similar technologies.

BMJ Open: first published as 10.1136/bmjopen-2018-025154 on 5 August 2019. Downloaded from http://bmjopen.bmj.com/ on June 12, 2025 at Agence Bibliographique de l Enseignement Superieur (ABES) . Protected by copyright, including for uses related to text and data mining, Al training, and similar technologies.

## 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  $\Delta$  HAZ (change in height-for-age Z-score) 0–24 months in the MINIMat cohort, Bangladesh.

For peer review only - http://bmjopen.bmj.com/site/about/guidelines.xhtml

Characteristics	n/n (%)	Stunted at 24 months <i>n/n</i> (%)	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  $\Delta$  HAZ were HAZ and WAZ at birth, pregnancy duration, head and chest circumference at birth, and maternal education.

BMJ Open: first published as 10.1136/bmjopen-2018-025154 on 5 August 2019. Downloaded from http://bmjopen.bmj.com/ on June 12, 2025 at Agence Bibliographique de l Enseignement Superieur (ABES)

Protected by copyright, including for uses related to text and data mining, Al training, and similar technologies

# 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  $\Delta$  HAZ 0-24 months due to distinctive combinations of levels of certain predictors. The conditional inference trees for stunting and  $\Delta$ 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  $\Delta$  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 ( $\leq$ -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  $\Delta$  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

#### **BMJ** Open

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

#### **BMJ** Open

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

For peer review only - http://bmjopen.bmj.com/site/about/guidelines.xhtml

#### **BMJ** Open

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

BMJ Open: first published as 10.1136/bmjopen-2018-025154 on 5 August 2019. Downloaded from http://bmjopen.bmj.com/ on June 12, 2025 at Agence Bibliographique de l Enseignement Superieur (ABES) . Protected by copyright, including for uses related to text and data mining, Al training, and similar technologies.

#### **BMJ** Open

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 lowmiddle 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

For peer review only - http://bmjopen.bmj.com/site/about/guidelines.xhtml

### **BMJ** Open

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 BMJ Open: first published as 10.1136/bmjopen-2018-025154 on 5 August 2019. Downloaded from http://bmjopen.bmj.com/ on June 12, 2025 at Agence Bibliographique de l Enseignement Superieur (ABES)

Protected by copyright, including for uses related to text and data mining, AI training, and similar technologies

### **BMJ** Open

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.

#### **BMJ** Open

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].

BMJ Open: first published as 10.1136/bmjopen-2018-025154 on 5 August 2019. Downloaded from http://bmjopen.bmj.com/ on June 12, 2025 at Agence Bibliographique de l Enseignement Superieur (ABES) . Protected by copyright, including for uses related to text and data mining, Al training, and similar technologies.

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

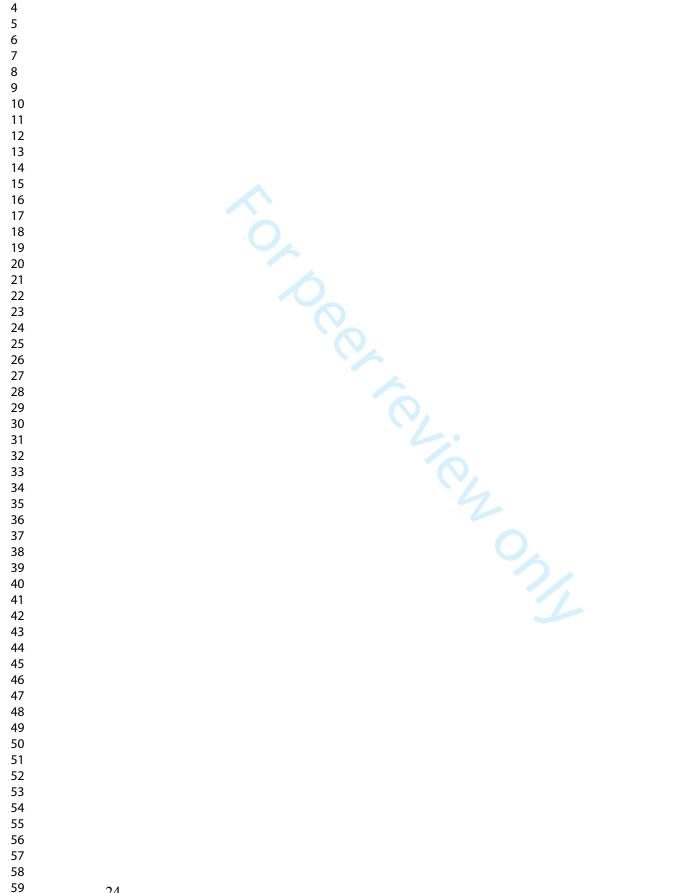
The MINIMat research study was funded by the icddr,b, United Nations Children's Emergency Fund, Swedish International Development Cooperation Agency (Sida), UK Medical Research Council, Swedish Research Council, Department for International Development (DFID), Japan Society for the Promotion of Science, 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 for perteries only

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  $(\Delta \text{HAZ}=\text{HAZ}_{24}-\text{HAZ}_{0})$  0–24 months within the MINIMat cohort in rural Bangladesh.

# References

4	
1	Pradhan M, Sahn DE, Younger SD. Decomposing world health inequality. <i>Journal</i> <i>Health Economics</i> 2003; <b>22</b> :271–93. doi:10.1016/S0167-6296(02)00123-6
2	Victora CG, Adair L, Fall C, <i>et al.</i> Maternal and child undernutrition: consequence for adult health and human capital. <i>Lancet</i> 2008; <b>371</b> :340–57. doi:10.1016/S014 6736(07)61692-4
3	Black RE, Victora CG, Walker SP, <i>et al.</i> Maternal and child undernutrition and overweight in low-income and middle-income countries. <i>The Lancet</i> 2013.
4	General Assembly, UNICEF. Transforming our world: the 2030 Agenda for Sustainable Development. <i>United Nations</i> 2015;:1–35.
5	Victora CG, de Onis M, Hallal PC, <i>et al.</i> Worldwide timing of growth faltering: revisiting implications for interventions. <i>Pediatrics</i> 2010; <b>125</b> :e473–80. doi:10.1542/peds.2009-1519
6	Black RE, Allen LH, Bhutta ZA, <i>et al.</i> Maternal and child undernutrition: global an regional exposures and health consequences. <i>The Lancet</i> 2008; <b>371</b> :243–60. doi:10.1016/S0140-6736(07)61690-0
7	Martorell R, Ramakrishnan U, Schroeder DG, <i>et al.</i> Intrauterine growth retardation, body size, body composition and physical performance in adolescence. <i>Eur J Clin Nutr</i> 1998; <b>52 Suppl 1</b> :S43–52–discussionS52–3.
8	Frongillo EA Jr. <i>Symposium: Causes and Etiology of Stunting. J Nutr</i> 1999; <b>129</b> :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. <i>Trans R Soc Trop Med Hyg</i> 1977; <b>71</b> :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 coho study. <i>PLoS Medicine</i> 2017;:1–21. doi:10.1371/journal.pmed.1002408
11	Espo M, Kulmala T, Maleta K, <i>et al.</i> Determinants of linear growth and predictors of severe stunting during infancy in rural Malawi. <i>Acta Paediatrica</i> 2002; <b>91</b> :136470. doi:10.1111/j.1651-2227.2002.tb02835.x
12	Rahman A, Chowdhury S. Determinants of Chronic Malnutrtion Among Preschoo Children In Bangladesh. <i>J Biosoc Sci</i> 2006; <b>39</b> :161. doi:10.1017/S0021932006001295
	MS MRAM, RN MWMPL, BS AFMC. Factors Associated with Stunting Among 0-23

	BMJ Open
14	Hong R, Banta JE, Betancourt JA. Relationship between household wealth inequality and chronic childhood under-nutrition in Bangladesh. <i>International Journal for Equity in Health</i> 2006; <b>5</b> :1–10. doi:10.1186/1475-9276-5-15
15	Kim R, Mejía-Guevara I, Corsi DJ, <i>et al.</i> Relative importance of 13 correlates of child stunting in South Asia: Insights from nationally representative data from Afghanistan, Bangladesh, India, Nepal, and Pakistan. <i>Social Science &amp; Medicine</i> 2017; <b>187</b> :144–54. doi:10.1016/j.socscimed.2017.06.017
16	Christian P, Lee SE, Donahue Angel M, <i>et al.</i> Risk of childhood undernutrition related to small-for-gestational age and preterm birth in low- and middle-income countries. <i>Int J Epidemiol</i> 2013; <b>42</b> :1340–55. doi:10.1093/ije/dyt109
17	World Health Organization. WHO Guidline: Daily iron and folic acid supplementation in pregnant women. <i>WHO</i> 2012;:1–32.
18	World Health Organization. Essential Nutrition Actions: improving maternal, newborn, infant and young child health and nutrition. <i>WHO</i> 2013;:1–144.
19	Bhutta ZA, Das JK, Arjumand R, <i>et al.</i> Maternal and Child Nutrition 2 Evidence- based interventions for improvement of maternaland child nutrition: what can be done and at what cost? <i>The Lancet</i> 2013; <b>382</b> :452–77. doi:10.1016/S0140- 6736(13)60996-4
20	Stewart CP, Iannotti L, Dewey KG, <i>et al.</i> Contextualising complementary feeding in a broader framework for stunting prevention. <i>Matern Child Nutr</i> 2013; <b>9</b> :27–45. doi:10.1111/mcn.12088
21	Breiman L, Friedman J, Stone CJ, <i>et al. Classification and Regression Trees</i> . Chapman and Hall/CRC 1984.
22	Persson LÅ, Arifeen S, Ekström E-C, <i>et al.</i> Effects of Prenatal Micronutrient and Early Food Supplementation on Maternal Hemoglobin, Birth Weight, and Infant Mortality Among Children in Bangladesh: The MINIMat Randomized Trial. <i>JAMA</i> 2012; <b>307</b> :2050–9. doi:10.1001/jama.2012.4061
23	Khan AI, Hawkesworth S, Ekström E-C, <i>et al.</i> Effects of exclusive breastfeeding intervention on child growth and body composition: the MINIMat trial, Bangladesh. <i>Acta Paediatrica</i> 2013; <b>102</b> :815–23. doi:10.1111/apa.12282
24	Gwatkin DR, Rustein S, Johnson K, <i>et al. Socio-economic differences in health nutrition and population in Bangladesh.</i> Washington, DC: HNP/Poverty Thematic Group of the World Bank 2000.
25	Frongillo EA, Chowdhury N, Ekström E-C, <i>et al.</i> 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. <i>Violence and Victims</i> 2004; <b>19</b> :1–15.

BMJ Open: first published as 10.1136/bmjopen-2018-025154 on 5 August 2019. Downloaded from http://bmjopen.bmj.com/ on June 12, 2025 at Agence Bibliographique de I Enseignement Superieur (ABES) . Protected by copyright, including for uses related to text and data mining, Al training, and similar technologies.

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16 17
1/
18
19
20
21
22
23 24
24 25
26 27
27 28
28 29
29 30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58

59

27	WHO. WHO Multi-country Study on Women's Health and Domestic Violence against Women. <i>WHO</i> 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. <i>Arch Dis Child</i> 2009; <b>94</b> :775–9. doi:10.1136/adc.2008.144444
29	Rahman A, Vahter M, Smith AH, <i>et al.</i> Arsenic Exposure During Pregnancy and Size at Birth: A Prospective Cohort Study in Bangladesh. <i>American Journal of Epidemiology</i> 2008; <b>169</b> :304–12. doi:10.1093/aje/kwn332
30	Arifeen SE, Black RE, Caulfield LE, <i>et al.</i> Infant growth patterns in the slums of Dhaka in relation to birth weight, intrauterine growth retardation, and prematurity. <i>Am J Clin Nutr</i> 2000; <b>72</b> :1010–7.
31	World Health Organization. Indicators for assessing breastfeeding practices. <i>WHO</i> 1991;:1–16.
32	Moore SE, Prentice AM, Coward WA, <i>et al.</i> Use of stable-isotope techniques to validate infant feeding practices reported by Bangladeshi women receiving breastfeeding counseling. <i>Am J Clin Nutr</i> 2007;:1075–82.
33	Kallioinen M, Ekström E-C, Khan AI, <i>et al.</i> Prenatal early food and multiple micronutrient supplementation trial reduced infant mortality in Bangladesh, but did not influence morbidity. <i>Acta Paediatrica</i> 2017; <b>382</b> :427–8. doi:10.1111/apa.14009
34	Arnold BF, Galiani S, Ram PK, <i>et al.</i> Optimal Recall Period for Caregiver-reported Illness in Risk Factor and Intervention Studies: A Multicountry Study. <i>American</i> <i>Journal of Epidemiology</i> 2013; <b>177</b> :361–70. doi:10.1093/aje/kws281
35	Onis M de, Onyango AW, Borghi E, <i>et al.</i> Development of a WHO growth reference for school-aged children and adolescents. <i>Bulletin of the World Health Organization</i> 2007; <b>85</b> :660–7. doi:10.1590/S0042-96862007000900010
36	Jönsson P, Wohlin C. An Evaluation of K-Nearest Neighbour Imputation Using Likert Data. <i>Proceedings of th International Symposium on Software Metrics</i> 2004;:108–18.
37	Stekhoven DJ, Buhlmann P. MissForest—non-parametric missing value imputation for mixed-type data. <i>Bioinformatics</i> 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. <i>The Elements of Statistical Learning</i> . 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.
28	

### **BMJ** Open

Enseignement Superieur (ABES) . Protected by copyright, including for uses related to text and data mining, Al training, and similar technologies.	BMJ Open: first published as 10.1136/bmjopen-2018-025154 on 5 August 2019. Downloaded from http://bmjopen.bmj.com/ on June 12, 2025 at Agence Bibliographiq
---	---

42	Cumming O, Cairncross S. Can water, sanitation and hygiene help eliminate stunting? Current evidence and policy implications. <i>Matern Child Nutr</i> 2016; <b>12</b> :91–105. doi:10.1111/mcn.12258
43	Wills AK, Chinchwadkar MC, Joglekar CV, <i>et al.</i> Maternal and paternal height and BMI and patterns of fetal growth: The Pune Maternal Nutrition Study. <i>Early Human Development</i> 2010; <b>86</b> :535–40. doi:10.1016/j.earlhumdev.2010.07.002
44	Christian P, Lee SE, Donahue Angel M, <i>et al.</i> Risk of childhood undernutrition related to small-for-gestational age and preterm birth in low- and middle-income countries. <i>Int J Epidemiol</i> 2013; <b>42</b> :1340–55. doi:10.1093/ije/dyt109
45	Schmidt MK, Muslimatun S, West CE, <i>et al.</i> Nutritional Status and Linear Growth of Indonesian Infants in West Java Are Determined More by Prenatal Environment than by Postnatal Factors. <i>J Nutr</i> 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. <i>Social Science &amp; Medicine</i> 2016; <b>157</b> :165–85. doi:10.1016/j.socscimed.2015.11.014
47	Kramer MS. The Epidemiology of Adverse Pregnancy Outcomes: An Overview. <i>J</i> <i>Nutr</i> 2003;:1592S–1596S.
48	A WHO Collaborative Study of Maternal Anhropomehtry and Pregnancy Outcomes. <i>International Journal of Gynecology and Obstetrics</i> 1996; <b>57</b> :1–15.
49	Özaltin E. Association of Maternal Stature With Offspring Mortality, Underweight, and Stunting in Low- to Middle-Income Countries. <i>JAMA</i> 2010; <b>303</b> :1507–21. doi:10.1001/jama.2010.450
50	Svefors P, Rahman A, Ekström E-C, <i>et al.</i> Stunted at 10 Years. Linear Growth Trajectories and Stunting from Birth to Pre-Adolescence in a Rural Bangladeshi Cohort. <i>PLoS ONE</i> 2016; <b>11</b> :e0149700–18. doi:10.1371/journal.pone.0149700
51	Van de Poel E, Hosseinpoor AR, Speybroeck N, <i>et al.</i> Socioeconomic inequality in malnutrition in developing countries. <i>Bulletin of the World Health Organization</i> 2008; <b>86</b> :282–91. doi:10.1590/S0042-96862008000400013
52	Bradley RH, Corwyn RF. Socioeconomic Status and Child Development. <i>Annual Review of Phsycology</i> 2002; <b>53</b> :371–99.
53	Ramalingaswami V, U J, J R. The Asian enigma. 1996.
54	Shroff M, Griffiths P, Adair L, <i>et al.</i> Maternal autonomy is inversely related to child stunting in Andhra Pradesh, India. <i>Matern Child Nutr</i> 2009; <b>5</b> :64–74. doi:10.1111/j.1740-8709.2008.00161.x
55	Chowdhury AMR, Bhuiya A, Chowdhury ME, <i>et al.</i> The Bangladesh paradox: exceptional health achievement despite economic poverty. <i>The Lancet</i> 2013; <b>382</b> :1734–45. doi:10.1016/S0140-6736(13)62148-0

BMJ Open: first published as 10.1136/bmjopen-2018-025154 on 5 August 2019. Downloaded from http://bmjopen.bmj.com/ on June 12, 2025 at Agence Bibliographique de I Enseignement Superieur (ABES) . Protected by copyright, including for uses related to text and data mining, Al training, and similar technologies.

2	
3	
4 5	
2	
6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 23 24 25 26	
7	
8	
9	
10	
11	
11	
12	
13	
14	
15	
16	
17	
17	
18	
19	
20	
21	
22	
22	
∠⊃ >4	
24	
25	
26	
26 27 28 29 30	
28	
20	
29	
30	
31 32 33 34 35	
32	
33	
34	
25	
22	
36	
36 37 38	
38	
39	
40	
41	
41	
43	
44	
45	
46	
47	
48	
49	
50	
51	
52	
53	
55 54	
55	
56	
57	
58	
59	
22	

60

56	Asadullah MN, Chaudhury N. Girls school enrolement figures are good in Bangladesh but drop out rates are worrying. <i>Development and Cooperation</i> 2017; <b>59</b> :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, <i>et al.</i> Multi-country analysis of the effects of diarrhoea on childhood stunting. <i>Int J Epidemiol</i> 2008; <b>37</b> :816–30. doi:10.1093/ije/dyn099
59	Koster FT, Curlin GC, Aziz KMA, <i>et al.</i> Synergistic impact of measels and diarrhoea on nutrtion and mortality in Bangladesh. <i>Bulletin of the World Health Organization</i> 1981;:901–8.
60	Arifeen DrPH El S, MPH AC, ScD LR, <i>et al.</i> Community-based approaches and partnerships: innovations in health-service delivery in Bangladesh. <i>The Lancet</i> 2013; <b>382</b> :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, <i>et al.</i> The Bangladesh paradox: exceptional health achievement despite economic poverty. <i>The Lancet</i> 2013; <b>382</b> :1734–45. doi:10.1016/S0140-6736(13)62148-0
63	Khan AI, Kabir I, Ekstrom EC, <i>et al.</i> Effects of prenatal food and micronutrient supplementation on child growth from birth to 54 months of age: a randomized trial in Bangladesh. <i>Nutr J</i> 2011.
64	Martorell R. Physical growth and development of the malnourished child: Contributions from 50 years of research at INCAP. <i>Food Nutr Bull</i> 2010; <b>31</b> :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. <i>BMC Public Health</i> 2011; <b>11</b> :S25. doi:10.1186/1471-2458- 11-S3-S25
66	Weedon MN, Frayling TM. Reaching new heights: insights into the genetics of human stature. <i>Trends in Genetics</i> 2008; <b>24</b> :595–603. doi:10.1016/j.tig.2008.09.006
67	Silventoinen K, Sammalisto S, Perola M, <i>et al.</i> Heritability of Adult Body Height: A Comparative Study of Twin Cohorts in Eight Countries. <i>Twin Research</i> 2003; <b>6</b> :399–408. doi:10.1375/twin.6.5.399
68	Simeone P, Alberti S. Epigenetic heredity of human height. <i>Physiological Reports</i> 2014; <b>2</b> :e12047–7. doi:10.14814/phy2.12047
69	Silventoinen K, Kaprio J, Lahelma E, <i>et al.</i> Relative Effect of Genetic and Environmental Factors on Body Height: Differences Across Birth Cohorts Among Finnish Men and Women. <i>American Journal of Public Health</i> 2000; <b>90</b> :627–30.
30	For peer review only - http://bmjopen.bmj.com/site/about/guidelines.xhtml

### BMJ Open

70	Imdad A, Bhutta ZA. Effect of balanced protein energy supplementation during pregnancy on birth outcomes. <i>BMC Public Health</i> 2011; <b>11</b> :S17. doi:10.1186/1471-2458-11-S3-S17
71	Fall CHD, Fisher DJ, Osmond C, <i>et al.</i> Multiple Micronutrient Supplementation during Pregnancy in Low-Income Countries: A Meta-Analysis of Effects on Birth Size and Length of Gestation. <i>Food Nutr Bull</i> 2009; <b>30</b> :S533–46. doi:10.1177/15648265090304S408
72	Ramakrishnan U, Grant FK, Goldenberg T, <i>et al.</i> Effect of Multiple Micronutrient Supplementation on Pregnancy and Infant Outcomes: A Systematic Review. <i>Paediatric and Perinatal Epidemiology</i> 2012; <b>26</b> :153–67. doi:10.1111/j.1365- 3016.2012.01276.x
73	West KP, Shamim AA, Mehra S, <i>et al.</i> 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. <i>JAMA</i> 2014; <b>312</b> :2649–58. doi:10.1001/jama.2014.16819
74	Johnson W, Darboe MK, Sosseh F, <i>et al.</i> Association of prenatal lipid-based nutritional supplementation with fetal growth in rural Gambia. <i>Matern Child Nutr</i> 2017; <b>13</b> :e12367. doi:10.1111/mcn.12367
75	Ashorn P, Alho L, Ashorn U, <i>et al.</i> 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. <i>J Nutr</i> 2015; <b>145</b> :1345–53. doi:10.3945/jn.114.207225
76	Adu-Afarwuah S, Lartey A, Okronipa H, <i>et al.</i> Lipid-based nutrient supplement increases the birth size of infants of primiparous women in Ghana. <i>Am J Clin Nutr</i> 2015; <b>101</b> :ajcn.091546–846. doi:10.3945/ajcn.114.091546
77	Huybregts L, Roberfroid D, Lanou H, <i>et al.</i> Prenatal food supplementation fortified with multiple micronutrients increases birth length: a randomized controlled trial in rural Burkina Faso. <i>American Journal of Clinical Nutrition</i> 2009; <b>90</b> :1593–600. doi:10.3945/ajcn.2009.28253
78	Gluckman PD, Hanson MA, Cooper C, <i>et al.</i> Effect of In Utero and Early-Life Conditions on Adult Health and Disease. <i>http://dxdoiorg/101056/NEJMra0708473</i> 2009; <b>359</b> :61–73. doi:10.1056/NEJMra0708473
79	Potdar RD, Sahariah SA, Gandhi M, <i>et al.</i> Improving women's diet quality preconceptionally and during gestation: effects on birth weight and prevalence of low birth weighta randomized controlled efficacy trial in India (Mumbai Maternal Nutrition Project). <i>American Journal of Clinical Nutrition</i> 2014; <b>100</b> :1257–68. doi:10.3945/ajcn.114.084921
80	The World Bank. World Development Indicators: Reproductive health. 2016;:1–6.
	For peer review only - http://bmjopen.bmj.com/site/about/guidelines.xhtml

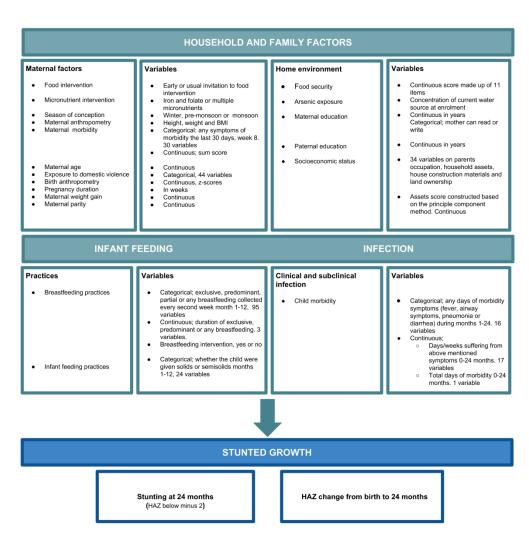


Figure 1. Factors, variables and outcomes included in the analysis of data from the MINIMat cohort, Bangladesh. Grouping according to the WHO conceptual framework on childhood stunting [20]

190x190mm (300 x 300 DPI)

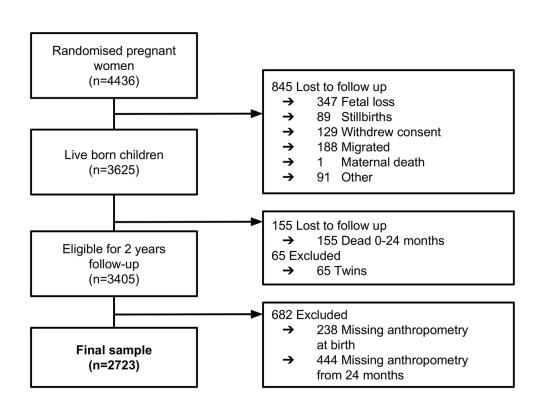


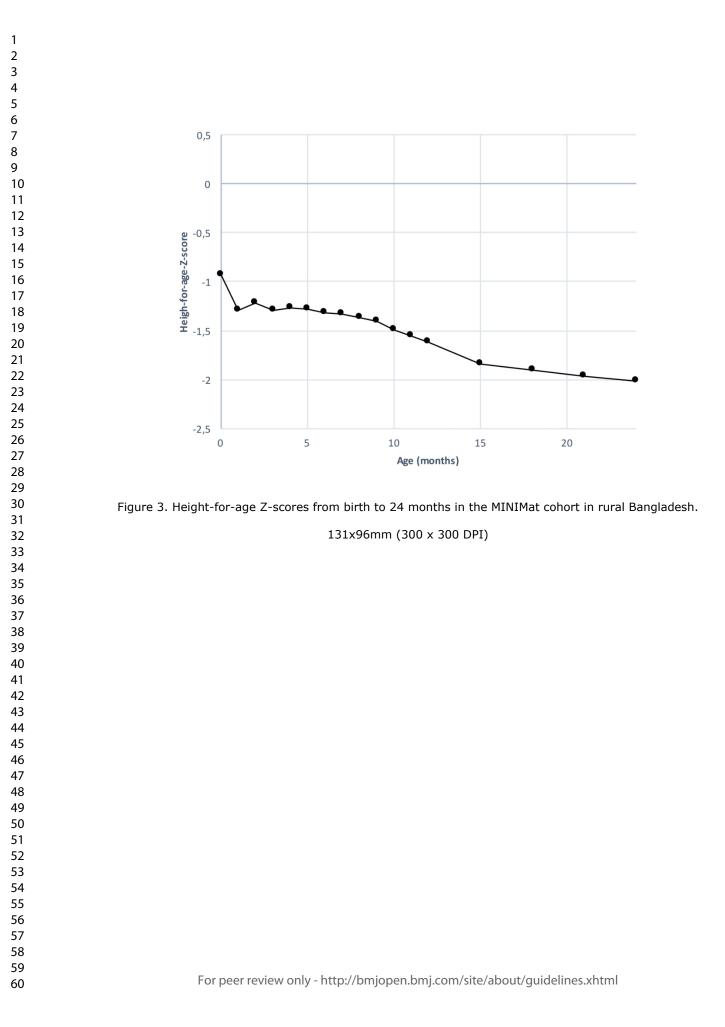
Figure 2. Flow chart of pregnant women and their children included in the data mining analyses of the MINIMat cohort from conception to two years of age.

BMJ Open: first published as 10.1136/bmjopen-2018-025154 on 5 August 2019. Downloaded from http://bmjopen.bmj.com/ on June 12, 2025 at Agence Bibliographique de l Enseignement Superieur (ABES) . Protected by copyright, including for uses related to text and data mining, Al training, and similar technologies.

106x80mm (300 x 300 DPI)

For peer review only - http://bmjopen.bmj.com/site/about/guidelines.xhtml

**BMJ** Open



1			
1 2			
3			
4 5			
6			
7			
8			
9	Height for age Z-score at birth		•
10	Weight for age Z-score at birth Maternal height	· · · · · · · · · · · · · · · · · · ·	
10	Small for gestational age Maternal weight at week 8		
12	Asset score Maternal education in years	•	
13	Chest circumference at birth Father's education in years	•	
14	Number of saris mother owns Head circumference at birth	•	
15	Household owns tashak or mattress Mother is literate	•	
16	Household owns clock or watch		
17	Number of pairs of shoes mother owns Household owns a television	•	
18	Household owns almirah Maternal BMI at 8 weeks	•	
19	Number of shalwar kameez mother owns Household owns cows		
20	Household member works on daily basis Household owns lep or quilt		
21	Food security score Household owns chair or table	•	
22	Household owns an electric fan Child received semisolids at 12 months		<ul> <li>Maternal factors</li> <li>Home environment</li> </ul>
23	Household has a stable source of income Sum of days with illness 0-24 months	•	Infant feeding practices
24	Number of days with chest indrawings 0–24m Number of days with rapid breathing 0-24 months	•	Inection
25 26		Ч <u> </u>	
20	0	000 0.002 0.004 0.00	6 0.008
28		Relative importance	
29			
30 Fig	gure 4. Conditional random forest plo		
31	ability to explain the presence of stur Colour coding according to	the WHO conceptual framework of	
32			
33		190x134mm (300 x 300 DPI)	
34			
35			
36			
37 38			
39			
40			
41			
42			
43			
43 44			
44 45 46			
44 45 46 47			
44 45 46 47 48			
44 45 46 47 48 49			
44 45 46 47 48 49 50			
44 45 46 47 48 49 50 51			
44 45 46 47 48 49 50 51 52			
44 45 46 47 48 49 50 51 52 53			
44 45 46 47 48 49 50 51 52 53 54			
44 45 46 47 48 49 50 51 52 53 54 55			
44 45 46 47 48 49 50 51 52 53 54			
44 45 46 47 48 49 50 51 52 53 54 55 56 57 58			
44 45 46 47 48 49 50 51 52 53 54 55 56 57 58 59			
44 45 46 47 48 49 50 51 52 53 54 55 56 57 58	For peer review only -	http://bmjopen.bmj.com/site/about	/guidelines.xhtml

Page 36 of 44

BMJ Open: first published as 10.1136/bmjopen-2018-025154 on 5 August 2019. Downloaded from http://bmjopen.bmj.com/ on June 12, 2025 at Agence Bibliographique de I Enseignement Superieur (ABES) . Protected by copyright, including for uses related to text and data mining, Al training, and similar technologies.

1						
2						
3						
4						
5						
6 7						
8						
9 Height for age Z-scor Weight for age Z-scor						• • • •
10 Pregnancy duration in Chest circumference	n weeks	•	•			
11 Head circumference a Preterm		•				
12 Maternal education in Small for gestational		•				
13 Parity 14 Father's education in Number of parity	years					
15 Asset score						
16 Household owns cow Maternal age	•••••					
17 Number of saris moth Household owns a tel Number of shalwar k	levision					
18 Mother is literate	•••••					
19 Housenoid owns an e Exclusive breastfeedi 20 Matemai height						
Material of floor	aw in the same household					
22 Arsenic concentration Household owns cloc				Materna	factors	
23 Exposure to lifetime p Partial breastfeeding	at 4 months			Home er Infant fe	nvironment eding practices	
24 Household owns chai Child received semisc	olids at 10 months			<ul> <li>Infectior</li> </ul>	1	
25 Domestic violence; ev 26		1 1	1			
27	0.00	0.01 0.0	02 0.03 Relative importance	0.04	0.05	0.06
28				-		
29 Figure 5, Conditional rand	lom forest plot rankin	a the relative	a importance	of 30 pr	dictors wi	th regard to the
2930Figure 5. Conditional rand ability to explain the v	ariation in change in I	HAZ (Δ HAZ)	from birth to	24 mon	ths of age	. The MINIMat
29Figure 5. Conditional rand30ability to explain the v31cohort in rural Bangla	dom forest plot rankin ariation in change in f adesh. Colour coding a	$HAZ (\Delta HAZ)$	from birth to	24 mon	ths of age	. The MINIMat
2930Figure 5. Conditional rand ability to explain the v	ariation in change in I	HAZ (Δ HAZ)	from birth to	24 mon	ths of age	. The MINIMat
29Figure 5. Conditional rand ability to explain the v30ability to explain the v31cohort in rural Bangla323334	ariation in change in H adesh. Colour coding a	$HAZ (\Delta HAZ)$	from birth to the WHO con	24 mon	ths of age	. The MINIMat
29Figure 5. Conditional rand ability to explain the v30ability to explain the v31cohort in rural Bangla32333435	ariation in change in H adesh. Colour coding a	HAZ (Δ HAZ) according to t stunting.	from birth to the WHO con	24 mon	ths of age	. The MINIMat
29Figure 5. Conditional rand ability to explain the v cohort in rural Bangla31cohort in rural Bangla323334353636	ariation in change in H adesh. Colour coding a	HAZ (Δ HAZ) according to t stunting.	from birth to the WHO con	24 mon	ths of age	. The MINIMat
29Figure 5. Conditional rand ability to explain the v cohort in rural Bangla31cohort in rural Bangla323334353637	ariation in change in H adesh. Colour coding a	HAZ (Δ HAZ) according to t stunting.	from birth to the WHO con	24 mon	ths of age	. The MINIMat
29Figure 5. Conditional rand ability to explain the v cohort in rural Bangla31cohort in rural Bangla32333435363738	ariation in change in H adesh. Colour coding a	HAZ (Δ HAZ) according to t stunting.	from birth to the WHO con	24 mon	ths of age	. The MINIMat
29Figure 5. Conditional rand ability to explain the v cohort in rural Bangla31cohort in rural Bangla323334353637383940	ariation in change in H adesh. Colour coding a	HAZ (Δ HAZ) according to t stunting.	from birth to the WHO con	24 mon	ths of age	. The MINIMat
29Figure 5. Conditional random30ability to explain the v cohort in rural Bangla3233333435363738394041	ariation in change in H adesh. Colour coding a	HAZ (Δ HAZ) according to t stunting.	from birth to the WHO con	24 mon	ths of age	. The MINIMat
29Figure 5. Conditional random30ability to explain the v cohort in rural Bangla323333343536373839404142	ariation in change in H adesh. Colour coding a	HAZ (Δ HAZ) according to t stunting.	from birth to the WHO con	24 mon	ths of age	. The MINIMat
29Figure 5. Conditional random30ability to explain the v cohort in rural Bangla3233333435363738394041424343	ariation in change in H adesh. Colour coding a	HAZ (Δ HAZ) according to t stunting.	from birth to the WHO con	24 mon	ths of age	. The MINIMat
29Figure 5. Conditional random30ability to explain the v cohort in rural Bangla323333343536373839404142	ariation in change in H adesh. Colour coding a	HAZ (Δ HAZ) according to t stunting.	from birth to the WHO con	24 mon	ths of age	. The MINIMat
29       Figure 5. Conditional random         30       ability to explain the v         31       cohort in rural Bangla         32       33         34       35         36       37         38       39         40       41         42       43         43       44         45       46	ariation in change in H adesh. Colour coding a	HAZ (Δ HAZ) according to t stunting.	from birth to the WHO con	24 mon	ths of age	. The MINIMat
29       Figure 5. Conditional random         30       ability to explain the vicohort in rural Bangla         32       cohort in rural Bangla         33       34         35       36         37       38         39       40         41       42         43       44         45       46         47       47	ariation in change in H adesh. Colour coding a	HAZ (Δ HAZ) according to t stunting.	from birth to the WHO con	24 mon	ths of age	. The MINIMat
29       Figure 5. Conditional random         30       ability to explain the v         31       cohort in rural Bangla         32       33         34       35         36       37         38       39         40       41         42       43         44       45         46       47         48       48	ariation in change in H adesh. Colour coding a	HAZ (Δ HAZ) according to t stunting.	from birth to the WHO con	24 mon	ths of age	. The MINIMat
29       Figure 5. Conditional random         30       ability to explain the v         31       cohort in rural Bangla         32       33         34       35         36       37         38       39         40       41         42       43         43       44         45       46         47       48         49       49	ariation in change in H adesh. Colour coding a	HAZ (Δ HAZ) according to t stunting.	from birth to the WHO con	24 mon	ths of age	. The MINIMat
29       Figure 5. Conditional random         30       ability to explain the volume         31       cohort in rural Bangla         32       33         34       35         36       37         38       39         40       41         42       43         43       44         45       46         47       48         49       50	ariation in change in H adesh. Colour coding a	HAZ (Δ HAZ) according to t stunting.	from birth to the WHO con	24 mon	ths of age	. The MINIMat
29       Figure 5. Conditional random         30       ability to explain the vicohort in rural Bangla         32       33         34       35         36       37         38       39         40       41         42       43         44       45         46       47         48       49         50       51         52       52	ariation in change in H adesh. Colour coding a	HAZ (Δ HAZ) according to t stunting.	from birth to the WHO con	24 mon	ths of age	. The MINIMat
29       Figure 5. Conditional random         30       ability to explain the vicohort in rural Bangla         32       33         34       35         36       37         38       39         40       41         42       43         44       45         46       47         48       49         50       51         52       53	ariation in change in H adesh. Colour coding a	HAZ (Δ HAZ) according to t stunting.	from birth to the WHO con	24 mon	ths of age	. The MINIMat
29       Figure 5. Conditional random         30       ability to explain the vicohort in rural Bangla         32       33         34       35         36       37         38       39         40       41         42       43         44       45         46       47         48       49         50       51         51       52         53       54	ariation in change in H adesh. Colour coding a	HAZ (Δ HAZ) according to t stunting.	from birth to the WHO con	24 mon	ths of age	. The MINIMat
29       Figure 5. Conditional random         30       ability to explain the vicohort in rural Bangla         32       33         34       35         36       37         38       39         40       41         42       43         44       45         46       47         48       49         50       51         51       52         53       54         55       54	ariation in change in H adesh. Colour coding a	HAZ (Δ HAZ) according to t stunting.	from birth to the WHO con	24 mon	ths of age	. The MINIMat
29       Figure 5. Conditional random         30       ability to explain the vicohort in rural Bangla         32       33         34       35         36       37         38       39         40       41         42       43         44       45         46       47         48       49         50       51         51       52         53       54	ariation in change in H adesh. Colour coding a	HAZ (Δ HAZ) according to t stunting.	from birth to the WHO con	24 mon	ths of age	. The MINIMat
29       Figure 5. Conditional random         30       ability to explain the vicohort in rural Bangla         32       33         34       35         36       37         38       39         40       41         42       43         44       45         46       47         48       49         50       51         51       52         53       54         55       56         57       58	ariation in change in H adesh. Colour coding a	HAZ (Δ HAZ) according to t stunting.	from birth to the WHO con	24 mon	ths of age	. The MINIMat
29       Figure 5. Conditional random         30       ability to explain the volume         31       cohort in rural Bangla         32       33         34       35         36       37         38       39         40       41         42       43         43       44         45       46         47       48         49       50         51       52         53       54         55       56         57       58         59	rariation in change in H adesh. Colour coding a 190x13	HAZ (Δ HAZ) according to t stunting. 4mm (300 x	from birth to the WHO con 300 DPI)	o 24 mon ceptual f	ths of age ramework	. The MINIMat on causes of
29       Figure 5. Conditional random         30       ability to explain the vicohort in rural Bangla         32       33         34       35         36       37         38       39         40       41         42       43         44       45         46       47         48       49         50       51         51       52         53       54         55       56         57       58         59       50	ariation in change in H adesh. Colour coding a	HAZ (Δ HAZ) according to t stunting. 4mm (300 x	from birth to the WHO con 300 DPI)	o 24 mon ceptual f	ths of age ramework	. The MINIMat on causes of

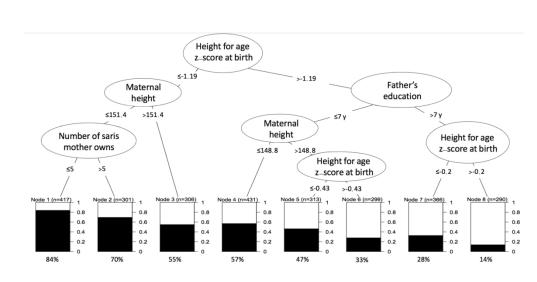


Figure 6. Conditional inference tree identifying sub-groups with different probabilities of stunting at 24 months. The MINIMat cohort in rural Bangladesh.

190x88mm (300 x 300 DPI)

BMJ Open: first published as 10.1136/bmjopen-2018-025154 on 5 August 2019. Downloaded from http://bmjopen.bmj.com/ on June 12, 2025 at Agence Bibliographique de l Enseignement Superieur (ABES) . Protected by copyright, including for uses related to text and data mining, Al training, and similar technologies.



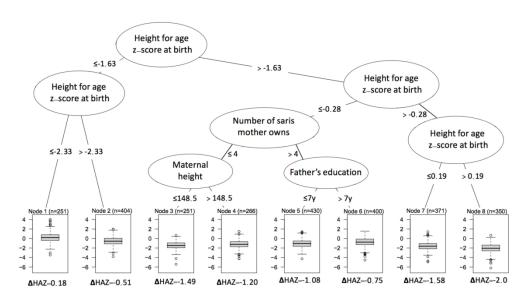


Figure 7. Conditional inference tree identifying sub-groups with different mean change in HAZ ( $\Delta$  HAZ=HAZ24-HAZ0) 0-24 months within the MINIMat cohort in rural Bangladesh.

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 ( $\alpha$ ) and deleted a proportion ( $\beta$ ) 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].

Enseignement Superieur (ABES) . Protected by copyright, including for uses related to text and data mining, Al training, and similar technologies.

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

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

<sup>1</sup> K-nearest neighbour <sup>2</sup>Missclassification rate

<sup>3</sup>Mean square error

Wear square error

 $\alpha$  = proportion of the non-missing values deleted

 $\beta$  = proportion of the original data sampled

Enseignement Superieur (ABES) . Protected by copyright, including for uses related to text and data mining, Al training, and similar technologies.

2	
3	
4	
5	
6	
7 8	
8	
9	
1	0
1	1
1	2
	3
1	4
1	5
	6
1	7
	, 8
1	9
ว	0
	1
2	2
	3
2	4
	5
2	6
2	7
2	8
2	9
	0
3	1
3	
3	3
3	
3	5
2	6
	7
3	8
	9
4	0
4	1
4	2
4	2
	4
4	5
	6
4	7
	8
	9
5	0
-	-
5	
5	2
	2
5	3
5	4
F	5
5	6
5	7
	8
5	9

60

1

<b>Table 2:</b> Means and Standard errors of the MR <sup>2</sup> and the MSE <sup>3</sup> for different
imputation methods, computed from m=100 samples, $lpha=0.05, eta=0.15$

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

<sup>1</sup> K-nearest neighbour

<sup>2</sup>Missclassification rate

<sup>3</sup>Mean square error

 $\alpha$  = proportion of the non-missing values deleted

 $\beta$  = proportion of the original data sampled

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

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

<sup>1</sup> K-nearest neighbour

<sup>2</sup>Missclassification rate

<sup>3</sup>Mean square error

- $\alpha$  = proportion of the non-missing values deleted
- $\beta$  = 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 <sup>1</sup>	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

<sup>1</sup> K-nearest neighbour

<sup>2</sup>Missclassification rate

<sup>3</sup>Mean square error

 $\alpha$  = proportion of the non-missing values deleted

 $\beta$  = 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.
- 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.

BMJ Open: first published as 10.1136/bmjopen-2018-025154 on 5 August 2019. Downloaded from http://bmjopen.bmj.com/ on June 12, 2025 at Agence Bibliographique de l Enseignement Superieur (ABES) . Protected by copyright, including for uses related to text and data mining, Al training, and similar technologies.

2	
2	
3	
4	
5	
6	
0	
7	
8	
9	
10	
11	
12	
13	
14	
15	
16	
17	
18	
19	
20	
21	
22	
23	
24	
25	
26	
27	
28	
29	
30	
31	
33	
34	
35	
36	
37	
38	
39	
40	
41	
42	
43	
44	
45	
46	
47	
48	
49	
50	
51	
52	
53	
54	
55	
56	
57	
58	
59	
60	

1

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

	Item No	Recommendation	
Title and abstract	1	( <i>a</i> ) Indicate the study's design with a commonly used term in the	1
		title or the abstract	
		(b) Provide in the abstract an informative and balanced summary	2
		of what was done and what was found	
Introduction			
Background/rationale	2	Explain the scientific background and rationale for the	4-5
		investigation being reported	
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	6
c		periods of recruitment, exposure, follow-up, and data collection	
Participants	6	(a) Cohort study—Give the eligibility criteria, and the sources	6
•		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	
		(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	9 Figure
		confounders, and effect modifiers. Give diagnostic criteria, if	1
		applicable	
Data sources/ measurement	8*	For each variable of interest, give sources of data and details of	7-9
		methods of assessment (measurement). Describe comparability	
		of assessment methods if there is more than one group	
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
			applicabl
Quantitative variables	11	Explain how quantitative variables were handled in the analyses.	7-9
		If applicable, describe which groupings were chosen and why	
Statistical methods	12	(a) Describe all statistical methods, including those used to	10-11
		control for confounding	
		(b) Describe any methods used to examine subgroups and	10-11
		interactions	
		(c) Explain how missing data were addressed	10-11
		(d) Cohort study—If applicable, explain how loss to follow-up	10
		was addressed	
		Case-control study—If applicable, explain how matching of	

	<i>Cross-sectional study</i> —If applicable, describe analyt methods taking account of sampling strategy	
Continued on next page10	$(\underline{e})$ Describe any sensitivity analyses	10

BMJ Open: first published as 10.1136/bmjopen-2018-025154 on 5 August 2019. Downloaded from http://bmjopen.bmj.com/ on June 12, 2025 at Agence Bibliographique de l Enseignement Superieur (ABES)

Protected by copyright, including for uses related to text and data mining, Al training, and similar technologies

2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

1 2

Results			
Participants	13*	(a) Report numbers of individuals at each stage of study—eg numbers potentially	Fig 2
		eligible, examined for eligibility, confirmed eligible, included in the study,	12
		completing follow-up, and analysed	
		(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	Table 1,
		information on exposures and potential confounders	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	13
		time	
		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	14
		and their precision (eg, 95% confidence interval). Make clear which confounders	
		were adjusted for and why they were included	
		(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	14
		sensitivity analyses	
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	15,16,17
		imprecision. Discuss both direction and magnitude of any potential bias	
Interpretation	20	Give a cautious overall interpretation of results considering objectives, limitations,	18-22
		multiplicity of analyses, results from similar studies, and other relevant evidence	
Generalisability	21	Discuss the generalisability (external validity) of the study results	17
Other information	n		
Funding	22	Give the source of funding and the role of the funders for the present study and, if	23
-		applicable, for the original study on which the present article is based	

\*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.

For peer review only - http://bmjopen?bmj.com/site/about/guidelines.xhtml

# **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
<b>Primary Subject Heading</b> :	Global health
Secondary Subject Heading:	Paediatrics, Nutrition and metabolism
Keywords:	Nutrition < TROPICAL MEDICINE, PUBLIC HEALTH, EPIDEMIOLOGY, PAEDIATRICS



1 2 3 4 5 6	1	Relative importance of pre- and postnatal determinants of stunting; data mining
7 8 9 10 11	2	approaches to the MINIMat cohort, Bangladesh
12 13 14 15	3	
16 17 18	4	Pernilla Svefors1°, Oleg Sysoev², Eva-Charlotte Ekström¹, Lars-Åke Persson³, Shams El Arifeen⁴, Ruchira Naved⁴, Anisur Rahman⁴,
19 20 21 22	5	Ashraful Islam Khan <sup>4</sup> , Katarina Ekholm Selling <sup>1</sup>
23 24 25 26	6	<sup>1</sup> International Maternal and Child Health, Department of Women's and Children's Health, Uppsala University, Uppsala, Sweden. <sup>2</sup>
27 28 29 30	7	Department of Computer and Information Sciences, Linköping University, Linköping, Sweden, <sup>3</sup> Department of Disease Control,
31 32 33	8	London School of Hygiene and Tropical Medicine, London, UK, 4 International Centre for Diarrhoeal Disease Research, Bangladesh
34 35 36 37 38	9	(iccdr,b), Dhaka, Bangladesh.
30         39         40         41         42         43         44         45         46         47         48         49         50         51         52         53         54         55         56         57         58         59         60	10	*Corresponding author pernilla.svefors@kbh.uu.se

### 12 Abstract

1

6 7		
7 8 9 10	13	<b>Introduction</b> The WHO has set a goal to reduce the prevalence of stunted child growth by 40% by the year 2025. To reach this goal,
11 12 13	14	it is imperative to establish the relative importance of risk factors for stunting to deliver appropriate interventions. Currently, most
14 15 16	15	interventions take place in late infancy and early childhood. This study aimed to identify the most critical pre- and postnatal
17 18 19	16	determinants of linear growth 0–24 months and the risk factors for stunting at two years, and to identify subgroups with different
20 21 22 23	17	growth trajectories and levels of stunting at two years.
24 25		
26 27	18	Methods Conditional inference-tree-based methods were applied to the extensive Maternal and Infant Nutrition Interventions in
28 29 30	19	Matlab (MINIMat) trial database with 309 variables of 2,723 children, their parents, and living conditions, including socioeconomic,
31 32 33 34	20	nutritional and other biological characteristics of the parents; maternal exposure to violence; household food security; breast and
35 36 37	21	complementary feeding; and measurements of morbidity of the mothers during pregnancy and repeatedly of their children up to
38 39 40 41	22	24 months of age. Child anthropometry was measured monthly from birth to 12 months, thereafter quarterly to 24 months.
42 43 44 45	23	<b>Results</b> Birth length and weight were the most critical factors for linear growth 0–24 months and stunting at two years, followed by
46 47 48	24	maternal anthropometry and parental education. Conditions after birth, such as feeding practices and morbidity, were less
49 50 51 52	25	strongly associated with linear growth trajectories and stunting at two years.
53 54 55 56	26	<b>Conclusion</b> The results of this study, together with findings from recent reviews, motivate a change in policy and practice,
57 58 59 60	27	emphasizing the benefit of interventions before conception and during pregnancy to reach a substantial reduction in stunting.

1 2 3 4	28	Strength	s and limitations of this study
5 6 7 8 9	29	•	Assesses the relative public health importance of pre- and post-natal risk factors.
10 11 12 13	30	•	Includes high-quality longitudinal data with low rates of missing data on child growth and a wide range of pre and
14 15 16	31		postnatal household, family, and environmental factors, child characteristics at birth, infant feeding, and morbidity.
17 18 19 20 21	32	•	Some potential important determinants of linear growth were not present in the database.
22 23 24	33	•	Employs decision-tree-based methods that permit the inclusion of a high number of predictor variables, variables of
25 26 27 28	34		different types and automatically discover complex interactions between predictor variables and include them in the
29 30 31 32 33 4 35 36 37 38 9 40 41 42 43 44 54 47 48 9 51 52 54 55 57 58 9 60	35		nodel.

## 37 Introduction

1

7 8 9 10	38	Linear growth is considered to be the best overall indicator of children's present and future health[1, 2] and the reduction of growth
11 12 13	39	failure is one of the targets within the sustainable development agenda. Stunted growth is associated with short-term morbidity
14 15 16 17	40	and mortality, impaired cognitive development, lower future productivity, and increased risk of adult chronic diseases [3]. In 2012,
18 19 20	41	the WHO adopted a resolution on maternal and child undernutrition, targeting a reduction of stunting by 40% by 2025 [4]. Linear
21 22 23	42	growth is most susceptible to environmentally modifiable factors from conception up to two years of age, i.e., the first 1000 days
24 25 26 27	43	when most of the growth faltering takes place [5] [6]. To develop and deliver appropriate interventions, it is imperative to establish
28 29 30	44	the relative importance of stunting risk factors. In addition, the sustainable development health goal has emphasized the
31 32 33 34	45	personalized perspective under the universal coverage of health care. Identifying and targeting high-risk subgroups have thus been
35 36 37	46	highlighted as one of the strategies to reach this goal.
38 39 40 41 42 43 44 45 46 47 48	47	Previous studies employing classical statistical methods have identified a wide range of pre- and post-natal factors
	48	associated with impaired growth [7-12]. Low birth weight, maternal height, maternal education, poverty and inadequate
	49	complementary feeding practices have been recognized as important risk factors [13-15]. Some analyses emphasize the importance
49 50 51	50	of fetal growth restriction for later stunted growth, but rarely is the relative importance of pre- and post-natal factors assessed [16].
52 53 54 55	51	Despite these findings, policy documents and recommendations emphasize interventions especially after birth, and pre-natal
56 57 58 59 60	52	recommendations are usually limited to routine micronutrient supplementation for pregnant women [17-19].

1 2 3 4	53	Despite a wealth of literature relating to the determinants of stunting, studies with a holistic approach, which
5 6 7	54	concurrently account for household, environmental, nutritional, biological, and socioeconomic influences are few. Moreover,
8 9 10 11	55	individuals and groups may be stunted for various reasons and thus respond differently to interventions. Studies that identify risk
12 13 14	56	groups with different probabilities of stunting are, to the best of our knowledge, not yet available. The available studies with a
15 16 17 18	57	multifactorial approach have frequently had a cross-sectional design and have applied traditional statistical methods. As visualized
19 20 21	58	in the WHO's conceptual framework on childhood stunting [20], the causes of stunted linear growth are complex. The number of
22 23 24	59	risk factors and the complexity of the associations of these risk factors with linear growth restriction make traditional statistical
25 26 27 28	60	models ineffective from a predictive perspective. Moreover, classical statistical methods do not have the capacity to identify groups
29 30 31	61	with different risks based on combinations of predictors. Decision trees are popular data mining (DM) methods, which allows for
32 33 34 35	62	the inclusion of a high number of predictor variables, handling variables of different types, automatically discovering complex
36 37 38 39 40 41	63	interactions between predictor variables and including them in the model [21]. Decision-tree-based algorithms can be used to rank
	64	a high number of predictors according to their relative importance for the outcome and to identify subgroups with different risk
42 43 44 45	65	patterns.
46 47 48 49	66	The Maternal and Infant Nutrition Interventions in Matlab (MINIMat) was a randomized prenatal food and multiple
50 51 52	67	micronutrient trial carried out in rural Bangladesh. The frequent follow-up of mothers and children participating in this trial
53 54 55 56	68	resulted in an extensive database, including frequent pre- and post-natal anthropometric assessments, socioeconomic and
57 58 59	69	biological characteristics of the mother and father, information on maternal exposure to violence, household food security, breast-
60	70	and infant-feeding practices, and measurement of morbidity of the mothers during pregnancy and repeatedly of children up to 24

71	months of age. The aim of this study is to, within this Bangladeshi cohort, assess the relative importance of determinants of linear
72	growth from 0–24 months and risk factors for stunting at two years, and to identify risk groups with negative growth trajectories
73	and high prevalence of stunting at two years.
74	Methods
75	Study setting, participants and study design
76	The MINIMat trial (Maternal and Infant Nutrition Interventions in Matlab, isrctn.org identifier: ISRCTN16581394) was carried out
77	in Matlab, Bangladesh, a rural delta region located 57 km southeast of the capital Dhaka. In this area, a health and demographic
78	surveillance system enables early pregnancy identification and longitudinal follow-up. Pregnant women were enrolled in the
79	MINIMat trial and the follow-up included their offspring. MINIMat was a factorial randomized trial primarily evaluating the effect
80	of an early invitation to prenatal food supplementation (versus usual timing) combined with multiple micronutrient
81	supplementation (versus usual program iron-folate) to pregnant women on maternal hemoglobin, birth weight, gestational age at
82	birth, and infant mortality [22]. Further, the participating women were randomly assigned to either counselling for exclusive
83	breastfeeding or a different health education message of equivalent intensity [23]. The MINIMat trial recruited pregnant women
84	from November 2001 to October 2003. When a woman reported to a community health worker that her menstruation was delayed
85	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
86	was missing, the gestational age assessment was based on ultrasound examination. In total, 4436 pregnant women participated,
	<ol> <li>72</li> <li>73</li> <li>74</li> <li>75</li> <li>76</li> <li>77</li> <li>78</li> <li>79</li> <li>80</li> <li>81</li> <li>82</li> <li>83</li> <li>84</li> <li>85</li> </ol>

1 2	~	
3 4	87	giving birth to 3625 live born infants from April 2002 to June 2004. The pregnant women were enrolled at around gestational week
5 6 7 8	88	8. In this analysis, the mothers and children were followed through pregnancy, birth, and up to two years of age.
9 10 11 12	89	Written and oral informed consent was obtained from all participating women and from the parents of the participating
13 14 15	90	children. The Ethical Review Committee at the International Centre for Diarrhoeal Disease Research, Bangladesh, approved the
16 17 18 19 20	91	study (approval registration numbers 2000-025; 2002-031; 2005-004)
21 22 23 24 25	92	
26 27	93	Data collection
28 29 30	94	Predictor and outcome variables are presented in Figure 1, grouped according to the WHO conceptual framework of stunting [20].
31 32 33 34	95	Data were collected using questionnaires, physical examinations, and laboratory analyses. At enrolment, well-trained field workers
35 36 37	96	collected information on women's age, parity, marital status, educational level, occupation, maternal morbidity, socioeconomic
38 39 40 41	97	characteristics, and household food security. Socioeconomic status was assessed based on a range of household assets, and a
42 43 44	98	continuous household asset score, with a mean value of zero, was constructed based on a principal component analysis [24]. A
45 46 47	99	validated household food security scale was created from eleven items with data on frequency of food purchased, cooked, borrowed
48 49 50 51	100	or lent (food and money), and whether there was ready access to adequate meals and snacks [25]. The participating women were
52 53 54	101	also asked whether they had suffered any of thirty morbidity symptoms from twelve different categories, including airway, urinary
57	102	tract, fever, circulation, bowel, or pain symptoms during the last month. A sum score ranging from zero to twelve was created based
58 59 60	103	on absence of symptoms or those not recorded for each category.

Enseignement Superieur (ABES) . Protected by copyright, including for uses related to text and data mining, Al training, and similar technologies.

1 2 3 4	104	Home visits were followed by clinic visits at local health sub-centers. Maternal height and weight were measured at
7	105	around eight weeks of gestation using a stadiometer to the nearest 0.1 cm and an electronic scale (Uniscale; SECA) with a precision
8 9 10 11	106	of 0.10 kg. In the third trimester, paramedics interviewed the participating women in privacy regarding their experiences of
14	107	domestic violence. A modified version of the WHO collaborative study questionnaire was used [26,27], based on the conflict tactic
15 16 17 18	108	scale covering physical, sexual and emotional violence and controlling behavior [28]. Household drinking water was analyzed for
19 20 21	109	arsenic concentration [29].
25	110	A birth notification system allowed birth anthropometry to be measured within 72 hours. In the few cases where the
28	111	newborns were reached after 72 hours, the measurements were adjusted to the time of birth using an SD score transformation,
29 30 31 32	112	assuming that the infants remained in the same relative position in the anthropometric distribution during this period [30]. At
33 34 35 36	113	birth, data on sex, birth weight, length, and breastfeeding practices were collected. During the subsequent two-year study period,
37 38 39	114	the mother-and-child pairs were visited monthly in their homes during the first year, and every three months during the second
42	115	year. On these occasions, data on infant feeding practices, child morbidity and anthropometry were collected. The mothers were
43 44 45 46	116	interviewed about breastfeeding and complementary feeding practices. Breastfeeding practices were categorized into exclusive,
47 48 49	117	predominant, partial, or any breastfeeding for each month from one to twelve months. The total time for exclusive, predominant,
52	118	and any breastfeeding was calculated. The WHO recommendations guided the breastfeeding assessment [31] and results were
53 54 55 56	119	validated with a stable-isotope technique. The classification of exclusive breastfeeding was found to suffer from limited
57	120	misclassification in both directions and to be accurate at the group level [32]. The food given to the infant was categorized into

2	21	semi-solids and solids each month from one to twelve months. The data collection did not include full dietary assessments or
4 5 6 1 7	22	classification of dietary diversity and meal frequency.
8 9 10 [ 11	23	The mothers were also asked whether the child had had any of the following symptoms during the last week; fever, cough,
12 13 14 15	24	difficult breathing, chest in-drawing, rapid breathing, diarrhea, bloody diarrhea and the duration of these symptoms [33].
16 17 1 18	25	Categories were created based on whether the child had suffered from fever, respiratory symptoms, suspected pneumonia, or
19 20 21 22	26	diarrhea, and the sum of days with each symptom and total morbidity calculated from birth to 24 months. To reduce the risk of
23 24 1 25	27	recall bias the mothers were visited monthly with an interview recall period of seven days for child morbidity. One week has been
26 27 1 28 29	28	found to be optimal for this kind of morbidity recall assessment [34].
30 31 32 33	29	Children's weight was measured by SECA beam and electronic scales (UNICEF Uniscale; SECA Gmbh & Co, Hamburg,
34 35 1 36	30	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
37 38 1 39 40	31	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
41 42 43	32	stadiometer. Head and chest circumference were measured with a measuring tape. Two measurements were recorded on each
44 45 1 46 47	33	occasion and the mean was calculated. The equipment was calibrated daily and refresher training on data collection methods,
48 1 49 50 51	34	including the standardization of anthropometric measurements, was conducted periodically.
52 53 54 55		
56 57 58		
59 60		

1		
2 3 4	135	Outcomes
5 6 7	136	Height-for-age z-scores (HAZ) were calculated from the measured length and height data using the program WHOAnthro, based
8 9 10 11	137	on the WHO growth reference for children [35]. Children with a HAZ below minus two SD-scores were classified as stunted. Two
12 13 14 15	150	outcomes were analyzed: stunting at 24 months and the change in HAZ from birth to 24 months, referred to as $\Delta$ HAZ.
17 18	139	
19 20 21	140	Statistical analysis
22 23 24		A database was created with 309 variables characterizing mothers and children in the MINIMat cohort from enrolment in early
25 26 27 28	142	pregnancy up to the time when the children were 24 months of age. The sub-set of records that had height measurements at birth
	143	and 24 months was selected ( <i>n</i> =2 723). The average percent of missing values among all the predictors were 4 %. The highest
32 33 34	111	percent missing were among maternal morbidity data during pregnancy (22%) and categorical monthly child morbidity data (ill
35 36 37 38	145	or not), ranging from 0% to 35% with the highest number of missing observations in the first months. The continuous child
	146	morbidity data however (sum of days with different types illnesses), had no missing values. The most important variables identified
42 43 44		by the random forest analyses and the variables included by the conditional inference trees had less than 1% missing values. The
45 46 47 48	148	missing values of the predictor variables were imputed. To find the best method to impute the missing data we made a simulation
49 50 51	149	study of the performance of the following imputation methods: imputation by variable mean, K-nearest neighbor imputation [36],
54	150	and random forest imputation [37]. The design of the study followed a procedure similar to the strategy described in Jonsson et al.
55 56 57 58	151	[36], see S appendix. Accordingly, we imputed the data by use of the random forest as the simulation study revealed that this
	152	method provided the most accurate imputations.

1 2 3 4	153	Decision trees [21] are data mining methods that allow for specifying an arbitrarily high number of predictor variables,
5 6 7	154	handle variables of different types, automatically discover complex interactions between predictor variables, and include them in
8 9 10 11	155	the model. Traditional decision trees, such as Classification and Regression Trees (CART) have been shown to be biased [38]. This
12	156	motivated us to select the Conditional Inference Trees (CIT) framework, a method that embeds a statistical hypothesis-testing
15 16 17 18	157	framework into a recursive partitioning algorithm used for model building [38]. Conditional inference trees were used in order to
19 20 21		identify sub-groups characterized by combinations of levels of certain predictors with distinct values of $\Delta$ HAZ or prevalence of
24	159	stunting at 24 months. Cross-validation, a well-established model selection method that selects a tree with an optimal predictive
31 32 33 34 35 36 37 38 39 40 41 42	160	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
	162	ensure public health relevance, the minimum number of observations in each terminal node (subgroup) was set to 250.
	163	Conditional random forest (CRF) analyses were performed to assess and rank the importance of predictors with regard to
	164	their ability to explain the variation of the continuous outcome of the change in HAZ from birth to 24 months and the presence of
43 44 45 46	165	stunting at 24 months of age. In conditional random forest analysis, an ensemble of conditional inference trees is created by means
40 47 48 49	166	of drawing subsamples from the original data and fitting a unique randomized conditional inference tree to each sample. Possible
50 51 52 53 54 55 56 57	167	predictors at each split are selected randomly from the complete set of predictors, which leads to a better predictive performance of
	168	the tree ensemble [39]. The importance of a variable is computed by comparing the predictive mean squared error (MSE) from the
	169	original data and a dataset where the corresponding variable values are specified incorrectly, which makes the variable irrelevant
60	170	for the prediction. If the variable does not contribute to the prediction, the MSE is expected to be small when the values of the

1 2 3 171 4	variable are permutated. An aggregated difference between the MSE values over the given ensemble of trees makes up the relative
5 6 172 7	importance of a variable. The random forests analyses were created based on 3000 trees, and the 30 variables with the highest
8 9 173 10 11	importance measure are presented. The programming language R version 3.2.4 [40] and the 'party' package [41] were used for all
12 13 174 14	analyses.
15 16 17 175 18	Patient and public involvement
19 20 176 21 22	No participants were involved in developing the hypothesis, the specific aims or the research questions, nor were they involved in
23 177 24 177 25	developing plans for design or implementation of the study. No participants were involved in the interpretation of study results or
26 27 178 28	write up of the manuscript. There are no plans to disseminate the results of the research to study participants.
29 30 31 179 32 33	
34 35 36 180 37 38 39	Results
40 41 181 42 43	There were 4436 women enrolled into the MINIMat trial, of whom 845 were lost to follow-up before delivery, mainly due to fetal
44 45 182 46	loss, outmigration, or because they withdrew their consent. Of the 3625 live born children, 155 died between birth and two years and
47 48 183 49 50	682 were excluded because of missing anthropometry, at birth or at two years, resulting in 2723 children available for analysis
51 184 52 53	(Figure 2). In the non-analyzed group there was a slightly higher percentage of mothers with more than five years of education,
54 55 185 56 57 58 59 60	younger than 20 years, and belonging to the lowest socioeconomic tertile, and preterm births of children (data not shown).

<sup>2</sup> 186 4		The characteristics of the households, mot	hers, fathers at eight weeks of ges	tation, and children at birth	are given in Table
5 6 187 7 8	1. The par	ticipating mothers had an average age of 2	6 years (SD 5·6), a mean height o	f 150 cm (SD 5·3) and a mea	n weight of 45 kg
9 188 10 11	(SD 6·8) a	it recruitment. One-third of the women we	re underweight, with a BMI belov	v 18·5 at pregnancy week eigl	nt. The average
12 13 189 14	number o	f years of education was similar for mothe	rs and fathers (5 years). The samp	ple of children comprised an	equal proportion
15 16 190 17 18	of girls ar	nd boys, and the average birth length was 2	;7·8 cm (SD 2·2), and of birth wei	ght, 2676 grams (SD 410·5). /	At birth, HAZ was
19 20 191 21	low (mear	n -0·94), and declined further at up to two	years of age with a mean change	of -1 HAZ, resulting in a mea	an HAZ at two
22 23 192 24 25	years of -2	2·0 (Figure 3) and 50% being stunted (girl	s 51·1%, boys 48·5%)		
26 27 193 28 29 30					
31 194 32 33	Table 1. B	aseline characteristics, prevalence of stunti	ng at 24 months, and mean $\Delta$ H/	NZ (change in height-for-age	e Z-score) 0—24
34 35 195 36 37	months ir	ı the MINIMat cohort, Bangladesh.			
38 39 40 41		Characteristics	n/n (%)	Stunted at 24 months <i>n/n</i> (%)	Mean Δ HAZ 0-24 months
42		Mother's age (years)			
43		<20	395/2723 (14.5)	199/395 (50.4)	-0.74
44		20–29	1556/2723 (57.1)	753/1556 (48.4)	-1.05
45		>30	772/2723 (28.4)	417/772 (54.0)	-1.28
46		Mother's education			
47 48		No education	913/2723 (33.5)	556/913 (60.9)	-1.27
48 49		Enrolled in primary school (1-5y)	624/2723 (22.9)	364/624 (58.3)	-1.24
50		Completed primary school (>5y)	1186/2723 (43.6)	449/1186 (37.9)	-0.83
51		Father's education	007/0700 (04.0)	E00/007 (04 4)	1.00
52		No education	867/2723 (31.8)	532/867 (61.4)	-1.29
53		Enrolled in primary school (1-5y)	670/2723 (24.6)	369/670 (55.1)	-1.12
54		Completed primary school (>5y)	1186/2723 (43.6)	468/1186 (39.5)	-0.89
55		Parity	701/2722 (20.0)	249/701 (44.0)	0.76
56		First child	791/2723 (29.0)	348/791 (44.0)	-0.76
57		Second child	774/2723 (28.4)	385/774 (49.7)	-1.09
58		Third or more child Number of saris mother owns	1158/2723 (42.5)	636/1158 (54.9)	-1.28
59		<pre><pre>Number of saris mother owns </pre></pre>	1078/2723 (39.6)	665/1078 (61.5)	-1.26
60		5–8	865/2723 (31.8)	427/865 (49.4)	-1.03
		<u> </u>	000/2120 (01.0)	721/000 (43.4)	-1.00

1				
2	>8	780/2723 (28.6)	277/780 (35.5)	-0.87
<sup>3</sup> 196	Child at birth			
4	Small for Gestational Age (SGA)	1606/2723 (59.0)	972/1606 (60.5)	-1.26
5	Appropriate for Gestational Age (AGA)	1117/2723 (41.0)	397/1117 (35.5)	-0.94
6	Low Birth Weight (LBW)	797/2723 (29.3)	546/797 (68.5)	-0.56
<sup>7</sup> <sub>8</sub> 197	Normal birth weight	1926/2723 (70.7)	823/1926 (42.7)	-1.29
9	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
$\begin{array}{cccccccccccccccccccccccccccccccccccc$	Term	ing at 24 months and ability to explain the proba Figure 4 and 5. HAZ and w wed by maternal height, Sm and parental education. The	d change in height sco ability of stunting at 24 month reight-for-age Z-scores (WAZ) nall for Gestational Age (SGA) e most important factors for a	<b>res from</b> Ins and the I) at birth were the I), maternal

1	
2 3 216 4	Subgroups with different levels of stunting at 24 months and levels of change in height scores from
5 217 6	birth to 24 months
7 8 218 9 10	The conditional inference trees presented in Figure 6 and 7 display subgroups with different probability of stunting at 24 months
11 219 12 13	and levels of $\Delta$ HAZ 0-24 months due to distinctive combinations of levels of certain predictors. The conditional inference trees for
14 15 220 16 17	stunting and $\Delta$ HAZ were composed of subgroups defined by the same predictors, specifically; HAZ at birth, maternal height,
18 221 19 20	father's educational level, and the number of saris owned by the mother. The probability of stunting ranged from 14% to 84%.
<sup>21</sup> 22 22 23 24	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
25 223 26 27	highest probability of stunting at 24 months, at 84%. Children of a father with more than seven years of education, who had HAZ
<sup>28</sup> 224 29 30	at birth above -0·2, had the lowest probability of stunting at 24 months, at 14%. The difference in <b><math>\Delta</math></b> HAZ between the identified
31 32 225 33 34	subgroups of children with the most negative change and the subgroup with the most positive change was 2·22 HAZ. Children who
35 226 36 37	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
<sup>38</sup> <sub>39</sub> 227 40 41 42 43 228 44 45 46	children who were born with a HAZ above 0.19 had the most negative $\Delta$ HAZ (-2·04 HAZ).
47 48 49 50 51	Discussion
52 53 230 54 55	In our analysis of 309 predictors characterizing household, environmental, biological, and socioeconomic factors, we found birth
56 231 57 58	size, maternal anthropometry and parental education to be the most influential for linear growth up to and stunting at 24 months.
<sup>59</sup> <sub>60</sub> 232	Conditions after birth, such as feeding practices and morbidity, were less important for linear growth trajectories and stunting at

Enseignement Superieur (ABES) . Protected by copyright, including for uses related to text and data mining, Al training, and similar technologies.

<sup>2</sup> <sub>3</sub> 233 4	two years. The difference between the identified subgroups of children with the highest and lowest probabilities of stunting was as
5 6 234 7 8	high.
9 10 235 11 12	The extensive database that was available for our analysis covered a wide range of household, family, and environmental
<sup>13</sup> <sub>14</sub> 236 15	factors, child characteristics at birth, feeding, and morbidity. Infant and young child growth was carefully assessed from birth up
16 17 237 18 19	to two years. The MINIMat cohort was implemented in an excellent research infrastructure that fulfills the prerequisites for
<sup>20</sup> 238 21 22	obtaining high-quality longitudinal data. Experienced field workers and study nurses collected data on the 309 variables during
23 24 239 25 26	pregnancy and the following two years. They received repeated training, including standardization exercises, and were supervised
27 240 28 29	by senior medical doctors.
30 31 32 33	Some potential determinants were not present in the database. Household water, sanitation, and hygiene (WASH)
34 35 242 36 37	characteristics were limited to information on arsenic contamination of the drinking water, but diarrhea and other morbidity
38 243 39 40	information were included in our analyses. Further, the cohort did not include the collection of stools for the study of
41 42 43 44	enteropathogens in the child, which may be associated with the risk of stunting [10]. Paternal height, which may be related to fetal
45 245 46 47	growth, was not available [42]. The mothers' smoking habits were not represented in the data, as smoking was extremely rare
<sup>48</sup> 246 49 50 51	among women in the study area.
52 53 54	There were slight differences in basic characteristics of the analyzed and non-analyzed groups. These differences had
55 56 248 57 58	most likely no influence on the primary outcomes of this study. There were no or few missing values of the critical variables that
<sup>59</sup> 249	ranked high in the random forest and defined the sub-groups in the conditional inference trees. A sub-study was carried out to

$^{1}_{3}$ 250	ensure the most accurate method to impute missing data. Thus, it is also highly unlikely that missing data influenced the main
4 5 6 251 7	findings.
8 9 10 252 11	A benefit of applying random forest modelling compared to using conventional models with relative risks or odds ratios
12 13 14 253 15	is that it ranks the predictors according to how important these are for the explaining the outcome. The random forest analysis,
16 17 254 18 19	however, does not provide information on whether the predictors have a positive or negative relation to the outcome. The
<sup>20</sup> 255 21 22	conditional inference trees, on the other hand, display precise information on the priority, size, and direction of the association of
23 24 256 25	the predictors with the outcome. The risk group identification, including the prioritization and relevant cut-offs of risk factors, can
26 27 257 28 29	be of high public health relevance for the design and targeting of appropriate interventions with the most significant benefit.
30 31 32 33	A potential limitation of the conditional inference tree method is that if the data contain two essential and highly
34 35 259 36	correlated predictors, the conditional inference tree method may select only one of them in the analysis, although the other
37 38 260 39 40	predictor might be as important. Further, decision trees do not deliver <i>p</i> -values or confidence intervals to the results. The cross-
41 42 43	validation method, however, ensures that the selected tree is optimal. This validation method was chosen superior to other model
44 45 262 46 47	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
48 263 49 50	computational burden.
51 52 53 264 54	The study setting was a low socioeconomic area in rural Bangladesh, where maternal and child undernutrition in early
55 56 265 57	life still is widespread. The growth trajectories of our cohort were consistent with established growth trajectories in South Asia,
58 59 60 266	where children are born below the WHO growth reference and falter dramatically up to 24 months of age [5]. The sub-continents of

1		
2 3 26 4	7	South Asia and Sub-Saharan Africa share similar proportions of stunted children and faltering patterns. The sub-Saharan African
5 6 26 7	8	children are however, on average born slightly bigger than children in South Asia [5], which makes our results mainly relevant for
8 9 26 10 11	9	the South Asian context.
12 13 14 27 15	0	The most important predictors of stunting at 24 months were different indicators of size at birth, maternal height, asset
16 17 27 18 19	1	score and maternal education. These findings are in line with a multi-country longitudinal study that found birth or enrollment
<sup>20</sup> 27 21 22	2	weight of the infant and maternal height to have the highest cumulative odds ratios for linear growth deficit up to two years of age
23 24 27 25	3	[10]. These results add to the growing evidence that a large part of linear growth faltering already originates in fetal life [10,43,44].
26 27 <u>2</u> 7 28 29	4	In a pooled analysis of 19 birth cohorts with longitudinal follow-up, 20% of stunting was attributable to small-for-gestational-age
30 31 32	5	weight at birth [16]. That study did not include any post-natal factors in the analysis. In a study in Indonesia, neonatal length and
33 34 27 35	6	weight were the strongest predictors of nutritional status and increases in weight and length during infancy [44]. Our study
36 37 27 38 39	7	included both pre- and post-natal factors and, in contrast to most other studies, assessed not only the relative importance of
40 41 27 42	8	different potential predictors, but also the public health importance of each element.
43 44 45 27 46 47	9	In a study with pooled data from five Demographic and Health Surveys in South Asia, maternal height and underweight,
48 49 50	0	household wealth, maternal education, and minimum dietary diversity were found to be the most important factors among
51 52 28 53	1	children aged 6—23 months [15]. Similar results were reported from a study in India [45]. These studies were, however, cross-
54 55 28 56 57 58	2	sectional, without access to birth characteristics.
59 60		

Page 19 of 46

5	.83	Maternal height is a strong determinant of fetal growth [46] that indirectly reflect the epigenetic heredity. Maternal
4 5 6 2 7	.84	height is directly associated with the uterine volume [47], cephalo-pelvic disproportion and subsequent infant and childhood
8 9 2 10	85	stunting, and child mortality [48,49]. In a previous analysis of the MINIMat cohort, a short maternal height was strongly associated
11 12 13 2 14	86	with stunting all the way up to 10 years of age [49]. Thus, factors that well precede pregnancy generate a vicious intergenerational
15 16 2 17	87	cycle, where small mothers give birth to small children of whom a high proportion become and remain stunted. In the conditional
18 19 20 21	88	inference trees for stunting at 24 months, children who were born with a higher HAZ but who had shorter mothers were as likely to
22 23 2 24	.89	be stunted as children with lower HAZ at birth but with a taller mother. This finding suggests that intergenerational improvements
25 26 2 27 28	90	in height are achievable and that interventions with a particular focus on adolescents and women of reproductive health are needed
29 30 2 31	.91	to break the vicious intergenerational cycle.
32 33 34 2 35	.92	A strong relationship between stunting and poverty has been reported from many low-middle income settings [50]. Asset
36 37 2 38 39	.93	score and other socioeconomic markers, such as the number of shoes and saris the mother owned, were highly ranked in the
40 41 42	.94	random forest analysis and categorized subgroups with a higher probability of stunting and undesirable linear growth trajectories.
43 44 2 45 46	.95	Poverty is associated with unfavorable food and sanitation practices that can lead to poor nutrition and an increased occurrence of
47 48 49	96	infections during pregnancy, infancy, and childhood. Poverty increases the risk of maternal stress, depression [51] and weak
50 51 2 52 53	97	mother-to-child interaction and stimulation.
54 55 2 56	.98	The number of shoes and saris the mother owns might also be markers of the woman's status in the household. During
57 58 59 2 60	.99	the last few decades, the importance of women's position in household and society for child nutrition has been emphasized [52].

1 2 3 300 4	Maternal status is associated with food allocation to mother and child, and a higher level of maternal autonomy has been
5 6 301 7	associated with better child weight and lower levels of stunting [53]. The subordinate position of women in South Asia has been
8 9 302 10 11	suggested to be a contributor to the high prevalence of child undernutrition in the region, compared to other areas with equivalent
12 13 303 14	levels of economic growth and food security [52].
15 16 17 304 18 19	An acknowledged way of increasing women's position is through improved education. The remarkable health
<sup>20</sup> 305 <sup>21</sup> <sup>22</sup>	achievements in Bangladesh over the past two decades can partly be attributed to the progress in access to education, especially at
23 24 306 25	primary level and for girls [54]. However, there is a considerable risk of not completing primary school for both girls and boys [55].
26 27 307 28	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
29 30 31 308 32	conditional decision trees models for stunting and change in HAZ, the cut-off values for paternal and maternal education in the
33 34 309 35 36	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,
37 310 38 39	furthering the importance of girls and boys not only enrolling in but also continuing at school.
40 41 42 311 43	It may seem contradictory that children who were born with a very short length had the smallest change in HAZ. This
44 45 312 46 47	finding most likely reflects a situation where linear growth had already been severely restricted in fetal life.
48 49 313 50 51	A multi-country pooled analysis of cohort studies showed that a higher cumulative burden of diarrhea increased the risk
52 53 314 54	of stunting [57]. In situations, where measles still occurred, its impact on growth and mortality risks were repeatedly documented
55 56 315 57 58	[58]. One explanation to the discrepancy between our results and previous findings could be Bangladesh's remarkable success in
<sup>59</sup> 316 60	achieving the globally highest coverage of oral rehydration therapy in diarrhea [59], which may have reduced the impact on linear

1 2 3 317	growth. Another factor is the almost universal immunization coverage [60,61] that has reduced or partly eliminated immunization-
4 5 6 318	preventable morbidity and the subsequent effect on growth. Our previous publications on the MINIMat prenatal nutrition
7 8	preventable morbially and the subsequent effect on growth. Our previous publications on the Minimat prenatal nutrition
9 319 10 11	interventions' effects on child growth and mortality were not mediated through morbidity [22,62], further supporting the modest
12 13 320 14	impact of child morbidity on linear growth in our sample [33]. In other settings with lower coverage of diarrhea treatment and
15 16 321 17 18 19	immunization, the relative importance of these factors may be greater.
20 322 21 22	Suboptimal infant and early childhood feeding practices have, in earlier studies, been reported as significant risk factors
23 24 323 25 26	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–
27 324 28 29	24 months had been randomized to appropriate complementary foods [64]. The infant feeding variables included in our analysis
<sup>30</sup> 31 32	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
33 34 326 35 36	documented effects of complementary feeding programs on stunting, these interventions are often the priority in efforts to combat
37 327 38 39 40	stunting.
41 42 43	The nutrition interventions from pre-conception to two years of age currently recommended by the WHO include efforts
44 45 329 46 47	to ensure exclusive breastfeeding, adequate complementary feeding, appropriate nutritional care of sick and malnourished
48 330 49 50	children and proper intake of vitamin A, iron and iodine for women and children [18]. All of these, except micronutrient
51 52 331 53 54	supplementation to pregnant women, are focused on the postnatal period from birth up to two years. Our results strengthen the
55 332 56 57 58 59 60	evidence that the process of becoming stunted already begins in utero, as well as the importance of intergenerational effects.

1 2 3 333 4	Although worthwhile, the present focus on postnatal interventions results in missed opportunities to intervene before or during the
5 6 334 7 8	first nine months when the process of stunting is established.
9 10 335 11 12	So, what possibilities do we have to improve the postnatal linear growth trajectories prenatally? Attained height is mainly
13 14 15	dependent on one's genetic potential for linear growth, in turn determined by DNA sequence polymorphism [65,66] and epigenetic
16 17 337 18 19	heredity [67], and to some extent the environment. The modulation of non-DNA sequence epigenetic heredity has been proposed to
<sup>20</sup> 338 21 22	be one of the leading factors explaining variations in height and height changes over generations[67], especially in more deprived
23 24 339 25	populations [68]. Postnatal interventions can influence factors in the environment that constrain the ability to increase linear
26 27 340 28 29	growth, while prenatal interventions also have the potential to modulate the actual growth potential through an epigenetic
30 31 341 32	modification that results from changes to gene expression in response to the fetal environment.
33 34 35 342 36	Established prenatal nutritional interventions include balanced energy-protein supplementation, multiple micronutrient
37 38 343 39 40	supplements, and nutritional counseling and education. Unfortunately, most studies evaluating these interventions report only
41 42 43 44	birth weight, not length, which is why evidence to directly assess the effect on fetal linear growth is limited. Meta-analyses and
44 45 345 46 47	randomized trials evaluating these interventions report their positive impact on birth weight and a reduced risk of LBW [69-76].
48 346 49 50	Effect sizes vary from increases in birth weight of 20—200g, with the smallest effects seen in studies of multiple micronutrients and
51 52 347 53	bigger effects seen by balanced energy-protein and lipid-based nutrient supplements. Considerable heterogeneity in growth
54 55 348 56 57	response is common, and is related to the mother's nutritional status when entering pregnancy and possibly also to the genetic
<sup>58</sup> 349 59 360	potential to benefit. In the MINIMat food and micronutrient interventions, all women received food supplementation, but they were

1	
<sup>2</sup> <sub>3</sub> 350 4	randomized to an early invitation to supplementation (week 9) or the usual program start of supplementation (week 20). Children
5 6 351 7	of mothers who participated in food supplementation from early pregnancy (versus the usual start) had a 13% reduction in
8 9 352 10 11	stunting up to five years [62].
12 13 14 353 15	There is increasing evidence that preconception interventions may be even more appropriate[77]. A few trials examining
16 17 354 18	the effect of interventions initiated before pregnancy are underway, but few results have so far been published [78]. Preconception
19 20 355 21 22	interventions have the potential to bring about epigenetic modulation and improved growth in present and future generations.
23 24 356 25	Thus, the launch and evaluation of interventions targeting adolescent and women of reproductive age that focus on adequate
26 27 357 28 29	health, education, and nutrition before and during pregnancy is needed, especially in South Asia with its high burden of maternal
<sup>30</sup> 31 358 32	undernutrition and young age at first pregnancy [79]. Targeting high-risk subgroups, in this setting characterized by short, poor,
33 34 359 35 36	women with low education, can be another strategy to address the intractable problem of stunting.
37 38 39 360 40 41 42	Contributors
43 44 361 45 46	PS contributed to study design, data analysis and interpretation of the results and had the main responsibility of writing the paper.
47 48 49	LÅP and SEA were principal investigators of the MINIMat project. ECE, LÅP and KES contributed to the study design. ECE, RN, AR
50 51 363 52 53	and AIK took part in and supervised data collection. PS, OS, and KES analysed the data. All authors contributed to the preparation
54 364 55 56 57	of the database, interpretation of the results and reviewed and approved the final version of the manuscript.
58 59 365 60	Competing interests

$\frac{2}{3}$ 366	The authors declare that they have no competing interests
4	
5	
$^{6}_{7}$ 367	Data sharing statment
8 9 368	Data are available from the authors upon reasonable request and with permission from the principal investigator of the MINIMat
10 11 369 12	study.
$^{13}_{14}370$	
15	
16 17 18 371	Funding
19 20	
21	
22 372 23 24	The MINIMat research study was funded by the icddr,b, United Nations Children's Emergency Fund, Swedish International
25 373 26 373 27	Development Cooperation Agency (Sida), UK Medical Research Council, Swedish Research Council, Department for International
28 29 374 30	Development (DFID), Japan Society for the Promotion of Science, Child Health and Nutrition Research Initiative, Uppsala
31 32 375 33 34	University, and US Agency for International Development. The funding agencies had no role in the design and conduct of the
35 36 37	study, in the collection, analysis, and interpretation of the data, or in the preparation, review, or approval of the manuscript.
38	
39 40 41 377	Acknowledgements
42 43 44	
45 46 378 47	We thank the participants and their families in Matlab for their continuing involvement in the MINIMat trial, and the field-team
48 49 379 50	members and data management staff for their excellent work.
51 52 53	
54 55	
56 57 58	
59 60	

1 2 3 3 4 5 6 7 8 382 9 10 11 12 13	
14 15 16 17 18 19 20	
21 22 23 24 25 26	
27 28 29 30 31 32	
33 34 35 36 37 38 39	
40 41 42 43 44 45	
46 47 48 49 50 51	
52 53 54 55 56 57 58 59 60	

1 2 3	384	Legend to Figures
4 5 6	385	Figure 1. Factors, variables and outcomes included in the analysis of data from the MINIMat cohort, Bangladesh. Grouping
7 8 9 10	386	according to the WHO conceptual framework on childhood stunting [20]
11 12 13 14	387	Figure 2. Flow chart of pregnant women and their children included in the data mining analyses of the MINIMat cohort from
15	388	conception to two years of age.
19	389	<b>Figure 3.</b> Height-for-age Z-scores from birth to 24 months in the MINIMat cohort in rural Bangladesh.
23	390	Figure 4. Conditional random forest plot ranking the relative importance of 30 predictors with regard to their ability to explain the
27 28 29	391	presence of stunting at 24 months of age. The MINIMat cohort in rural Bangladesh. Colour coding according to the WHO
30 31 32 33	392	conceptual framework on causes of stunting.
34 35 36 37	393	Figure 5. Conditional random forest plot ranking the relative importance of 30 predictors with regard to their ability to explain the
38 39 40	394	variation in change in HAZ ( $\Delta$ HAZ) from birth to 24 months of age. The MINIMat cohort in rural Bangladesh. Colour coding
43 44	395	according to the WHO conceptual framework on causes of stunting.
45 46 47 48	396	Figure 6. Conditional inference tree identifying sub-groups with different probabilities of stunting at 24 months. The MINIMat
49 50 51 52	397	cohort in rural Bangladesh.
53	398	<b>Figure 7.</b> Conditional inference tree identifying sub-groups with different mean change in HAZ ( $\Delta$ HAZ=HAZ <sub>24</sub> –HAZ <sub>0</sub> ) 0–24
57 58 59 60	399 400	months within the MINIMat cohort in rural Bangladesh.
	TUU	26 For peer review only - http://bmjopen.bmj.com/site/about/guidelines.xhtml

1 2 3 401 4		
5		
7 402 8		
9		
10 11	<b>D</b> (	
12 <sup>403</sup>	Reference	2S
13 14 15		
$^{16}_{17}404$	1	Pradhan M, Sahn DE, Younger SD. Decomposing world health inequality. <i>Journal of Health Economics</i>
18 405		2003; <b>22</b> :271-93. doi:10.1016/S0167-6296(02)00123-6
19 20		
21 406	2	Victora CG, Adair L, Fall C, <i>et al.</i> Maternal and child undernutrition: consequences for adult health and human
$\frac{22}{23}407$		capital. <i>Lancet</i> 2008; <b>371</b> :340—57. doi:10.1016/S0140-6736(07)61692-4
24 25 408		
26	3	Black RE, Victora CG, Walker SP, <i>et al.</i> Maternal and child undernutrition and overweight in low-income and
27 409 28		middle-income countries. <i>The Lancet</i> 2013.
29 30 410	4	General Assembly, UNICEF. Transforming our world: the 2030 Agenda for Sustainable Development. <i>United</i>
$30^{31}$ 411	4	Nations 2015;:1–35.
32 <sup>111</sup> 33		
34 412	5	Victora CG, de Onis M, Hallal PC, <i>et al.</i> Worldwide timing of growth faltering: revisiting implications for
35 36 413		interventions. <i>Pediatrics</i> 2010; <b>125</b> :e473—80. doi:10.1542/peds.2009-1519
37		
<sup>38</sup> 414 39	6	Black RE, Allen LH, Bhutta ZA, <i>et al.</i> Maternal and child undernutrition: global and regional exposures and health
40 415 41		consequences. <i>The Lancet</i> 2008; <b>371</b> :243—60. doi:10.1016/S0140-6736(07)61690-0
42 43 416	7	Martorell R, Ramakrishnan U, Schroeder DG, <i>et al.</i> Intrauterine growth retardation, body size, body composition
44 45 417	,	and physical performance in adolescence. <i>Eur J Clin Nutr</i> 1998; <b>52 Suppl 1</b> :S43—52—discussionS52—3.
46		······································
47 418 48	8	Frongillo EA Jr. <i>Symposium: Causes and Etiology of Stunting. J Nutr</i> 1999 <b>,129</b> :529S–530S. doi:10.1016/S0140-
49 419		6736(94)91337-4
50 51		
52 420	9	Cole TJ, Parkin JM. Infection and its effect on the growth of young children: A comparison of The Gambia and
<sup>53</sup> 421 54 55		Uganda. <i>Trans R Soc Trop Med Hyg</i> 1977; <b>71</b> :196—8. doi:10.1016/0035-9203(77)90005-0
55 56 422 57	10	Investigators M-EN. Childhood stunting in relation to the pre- and postnatal environment during the first 2 years
<sub>58</sub> 423		of life: The MAL-ED longitudinal birth cohort study. <i>PLoS Medicine</i> 2017;:1—21. doi:10.1371/journal.pmed.1002408
59 60		

Enseignement Superieur (ABES) . Protected by copyright, including for uses related to text and data mining, Al training, and similar technologies.

	28	
<sup>60</sup> 451		body composition: the MINIMat trial, Bangladesh. <i>Acta Paediatrica</i> 2013; <b>102</b> :815—23. doi:10.1111/apa.12282
58 59 450	23	Khan AI, Hawkesworth S, Ekström E-C, <i>et al.</i> Effects of exclusive breastfeeding intervention on child growth and
56 449 57		Randomized Trial. <i>JAMA</i> 2012; <b>307</b> :2050—9. doi:10.1001/jama.2012.4061
<sup>54</sup> 448 55		Maternal Hemoglobin, Birth Weight, and Infant Mortality Among Children in Bangladesh: The MINIMat
52 53 447	22	Persson LÅ, Arifeen S, Ekström E-C, <i>et al.</i> Effects of Prenatal Micronutrient and Early Food Supplementation on
49 50 446 51	21	Breiman L, Friedman J, Stone CJ, <i>et al. Classification and Regression Trees</i> . Chapman and Hall/CRC 1984.
47 445 48		stunting prevention. <i>Matern Child Nutr</i> 2013; <b>9</b> :27—45. doi:10.1111/mcn.12088
44 45 46 444	20	Stewart CP, Iannotti L, Dewey KG, <i>et al.</i> Contextualising complementary feeding in a broader framework for
42 43 443		doi:10.1016/S0140-6736(13)60996-4
41 442		improvement of maternaland child nutrition: what can be done and at what cost? <i>The Lancet</i> 2013; <b>382</b> :452–77.
<sup>39</sup> 441	19	Bhutta ZA, Das JK, Arjumand R, <i>et al.</i> Maternal and Child Nutrition 2 Evidence-based interventions for
37 440 38		health and nutrition. WHO 2013;:1—144.
34 35 439 36	18	World Health Organization. Essential Nutrition Actions: improving maternal, newborn, infant and young child
<sup>32</sup> 438	,	2012;;1-32.
29 30 31 437	17	World Health Organization. WHO Guidline: Daily iron and folic acid supplementation in pregnant women. <i>WHO</i>
27 28 436		and preterm birth in low- and middle-income countries. <i>Int J Epidemiol</i> 2013; <b>42</b> :1340—55. doi:10.1093/ije/dyt109
24 25 26 435	16	Christian P, Lee SE, Donahue Angel M, <i>et al.</i> Risk of childhood undernutrition related to small-for-gestational age
<sup>23</sup> 24 434		<i>Medicine</i> 2017; <b>187</b> :144–54. doi:10.1016/j.socscimed.2017.06.017
21 22 433		from nationally representative data from Afghanistan, Bangladesh, India, Nepal, and Pakistan. Social Science &
19 20 432	15	Kim R, Mejía-Guevara I, Corsi DJ, <i>et al.</i> Relative importance of 13 correlates of child stunting in South Asia: Insights
<sup>17</sup> 431		nutrition in Bangladesh. International Journal for Equity in Health 2006; <b>5</b> :1—10. doi:10.1186/1475-9276-5-15
14 15 16 430	14	Hong R, Banta JE, Betancourt JA. Relationship between household wealth inequality and chronic childhood under-
11 428 12 13 429	13	MS MRAM, RN MWMPL, BS AFMC. Factors Associated with Stunting Among 0-23 Months-Old Children in Rural Bangladesh. <i>Journal of Nutrition Education and Behavior</i> 2017; <b>49</b> :S21. doi:10.1016/j.jneb.2017.05.061
9 427 10		<i>Biosoc Sci</i> 2006; <b>39</b> :161. doi:10.1017/S0021932006001295
7 426 <sup>8</sup> 427	12	Rahman A, Chowdhury S. Determinants of Chronic Malnutrtion Among Preschool Children In Bangladesh. J
4 425 5 6		infancy in rural Malawi. <i>Acta Paediatrica</i> 2002; <b>91</b> :1364—70. doi:10.1111/j.1651-2227.2002.tbo2835.x
3 424	11	Espo M, Kulmala T, Maleta K, <i>et al.</i> Determinants of linear growth and predictors of severe stunting during
1 2 424		

For peer review only - http://bmjopen.bmj.com/site/about/guidelines.xhtml

Page 29 of 46

1		
$^{2}_{3}$ 452	24	Gwatkin DR, Rustein S, Johnson K, et al. Socio-economic differences in health nutrition and population in
4 453		Bangladesh. Washington, DC: HNP/Poverty Thematic Group of the World Bank 2000.
5 6		
7 454	25	Frongillo EA, Chowdhury N, Ekström E-C, <i>et al.</i> Understanding the Experience of Household Food Insecurity in
8 9 455		Rural Bangladesh Leads to a Measure Different from That Used in Other Countries
10		
<sup>11</sup> 456 12	.]Nutr2003;:2	4158—62.
13 14 457	- (	
15	26	Straus MA, Douglas EM. A Short Form of the Revised Conflict Tactics Scales, and Typologies for Severity and
16 458 17		Mutuality. <i>Violence and Victims</i> 2004; <b>19</b> :1—15.
18 450	97	WHO. WHO Multi-country Study on Women's Health and Domestic Violence against Women. <i>WHO</i> 2005;:1–38.
19 <sup>439</sup> 20	27	wite. Wite mail county study on women's realitratic bomeste violence against women. Wite 2005, 1 '30.
21 460	28	Asling-Monemi K, Naved RT, Persson LA. Violence against women and the risk of fetal and early childhood growth
22 23 461		impairment: a cohort study in rural Bangladesh. <i>Arch Dis Child</i> 2009; <b>94</b> :775–9. doi:10.1136/adc.2008.144444
24		
<sup>25</sup> 462	29	Rahman A, Vahter M, Smith AH, <i>et al.</i> Arsenic Exposure During Pregnancy and Size at Birth: A Prospective
27 463		Cohort Study in Bangladesh. <i>American Journal of Epidemiology</i> 2008; <b>169</b> :304—12. doi:10.1093/aje/kwn332
28 29		
30 464 21	30	Arifeen SE, Black RE, Caulfield LE, <i>et al.</i> Infant growth patterns in the slums of Dhaka in relation to birth weight,
<sup>31</sup> 32 465		intrauterine growth retardation, and prematurity. <i>Am J Clin Nutr</i> 2000; <b>72</b> :1010—7.
33 34 ACC		
<sup>34</sup> 466 35	31	World Health Organization. Indicators for assessing breastfeeding practices. <i>WHO</i> 1991;:1–16.
36 37 467	32	Moore SE, Prentice AM, Coward WA, <i>et al.</i> Use of stable-isotope techniques to validate infant feeding practices
38 39 468	5-	reported by Bangladeshi women receiving breastfeeding counseling. <i>Am J Clin Nutr</i> 2007;:1075–82.
39 <del>4</del> 00 40		reported by Bangladeshi women receiving breastreading counseling. 7mg cun van 2007,.1075–02.
<sup>41</sup> 469	33	Kallioinen M, Ekström E-C, Khan AI, <i>et al.</i> Prenatal early food and multiple micronutrient supplementation trial
42 43 470		reduced infant mortality in Bangladesh, but did not influence morbidity. <i>Acta Paediatrica</i> 2017; <b>382</b> :427—8.
44 45 471		doi:10.1111/apa.14009
46		
47 48472	34	Arnold BF, Galiani S, Ram PK, <i>et al.</i> Optimal Recall Period for Caregiver-reported Illness in Risk Factor and
49 473		Intervention Studies: A Multicountry Study. <i>American Journal of Epidemiology</i> 2013; <b>177</b> :361—70.
50 51 474		doi:10.1093/aje/kws281
52		
<sup>53</sup> 54 475	35	Onis M de, Onyango AW, Borghi E, <i>et al.</i> Development of a WHO growth reference for school-aged children and
<sup>55</sup> 476		adolescents. <i>Bulletin of the World Health Organization</i> 2007; <b>85</b> :660—7. doi:10.1590/S0042-96862007000900010
56 57		
58 477 59	36	Jönsson P, Wohlin C. An Evaluation of K-Nearest Neighbour Imputation Using Likert Data. Proceedings of th
60 478		International Symposium on Software Metrics 2004;:108—18.

Enseignement Superieur (ABES) . Protected by copyright, including for uses related to text and data mining, Al training, and similar technologies.

#### BMJ Open

1		
$^2_3$ 479	37	Stekhoven DJ, Buhlmann P. MissForest—non-parametric missing value imputation for mixed-type data.
4 480 5		<i>Bioinformatics</i> 2012;:112–8.
6 7 481	38	Torsten Hothorn KHAZ. ctree: Conditional Inference Trees. www.cran.r-project.org. 2016;:1—21.https://cran.r-
8 9 482		project.org/web/packages/partykit/vignettes/ctree.pdf (accessed 24 Mar2017).
10 11 483	39	Hastie T, Tibshirani R, Friedman J. <i>The Elements of Statistical Learning</i> : New York, NY: : Springer Science &
12 13 484	55	Business Media 2013.
14 15 <sub>495</sub>		
16 485 17	40	R Core Team. R: A language and environment for stastical computing. 2017. httpa://www.R-project.org/
<sup>18</sup> 486 19	41	Achim Zeileis THKH. party: A Laboratory for Recursive Partytioning. 2017;:1–18.
20 21 487	42	Wills AK, Chinchwadkar MC, Joglekar CV, <i>et al.</i> Maternal and paternal height and BMI and patterns of fetal
22 23 488	-	growth: The Pune Maternal Nutrition Study. <i>Early Human Development</i> 2010; <b>86</b> :535—40.
<sup>24</sup> 25 489		doi:10.1016/j.earlhumdev.2010.07.002
25 <sup>10</sup> 26		
27 490	43	Christian P, Lee SE, Donahue Angel M, <i>et al.</i> Risk of childhood undernutrition related to small-for-gestational age
28 29 491 30		and preterm birth in low- and middle-income countries. <i>Int J Epidemiol</i> 2013; <b>42</b> :1340—55. doi:10.1093/ije/dyt109
30 31 32 492	44	Schmidt MK, Muslimatun S, West CE, <i>et al.</i> Nutritional Status and Linear Growth of Indonesian Infants in West
<sup>33</sup> 493		Java Are Determined More by Prenatal Environment than by Postnatal Factors. <i>J Nutr</i> 2002;:2202—7.
34 35		, , , , , , , , , , , , , , , , , , ,
36 494	45	Corsi DJ, Mejía-Guevara I, Subramanian SV. Risk factors for chronic undernutrition among children in India:
37 38 495		Estimating relative importance, population attributable risk and fractions. Social Science & Medicine 2016; <b>157</b> :165—
<sup>39</sup> 496 40		85. doi:10.1016/j.socscimed.2015.11.014
41 42 497	46	Kramer MS. The Epidemiology of Adverse Pregnancy Outcomes: An Overview. <i>J Nutr</i> 2003;:1592S—1596S.
43 44		
45 498 46	47	A WHO Collaborative Study of Maternal Anhropomehtry and Pregnancy Outcomes. <i>International Journal of</i>
46 47 499		<i>Gynecology and Obstetrics</i> 1996; <b>57</b> :1–15.
48 49 500	48	Özaltin E. Association of Maternal Stature With Offspring Mortality, Underweight, and Stunting in Low- to
50 51 501		Middle-Income Countries. <i>JAMA</i> 2010; <b>303</b> :1507—21. doi:10.1001/jama.2010.450
52 53 54 502	10	Svefors P, Rahman A, Ekström E-C, <i>et al.</i> Stunted at 10 Years. Linear Growth Trajectories and Stunting from Birth
<sup>54</sup> <sup>502</sup>	49	
50		to Pre-Adolescence in a Rural Bangladeshi Cohort. <i>PLoS ONE</i> 2016, <b>11</b> :e0149700—18.
57 504 58		doi:10.1371/journal.pone.0149700
59 60 505	50	Van de Poel E, Hosseinpoor AR, Speybroeck N, <i>et al.</i> Socioeconomic inequality in malnutrition in developing
506		countries. <i>Bulletin of the World Health Organization</i> 2008; <b>86</b> :282—91. doi:10.1590/S0042-96862008000400013
	30	For peer review only - http://bmjopen.bmj.com/site/about/guidelines.xhtml

Page 31 of 46

1		
$^2_3$ 507	51	Bradley RH, Corwyn RF. Socioeconomic Status and Child Development. Annual Review of Phsycology 2002; <b>53</b> :371–
4 508 5		99.
6 7 509 8	52	Ramalingaswami V, U J, J R. The Asian enigma. 1996.
9 10 510	53	Shroff M, Griffiths P, Adair L, <i>et al.</i> Maternal autonomy is inversely related to child stunting in Andhra Pradesh,
11 511 12 13		India. <i>Matern Child Nutr</i> 2009; <b>5</b> :64—74. doi:10.1111/j.1740-8709.2008.00161.x
14 512	54	Chowdhury AMR, Bhuiya A, Chowdhury ME, <i>et al.</i> The Bangladesh paradox: exceptional health achievement
15 16 513 17		despite economic poverty. <i>The Lancet</i> 2013; <b>382</b> :1734—45. doi:10.1016/S0140-6736(13)62148-0
<sup>18</sup> 514	55	Asadullah MN, Chaudhury N. Girls school enrolement figures are good in Bangladesh but drop out rates are
20 515 21 22		worrying. <i>Development and Cooperation</i> 2017; <b>59</b> :447—72. doi:10.1086/681929
<sub>23</sub> 516	56	UNESCO. EFA 2015 national review: Bangladesh; 2014. unesdoc.unesco.org.
<sup>24</sup> 517 25		http://unesdoc.unesco.org/images/0023/002305/230507e.pdf (accessed 13 Feb2018).
27 518 28	57	Checkley W, Buckley G, Gilman RH, <i>et al.</i> Multi-country analysis of the effects of diarrhoea on childhood stunting.
29 519 30		<i>Int J Epidemiol</i> 2008; <b>37</b> :816—30. doi:10.1093/ije/dyn099
$^{31}_{32}520$		Koster FT, Curlin GC, Aziz KMA, <i>et al.</i> Synergistic impact of measels and diarrhoea on nutrtion and mortality in
<sup>33</sup> 521 34 35		Bangladesh. <i>Bulletin of the World Health Organization</i> 1981;:901—8.
36 522	59	Arifeen DrPH El S, MPH AC, ScD LR, <i>et al.</i> Community-based approaches and partnerships: innovations in
37 38 523 39		health-service delivery in Bangladesh. <i>The Lancet</i> 2013; <b>382</b> :2012–26. doi:10.1016/S0140-6736(13)62149-2
40 524 41 42	60	WHO, UNICEF. Bangladesh: WHO and UNICEF estimates of immunization coverage: 2016 revision. 2016;:1—30.
43 525 44	61	Chowdhury AMR, Bhuiya A, Chowdhury ME, <i>et al.</i> The Bangladesh paradox: exceptional health achievement
45 526 46		despite economic poverty. <i>The Lancet</i> 2013; <b>382</b> :1734—45. doi:10.1016/S0140-6736(13)62148-0
47 48 527	62	Khan AI, Kabir I, Ekstrom EC, <i>et al.</i> Effects of prenatal food and micronutrient supplementation on child growth
49 528 50 51		from birth to 54 months of age: a randomized trial in Bangladesh. <i>Nutr J</i> 2011.
52 529		Martorell R. Physical growth and development of the malnourished child: Contributions from 50 years of research
<sup>53</sup> 54 55		at INCAP. <i>Food Nutr Bull</i> 2010; <b>31</b> :68—82.
<sup>56</sup> 531 57	64	Imdad A, Yakoob MY, Bhutta ZA. Impact of maternal education about complementary feeding and provision of
58 532		complementary foods on child growth in developing countries. <i>BMC Public Health</i> 2011 <b>;11</b> :S25. doi:10.1186/1471-
<sup>59</sup> 60 533		2458-11-S3-S25

Enseignement Superieur (ABES) . Protected by copyright, including for uses related to text and data mining, Al training, and similar technologies.

#### BMJ Open

$^2_3$ 534	65	Weedon MN, Frayling TM. Reaching new heights: insights into the genetics of human stature. Trends in Genetics
4 535 5		2008; <b>24</b> :595—603. doi:10.1016/j.tig.2008.09.006
6	66	
7 536	66	Silventoinen K, Sammalisto S, Perola M, <i>et al.</i> Heritability of Adult Body Height: A Comparative Study of Twin
9 537 10		Cohorts in Eight Countries. <i>Twin Research</i> 2003; <b>6</b> :399—408. doi:10.1375/twin.6.5.399
11 538	67	Simeone P, Alberti S. Epigenetic heredity of human height. <i>Physiological Reports</i> 2014; <b>2</b> :e12047–7.
12 13 539		doi:10.14814/phy2.12047
14 15		
16 540	68	Silventoinen K, Kaprio J, Lahelma E, <i>et al.</i> Relative Effect of Genetic and Environmental Factors on Body Height:
<sup>17</sup> 541 18		Differences Across Birth Cohorts Among Finnish Men and Women. American Journal of Public Health
19 542		2000; <b>90</b> :627–30.
20 21		
22 543 23	69	Imdad A, Bhutta ZA. Effect of balanced protein energy supplementation during pregnancy on birth outcomes.
<sup>23</sup> 544 24		BMC Public Health 2011;11:S17. doi:10.1186/1471-2458-11-S3-S17
25 26 545	70	Fall CHD, Fisher DJ, Osmond C, <i>et al.</i> Multiple Micronutrient Supplementation during Pregnancy in Low-Income
27	70	
28 546 29 5 47		Countries: A Meta-Analysis of Effects on Birth Size and Length of Gestation. <i>Food Nutr Bull</i> 2009; <b>30</b> :S533–46.
30 547 31		doi:10.1177/15648265090304S408
32 548	71	Ramakrishnan U, Grant FK, Goldenberg T, <i>et al.</i> Effect of Multiple Micronutrient Supplementation on Pregnancy
33 <sup>3</sup> 34 549	,	and Infant Outcomes: A Systematic Review. <i>Paediatric and Perinatal Epidemiology</i> 2012; <b>26</b> :153—67.
<sup>35</sup> 36 550		doi:10.1111/j.1365-3016.2012.01276.x
37		
<sup>38</sup> 551 39	72	West KP, Shamim AA, Mehra S, <i>et al.</i> Effect of Maternal Multiple Micronutrient vs Iron—Folic Acid
40 552		Supplementation on Infant Mortality and Adverse Birth Outcomes in Rural Bangladesh: The JiVitA-3 Randomized
41 42 553		Trial. JAMA 2014; <b>312</b> :2649–58. doi:10.1001/jama.2014.16819
43		
44 45 554	73	Johnson W, Darboe MK, Sosseh F, <i>et al.</i> Association of prenatal lipid-based nutritional supplementation with fetal
46 555 47		growth in rural Gambia. <i>Matern Child Nutr</i> 2017, <b>13</b> :e12367. doi:10.1111/mcn.12367
48		
49 556 50	74	Ashorn P, Alho L, Ashorn U, <i>et al.</i> Supplementation of Maternal Diets during Pregnancy and for 6 Months
<sup>50</sup> 51 557		Postpartum and Infant Diets Thereafter with Small-Quantity Lipid-Based Nutrient Supplements Does Not
<sup>52</sup> 558 53		Promote Child Growth by 18 Months of Age in Rural Malawi: A Randomized Controlled Trial. <i>J Nutr</i>
54 559 55		2015; <b>145</b> :1345—53. doi:10.3945/jn.114.207225
56		Adu-Afarwuah S, Lartey A, Okronipa H, <i>et al.</i> Lipid-based nutrient supplement increases the birth size of infants
57 560	75	
<sup>58</sup> 561		of primiparous women in Ghana. <i>Am J Clin Nutr</i> 2015; <b>101</b> :ajcn.091546—846. doi:10.3945/ajcn.114.091546
60		

1		
$^2_3$ 562	76	Huybregts L, Roberfroid D, Lanou H, <i>et al.</i> Prenatal food supplementation fortified with multiple micronutrients
4 563		increases birth length: a randomized controlled trial in rural Burkina Faso. American Journal of Clinical Nutrition
5 6 564		2009; <b>90</b> :1593—600. doi:10.3945/ajcn.2009.28253
7		
8 9 565	77	Gluckman PD, Hanson MA, Cooper C, <i>et al.</i> Effect of In Utero and Early-Life Conditions on Adult Health and
<sup>10</sup> 566		Disease. <i>http://dxdoiorg/101056/NEJMra0708473</i> 2009; <b>359</b> :61—73. doi:10.1056/NEJMra0708473
11 12		
13 567	78	Potdar RD, Sahariah SA, Gandhi M, <i>et al.</i> Improving women's diet quality preconceptionally and during gestation:
14 15 568		effects on birth weight and prevalence of low birth weighta randomized controlled efficacy trial in India (Mumbai
<sup>16</sup> 569 17		Maternal Nutrition Project). American Journal of Clinical Nutrition 2014; <b>100</b> :1257—68. doi:10.3945/ajcn.114.084921
17 18		
19 570	79	The World Bank. World Development Indicators: Reproductive health. 2016;:1-6.
20 21		
22 571		
23 24		
25		
26 27		
28		
29 30		
31		
32 33		
34		
35 36		
37		
38 39		
40		
41 42		
43		
44 45		
46		
47 48		
49		
50 51		
52		
53 54		
55		
56 57		
58		
59 60		
00		

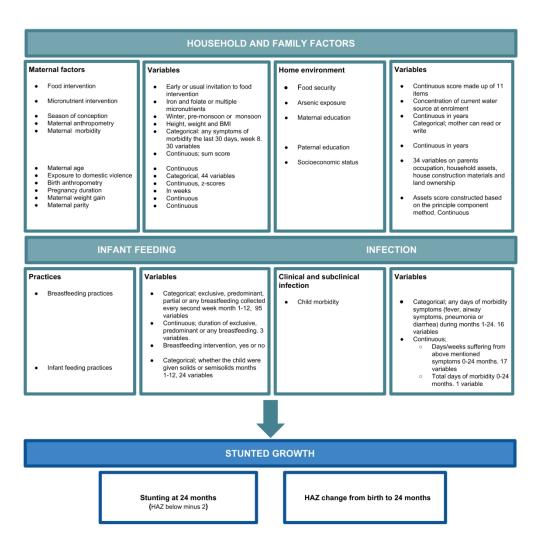


Figure 1. Factors, variables and outcomes included in the analysis of data from the MINIMat cohort, Bangladesh. Grouping according to the WHO conceptual framework on childhood stunting [20]

190x190mm (300 x 300 DPI)

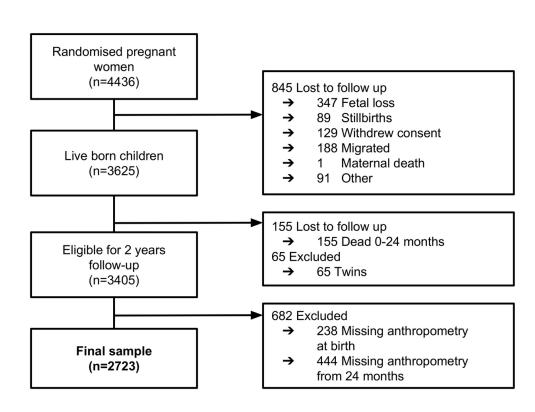
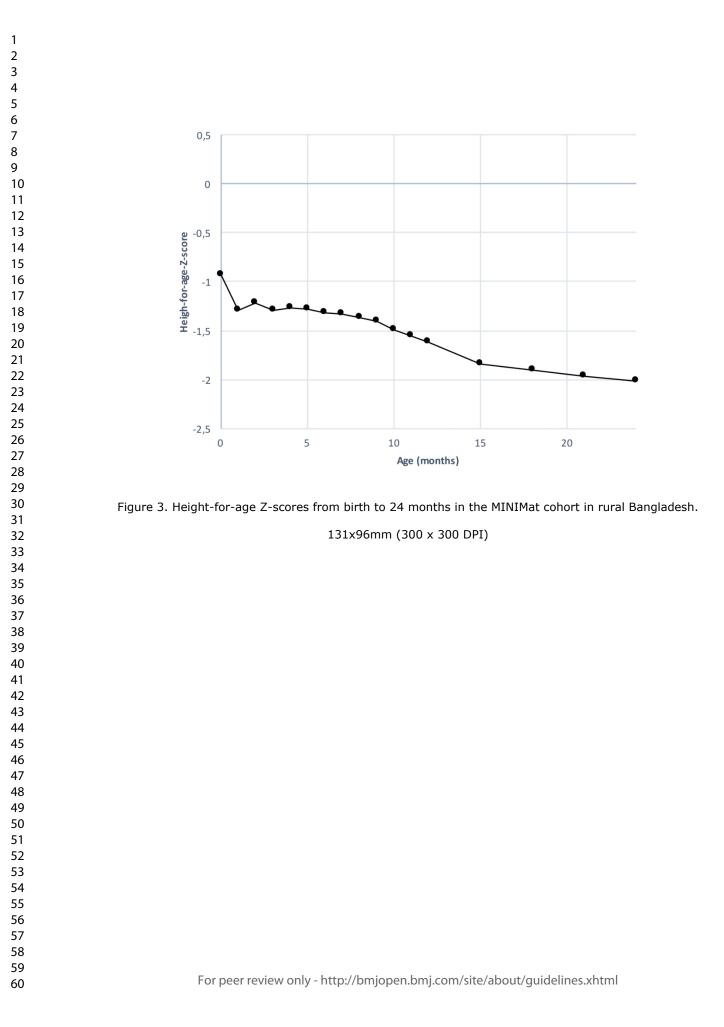


Figure 2. Flow chart of pregnant women and their children included in the data mining analyses of the MINIMat cohort from conception to two years of age.

106x80mm (300 x 300 DPI)

For peer review only - http://bmjopen.bmj.com/site/about/guidelines.xhtml

BMJ Open: first published as 10.1136/bmjopen-2018-025154 on 5 August 2019. Downloaded from http://bmjopen.bmj.com/ on June 12, 2025 at Agence Bibliographique de I Enseignement Superieur (ABES) . Protected by copyright, including for uses related to text and data mining, Al training, and similar technologies.



1 2			
3 4			
5 6			
7 8			
9	Height for age Z-score at birth Weight for age Z-score at birth	•	
10 11	Maternal height Small for gestational age Maternal weight at week 8 Asset score		
12 13	Maternal education in years Chest circumference at birth Father's education in years	•	
14 15	Number of saris mother owns Head circumference at birth Household owns tashak or mattress		
16	Mother is literate Household owns clock or watch Number of pairs of shoes mother owns		
17 18	Household owns a television Household owns almirah Maternal BMI at 8 weeks		
19 20	Number of shalwar kameez mother owns Household owns cows Household member works on daily basis Household owns lep or quilt		
21 22	Food security score Household owns chair or table Household owns an electric fan	Maternal factors	
23 24	Child received semisolids at 12 months Household has a stable source of income Sum of days with illness 0-24 months	Home environment     Infant feeding practices     Infection	
25	Number of days with chest indrawings 0-24m Number of days with rapid breathing 0-24 months		
26 27	0	.000 0.002 0.004 0.006 0.008 Relative importance	
28 29			
	bility to explain the presence of stur	t ranking the relative importance of 30 predictors with regard to the nting at 24 months of age. The MINIMat cohort in rural Bangladesh	
32 33	Colour cooling according to	o the WHO conceptual framework on causes of stunting.	
34		190x134mm (300 x 300 DPI)	
25			
35 36			
36 37			
36 37 38 39			
36 37 38 39 40 41			
36 37 38 39 40 41 42 43			
36 37 38 39 40 41 42			
36 37 38 39 40 41 42 43 44 45 46			
36 37 38 39 40 41 42 43 44 45 46 47 48			
36 37 38 39 40 41 42 43 44 45 46 47 48 49 50			
36 37 38 39 40 41 42 43 44 45 46 47 48 49 50 51 52			
36 37 38 39 40 41 42 43 44 45 46 47 48 49 50 51 52 53 54			
36 37 38 39 40 41 42 43 44 45 46 47 48 49 50 51 52 53 54 55			
36 37 38 39 40 41 42 43 44 45 46 47 48 49 50 51 52 53 54			

Page 38 of 46

BMJ Open: first published as 10.1136/bmjopen-2018-025154 on 5 August 2019. Downloaded from http://bmjopen.bmj.com/ on June 12, 2025 at Agence Bibliographique de I Enseignement Superieur (ABES) . Protected by copyright, including for uses related to text and data mining, Al training, and similar technologies.

1								
2								
3								
4								
5 6								
7								
8		[						
9	Height for age Z-score at birth Weight for age Z-score at birth			•			•	
10	Pregnancy duration in weeks Chest circumference at birth		•					
11 12	Head circumference at birth Preterm Maternal education in years							
13	Small for gestational age Parity							
14	Father's education in years Number of pairs of shoes mother owns	•						
15	Asset score Household owns cows							
16 17	Maternal age Number of saris mother owns Household owns a television	•						
18	Number of shalwar kamiz mother owns Mother is literate							
19	Household owns an electric fan Exclusive breastfeeding at age 4 months							
20	Maternal height Material of floor	•						
21 22	Father or brother-in-law in the same household Arsenic concentration in drinking water					l fe et		
22	Household owns clock or watch Exposure to lifetime physical violence	•				nvironment		
24	Partial breastfeeding at 4 months Household owns chair or table Child received semisolids at 10 months	•			<ul> <li>Infaction</li> </ul>	eeding practic n	es	
25	Domestic violence; ever been slapped	•						
26		0.00 0.01	0.02	0.03	0.04	0.05	0.06	
27 28			R	elative importanc	e			
20								
29								
Fig	ure 5. Conditional random forest plot							
30 Fig 31	ure 5. Conditional random forest plot ability to explain the variation in char cohort in rural Bangladesh. Colour c	nge in HAZ ( $\Delta$ H coding according	IAZ) fron g to the V	n birth to	24 mon	ths of ag	ge. The	MINIMat
30 Fig 31 32	ability to explain the variation in char	nge in HAZ (Δ H	IAZ) fron g to the V	n birth to	24 mon	ths of ag	ge. The	MINIMat
30 Fig 31	ability to explain the variation in char cohort in rural Bangladesh. Colour c	nge in HAZ ( $\Delta$ H coding according	IAZ) fron to the V ng.	n birth to NHO con	24 mon	ths of ag	ge. The	MINIMat
30 Fig 31 32 33 34 35	ability to explain the variation in char cohort in rural Bangladesh. Colour c	nge in HAZ (Δ H oding according stuntin	IAZ) fron to the V ng.	n birth to NHO con	24 mon	ths of ag	ge. The	MINIMat
30 Fig 31 32 33 34 35 36	ability to explain the variation in char cohort in rural Bangladesh. Colour c	nge in HAZ (Δ H oding according stuntin	IAZ) fron to the V ng.	n birth to NHO con	24 mon	ths of ag	ge. The	MINIMat
30 Fig 31 32 33 34 35 36 37	ability to explain the variation in char cohort in rural Bangladesh. Colour c	nge in HAZ (Δ H oding according stuntin	IAZ) fron to the V ng.	n birth to NHO con	24 mon	ths of ag	ge. The	MINIMat
30 Fig 31 32 33 34 35 36 37 38	ability to explain the variation in char cohort in rural Bangladesh. Colour c	nge in HAZ (Δ H oding according stuntin	IAZ) fron to the V ng.	n birth to NHO con	24 mon	ths of ag	ge. The	MINIMat
30 Fig 31 32 33 34 35 36 37 38 39 40	ability to explain the variation in char cohort in rural Bangladesh. Colour c	nge in HAZ (Δ H oding according stuntin	IAZ) fron to the V ng.	n birth to NHO con	24 mon	ths of ag	ge. The	MINIMat
30 Fig 31 32 33 34 35 36 37 38 39 40 41	ability to explain the variation in char cohort in rural Bangladesh. Colour c	nge in HAZ (Δ H oding according stuntin	IAZ) fron to the V ng.	n birth to NHO con	24 mon	ths of ag	ge. The	MINIMat
30 Fig 31 32 33 34 35 36 37 38 39 40 41 42	ability to explain the variation in char cohort in rural Bangladesh. Colour c	nge in HAZ (Δ H oding according stuntin	IAZ) fron to the V ng.	n birth to NHO con	24 mon	ths of ag	ge. The	MINIMat
30 Fig 31 32 33 34 35 36 37 38 39 40 41 42 43	ability to explain the variation in char cohort in rural Bangladesh. Colour c	nge in HAZ (Δ H oding according stuntin	IAZ) fron to the V ng.	n birth to NHO con	24 mon	ths of ag	ge. The	MINIMat
30     Fig.       31     32       33     34       35     36       37     38       39     40       41     42       43     44       45     45	ability to explain the variation in char cohort in rural Bangladesh. Colour c	nge in HAZ (Δ H oding according stuntin	IAZ) fron to the V ng.	n birth to NHO con	24 mon	ths of ag	ge. The	MINIMat
30     Fig.       31     32       33     34       35     36       37     38       39     40       41     42       43     44       45     46	ability to explain the variation in char cohort in rural Bangladesh. Colour c	nge in HAZ (Δ H oding according stuntin	IAZ) fron to the V ng.	n birth to NHO con	24 mon	ths of ag	ge. The	MINIMat
30     Fig.       31     32       33     34       35     36       37     38       39     40       41     42       43     44       45     46       47	ability to explain the variation in char cohort in rural Bangladesh. Colour c	nge in HAZ (Δ H oding according stuntin	IAZ) fron to the V ng.	n birth to NHO con	24 mon	ths of ag	ge. The	MINIMat
30     Fig       31     32       33     34       35     36       37     38       39     40       41     42       43     44       45     46       47     48	ability to explain the variation in char cohort in rural Bangladesh. Colour c	nge in HAZ (Δ H oding according stuntin	IAZ) fron to the V ng.	n birth to NHO con	24 mon	ths of ag	ge. The	MINIMat
30     Fig.       31     32       33     34       35     36       37     38       39     40       41     42       43     44       45     46       47     48       49     50	ability to explain the variation in char cohort in rural Bangladesh. Colour c	nge in HAZ (Δ H oding according stuntin	IAZ) fron to the V ng.	n birth to NHO con	24 mon	ths of ag	ge. The	MINIMat
30     Fig.       31     32       33     34       35     36       37     38       39     40       41     42       43     44       45     46       47     48       49     50       51     51	ability to explain the variation in char cohort in rural Bangladesh. Colour c	nge in HAZ (Δ H oding according stuntin	IAZ) fron to the V ng.	n birth to NHO con	24 mon	ths of ag	ge. The	MINIMat
30     Fig.       31     32       33     34       35     36       37     38       39     40       41     42       43     44       45     46       47     48       49     50       51     52	ability to explain the variation in char cohort in rural Bangladesh. Colour c	nge in HAZ (Δ H oding according stuntin	IAZ) fron to the V ng.	n birth to NHO con	24 mon	ths of ag	ge. The	MINIMat
30     Fig.       31     32       33     34       35     36       37     38       39     40       41     42       43     44       45     46       47     48       49     50       51     52       53	ability to explain the variation in char cohort in rural Bangladesh. Colour c	nge in HAZ (Δ H oding according stuntin	IAZ) fron to the V ng.	n birth to NHO con	24 mon	ths of ag	ge. The	MINIMat
30     Fig.       31     32       33     34       35     36       37     38       39     40       41     42       43     44       45     46       47     48       49     50       51     52	ability to explain the variation in char cohort in rural Bangladesh. Colour c	nge in HAZ (Δ H oding according stuntin	IAZ) fron to the V ng.	n birth to NHO con	24 mon	ths of ag	ge. The	MINIMat
30       Fig.         31       32         33       34         35       36         37       38         39       40         41       42         43       44         45       46         47       48         49       50         51       52         53       54         55       56	ability to explain the variation in char cohort in rural Bangladesh. Colour c	nge in HAZ (Δ H oding according stuntin	IAZ) fron to the V ng.	n birth to NHO con	24 mon	ths of ag	ge. The	MINIMat
30       Fig.         31       32         33       34         35       36         37       38         39       40         41       42         43       44         45       46         47       48         49       50         51       52         53       54         55       56         57	ability to explain the variation in char cohort in rural Bangladesh. Colour c	nge in HAZ (Δ H oding according stuntin	IAZ) fron to the V ng.	n birth to NHO con	24 mon	ths of ag	ge. The	MINIMat
30       Fig.         31       32         33       34         35       36         37       38         39       40         41       42         43       44         45       46         47       48         49       50         51       52         53       54         55       56         57       58	ability to explain the variation in char cohort in rural Bangladesh. Colour c	nge in HAZ (Δ H oding according stuntin	IAZ) fron to the V ng.	n birth to NHO con	24 mon	ths of ag	ge. The	MINIMat
30       Fig.         31       32         33       34         35       36         37       38         39       40         41       42         43       44         45       46         47       48         49       50         51       52         53       54         55       56         57	ability to explain the variation in char cohort in rural Bangladesh. Colour c	nge in HAZ (Δ H coding according stunti 190x134mm (30	IAZ) fron g to the V ng. D0 x 300	n birth tc WHO con DPI)	o 24 mon ceptual f	iths of a <u>c</u>	ge. The rk on ca	MINIMat
30       Fig.         31       32         33       34         35       36         37       38         39       40         41       42         43       44         45       46         47       48         49       50         51       52         53       54         55       56         57       58         59       59	ability to explain the variation in char cohort in rural Bangladesh. Colour c 1	nge in HAZ (Δ H coding according stunti 190x134mm (30	IAZ) fron g to the V ng. D0 x 300	n birth tc WHO con DPI)	o 24 mon ceptual f	iths of a <u>c</u>	ge. The rk on ca	MINIMat

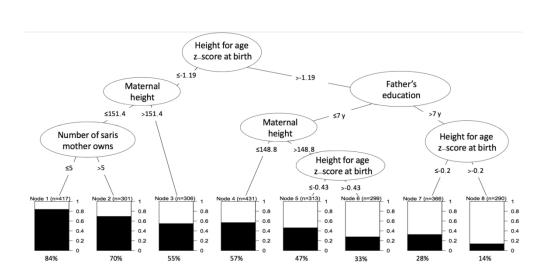
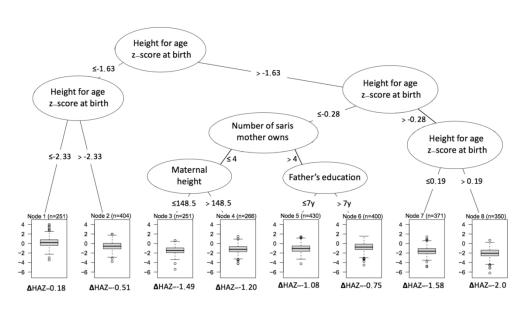


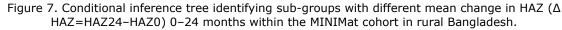
Figure 6. Conditional inference tree identifying sub-groups with different probabilities of stunting at 24 months. The MINIMat cohort in rural Bangladesh.

190x88mm (300 x 300 DPI)

BMJ Open: first published as 10.1136/bmjopen-2018-025154 on 5 August 2019. Downloaded from http://bmjopen.bmj.com/ on June 12, 2025 at Agence Bibliographique de l Enseignement Superieur (ABES) . Protected by copyright, including for uses related to text and data mining, Al training, and similar technologies.







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 ( $\alpha$ ) and deleted a proportion ( $\beta$ ) 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].

Enseignement Superieur (ABES) . Protected by copyright, including for uses related to text and data mining, Al training, and similar technologies.

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

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

<sup>1</sup> K-nearest neighbour <sup>2</sup>Missclassification rate

<sup>3</sup>Mean square error

weatt square error

 $\alpha$  = proportion of the non-missing values deleted

 $\beta$  = proportion of the original data sampled

Enseignement Superieur (ABES) . Protected by copyright, including for uses related to text and data mining, Al training, and similar technologies.

י ר	
2	
3	
4	
5	
6	
7	
8	
9	
10	
11	
12	
12	
14	
15 16	
16	
17 18	
18	
19	
20	
20	
22	
22	
23	
24	
25 26	
26	
27	
28	
29	
30	
31	
31 32	
32	
33	
34	
35 36	
36	
37	
37 38	
39	
40	
41	
41	
43	
44	
45	
46	
47	
48	
49	
50	
51	
52	
53	
53 54	
55	
56	
57	
58	
59	

60

1

<b>Table 2:</b> Means and Standard errors of the MR <sup>2</sup> and the MSE <sup>3</sup> for different	
imputation methods, computed from m=100 samples, $lpha=0.05, eta=0.15$	

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

<sup>1</sup> K-nearest neighbour

<sup>2</sup>Missclassification rate

<sup>3</sup>Mean square error

 $\alpha$  = proportion of the non-missing values deleted

 $\beta$  = proportion of the original data sampled

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

Variable mean	1	
variable mean	KNN⁺	Random forest
0.1625007370	0.1608280983	0.094319580
0.0005210379	0.0005181798	0.000367369
1.0023969039	0.7975006166	0.450253626
0.0068209597	0.0066997794	0.006069386
	0.1625007370 0.0005210379 1.0023969039	0.16250073700.16082809830.00052103790.00051817981.00239690390.7975006166

<sup>1</sup> K-nearest neighbour

<sup>2</sup>Missclassification rate

<sup>3</sup>Mean square error

- $\alpha$  = proportion of the non-missing values deleted
- $\beta$  = 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 <sup>1</sup>	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

<sup>1</sup> K-nearest neighbour

<sup>2</sup>Missclassification rate

<sup>3</sup>Mean square error

 $\alpha$  = proportion of the non-missing values deleted

 $\beta$  = 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.
- 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.

2	
2	
3	
4	
5	
6	
7	
8	
9	
10	
11	
11	
12	
13	
14	
15	
16	
17 18	
10	
10	
19	
20	
21	
22	
23	
24	
25	
25	
26	
27	
28	
29	
30	
31	
32	
33	
34	
35	
36	
37	
38	
39	
40	
41	
42	
43	
44	
45	
46	
47	
48	
49	
50	
51	
52	
53	
54	
55	
56	
57	
58	
59	
60	
00	

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	1
		title or the abstract	
		(b) Provide in the abstract an informative and balanced summary	2
		of what was done and what was found	
Introduction			
Background/rationale	2	Explain the scientific background and rationale for the	4-5
-		investigation being reported	
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	6
Setting		periods of recruitment, exposure, follow-up, and data collection	0
Participants	6	(a) Cohort study—Give the eligibility criteria, and the sources	6
i unicipanto	Ű	and methods of selection of participants. Describe methods of	0
		follow-up	
		<i>Case-control study</i> —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	
		<i>Cross-sectional study</i> —Give the eligibility criteria, and the	
		sources and methods of selection of participants	
		(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	9 Figure
		confounders, and effect modifiers. Give diagnostic criteria, if	1
		applicable	
Data sources/ measurement	8*	For each variable of interest, give sources of data and details of	7-9
		methods of assessment (measurement). Describe comparability	
		of assessment methods if there is more than one group	
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.	7-9
		If applicable, describe which groupings were chosen and why	
Statistical methods	12	(a) Describe all statistical methods, including those used to	10-11
		control for confounding	
		(b) Describe any methods used to examine subgroups and	10-11
		interactions	
		(c) Explain how missing data were addressed	10-11
		(d) Cohort study—If applicable, explain how loss to follow-up	10
		(a) Conort study—It applicable, explain now loss to follow-up was addressed	10
		<i>Case-control study</i> —If applicable, explain how matching of	
		Cuse-control study—II applicable, explain now matching of	
		cases and controls was addressed	

	<i>Cross-sectional study</i> —If applicable, describe analyt methods taking account of sampling strategy	
Continued on next page10	( $\underline{e}$ ) Describe any sensitivity analyses	10

BMJ Open: first published as 10.1136/bmjopen-2018-025154 on 5 August 2019. Downloaded from http://bmjopen.bmj.com/ on June 12, 2025 at Agence Bibliographique de l Enseignement Superieur (ABES)

Protected by copyright, including for uses related to text and data mining, Al training, and similar technologies

2	
3	
4	
5	
6	
7	
8	
9	
10	
11	
12	
13	
14	
15	
16	
17	
18	
19	
20	
21	
22	
23	
24	
25	
26	
27	
28	
20	
29 30	
31	
32	
33	
34	
35	
36	
37	
38	
39	
40	
41	
42	
43	
44	
45	
46	
40 47	
47 48	
49 50	
50	
51	
52	
53	
54	
55	
56	
57	
58	
59	
60	
50	

1 2

Participants	13*	(a) Report numbers of individuals at each stage of study—eg numbers potentially	Fig 2
1 un no spunds	10	eligible, examined for eligibility, confirmed eligible, included in the study,	12
		completing follow-up, and analysed	
		(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	Table 1,
-		information on exposures and potential confounders	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	13
		time	
		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	14
		and their precision (eg, 95% confidence interval). Make clear which confounders	
		were adjusted for and why they were included	
		(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	14
		sensitivity analyses	
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	15,16,17
		imprecision. Discuss both direction and magnitude of any potential bias	
Interpretation	20	Give a cautious overall interpretation of results considering objectives, limitations,	18-22
		multiplicity of analyses, results from similar studies, and other relevant evidence	
Generalisability	21	Discuss the generalisability (external validity) of the study results	17
Other information	on		
Funding	22	Give the source of funding and the role of the funders for the present study and, if	23
		applicable, for the original study on which the present article is based	

\*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
<b>Primary Subject Heading</b> :	Global health
Secondary Subject Heading:	Paediatrics, Nutrition and metabolism
Keywords:	Nutrition < TROPICAL MEDICINE, PUBLIC HEALTH, EPIDEMIOLOGY, PAEDIATRICS



2 3 4 5	1	Relative importance of pre- and postnatal determinants of stunting: data mining
6 7 8 9 10 11	2	approaches to the MINIMat cohort, Bangladesh
12 13 14 15	3	
16 17 18	4	Pernilla Svefors1°, Oleg Sysoev², Eva-Charlotte Ekström¹, Lars-Åke Persson³, Shams El Arifeenª, Ruchira Navedª, Anisur Rahmanª,
19 20 21 22	5	Ashraful Islam Khan <sup>4</sup> , Katarina Ekholm Selling <sup>1</sup>
23 24 25 26	6	'International Maternal and Child Health, Department of Women's and Children's Health, Uppsala University, Uppsala, Sweden. ²
27 28 29	7	Department of Computer and Information Sciences, Linköping University, Linköping, Sweden, <sup>3</sup> Department of Disease Control,
30 31 32 33	8	London School of Hygiene and Tropical Medicine, London, UK, 4 International Centre for Diarrhoeal Disease Research, Bangladesh
34 35 36 37	9	(iccdr,b), Dhaka, Bangladesh.
<ul> <li>38</li> <li>39</li> <li>40</li> <li>41</li> <li>42</li> <li>43</li> <li>44</li> <li>45</li> <li>46</li> <li>47</li> <li>48</li> <li>49</li> <li>50</li> <li>51</li> <li>52</li> <li>53</li> <li>54</li> <li>55</li> <li>56</li> <li>57</li> <li>58</li> <li>59</li> <li>60</li> </ul>	10	*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 o-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. 

1 2 3 4 5	29	Strengths :	and limitations of this study
6 7 8	30	• 1	includes high-quality longitudinal data with low rates of missing data on child growth and a wide range of pre and
9 10 11 12	31	ł	postnatal household, family, and environmental factors, child characteristics at birth, infant feeding, and morbidity.
13 14 15 16	32	• E	Employs decision-tree-based methods that permit the inclusion of a high number of predictor variables, variables of
17 18 19	33	C	different types and automatically discover complex interactions between predictor variables and include them in the
20 21 22 23	34	Y	nodel.
24 25 26	35	• 5	Some potentially important determinants of linear growth were not present in the database.
27 28 29 30	36	• 1	The study does not include stratified analyses for girls and boys
31 32 33 34			
35 36 37			
38 39 40 41			
42 43 44			
45 46 47 48			
48 49 50 51			
52 53 54			
55 56 57 58			
59 60			

#### 4 2: 2: 3: 3: 50 50 50 50 50 50 5

### Introduction

	39	Linear growth is considered to be the best overall indicator of children's present and future health[1,2] and the reduction of growth
0	40	failure is one of the targets within the sustainable development agenda. Stunted growth is associated with short-term morbidity
2 3 4	41	and mortality, impaired cognitive development, lower future productivity, and increased risk of adult chronic diseases [3]. In 2012,
5 6 7	42	the WHO adopted a resolution on maternal and child undernutrition, targeting a reduction of stunting by 40% by 2025 [4]. Linear
8 9 0	43	growth is most susceptible to environmentally modifiable factors from conception up to two years of age, i.e., the first 1000 days
2 3 4	44	when most of the growth faltering takes place [5, 6]. To develop and deliver appropriate interventions, it is imperative to establish
5 6 7	45	the relative importance of stunting risk factors. In addition, the sustainable development health goal has emphasized the
8 9 0 1	46	personalized perspective under the universal coverage of health care. Precision public health interventions by identifying and
2 3 4	47	targeting high-risk subgroups can be one of the strategies to reach this goal[7] .
5 6 7	48	Previous studies employing classical statistical methods have identified a wide range of pre- and post-natal factors
8 9	10	r revious studies employing classical statistical methous have mentyled a white range of pre- and post-natal factors
0 1 2	49	associated with impaired growth [8-13]. Low birth weight, maternal height, maternal education, poverty and inadequate
3 4 5	50	complementary feeding practices have been recognized as important risk factors [14-16]. Some analyses emphasize the importance
6 7 8 9	51	of fetal growth restriction for later stunted growth, but rarely is the relative importance of pre- and post-natal factors assessed [17].
9 0 1 2	52	Despite these findings, policy documents and recommendations emphasize interventions especially after birth, and pre-natal
3 4 5	53	recommendations are usually limited to routine micronutrient supplementation for pregnant women [18-20].
6 7 8		
9 0		

1 2 3 4	54	Despite a wealth of literature relating to the determinants of stunting, studies with a holistic approach, which
5 6 7	55	concurrently account for household, environmental, nutritional, biological, and socioeconomic influences are few. Moreover,
8 9 10 11	56	individuals and groups may be stunted for various reasons and thus respond differently to interventions. Studies that identify risk
12 13 14	57	groups with different probabilities of stunting are, to the best of our knowledge, not yet available. The available studies with a
15 16 17	58	multifactorial approach have frequently had a cross-sectional design and have applied traditional statistical methods. As visualized
18 19 20 21	59	in the WHO's conceptual framework on childhood stunting [21], the causes of stunted linear growth are complex. The number of
22 23 24	60	risk factors and the complexity of the associations of these risk factors with linear growth restriction make traditional statistical
25 26 27	61	models ineffective from a predictive perspective. Moreover, classical statistical methods do not have the capacity to identify groups
28 29 30 31	62	with different risks based on combinations of predictors. Decision trees are popular data mining (DM) methods, which allows for
32 33 34	63	the inclusion of a high number of predictor variables, handling variables of different types, automatically discovering complex
35 36 37	64	interactions between predictor variables and including them in the model [22]. Decision-tree-based algorithms can be used to rank
38 39 40 41	65	a high number of predictors according to their relative importance for the outcome and to identify subgroups with different risk
41 42 43 44 45 46 47 48 49	66	patterns.
	67	The Maternal and Infant Nutrition Interventions in Matlab (MINIMat) was a randomized prenatal food and multiple
50 51 52	68	micronutrient trial carried out in rural Bangladesh. The frequent follow-up of mothers and children participating in this trial
53 54 55	69	resulted in an extensive database, including frequent pre- and post-natal anthropometric assessments, socioeconomic and
56 57 58 59	70	biological characteristics of the mother and father, information on maternal exposure to violence, household food security, breast-
60	71	and infant-feeding practices, and measurement of morbidity of the mothers during pregnancy and repeatedly of children up to 24

2 3 4	72	months of age. The aim of this study is to, within this Bangladeshi cohort, assess the relative importance of determinants of linear
5 6 7	73	growth from 0–24 months and risk factors for stunting at two years, and to identify risk groups with negative growth trajectories
8 9 10 11 12	74	and high prevalence of stunting at two years.
13 14 15	75	Methods
16 17 18	76	Study setting, participants and study design
19 20 21	77	The MINIMat trial (Maternal and Infant Nutrition Interventions in Matlab, isrctn.org identifier: ISRCTN16581394) was carried out
22 23 24 25	78	in Matlab, Bangladesh, a rural delta region located 57 km southeast of the capital Dhaka. In this area, a health and demographic
26 27 28	79	surveillance system enables early pregnancy identification and longitudinal follow-up. Pregnant women were enrolled in the
29 30 31 32	80	MINIMat trial and the follow-up included their offspring. MINIMat was a factorial randomized trial primarily evaluating the effect
33 34 35	81	of an early invitation to prenatal food supplementation (versus usual timing) combined with multiple micronutrient
36 37 38	82	supplementation (versus usual program iron-folate) to pregnant women on maternal hemoglobin, birth weight, gestational age at
39 40 41 42	83	birth, and infant mortality [23]. Further, the participating women were randomly assigned to either counselling for exclusive
43 44 45	84	breastfeeding or a different health education message of equivalent intensity [24]. The MINIMat trial recruited pregnant women
46 47 48 49	85	from November 2001 to October 2003. When a woman reported to a community health worker that her menstruation was delayed
49 50 51 52	86	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
53 54 55	87	was missing, the gestational age assessment was based on ultrasound examination. In total, 4436 pregnant women participated,
56 57 58 59	88	giving birth to 3625 live born infants from April 2002 to June 2004. The pregnant women were enrolled at around gestational week
60	89	8. In this analysis, the mothers and children were followed through pregnancy, birth, and up to two years of age.
		(

6

1 2 3 4	90	Written and oral informed consent was obtained from all participating women and from the parents of the participating
5 6 7 8	91	children. The Ethical Review Committee at the International Centre for Diarrhoeal Disease Research, Bangladesh, approved the
9 10 11	92	study (approval registration numbers 2000-025; 2002-031; 2005-004)
12 13 14 15 16	93	
17 18 19	94	Data collection
20 21 22 23	95	Predictor and outcome variables are presented in Figure 1, grouped according to the WHO conceptual framework of stunting [21].
24 25 26	96	Data were collected using questionnaires, physical examinations, and laboratory analyses. At enrolment, well-trained field workers
27 28 29 30	97	collected information on women's age, parity, marital status, educational level, occupation, maternal morbidity, socioeconomic
31 32 33	98	characteristics, and household food security. Socioeconomic status was assessed based on a range of household assets, and a
34 35 36	99	continuous household asset score, with a mean value of zero, was constructed based on a principal component analysis [25]. A
37 38 39 40	100	validated household food security scale was created from eleven items with data on frequency of food purchased, cooked, borrowed
41 42 43		or lent (food and money), and whether there was ready access to adequate meals and snacks [26]. The participating women were
44 45 46 47	102	also asked whether they had suffered any of thirty morbidity symptoms from twelve different categories, including airway, urinary
48 49 50		tract, fever, circulation, bowel, or pain symptoms during the last month. A sum score ranging from zero to twelve was created based
51 52 53 54	104	on absence of symptoms or those not recorded for each category.
55 56 57	105	Home visits were followed by clinic visits at local health sub-centers. Maternal height and weight were measured at
58 59 60	106	around eight weeks of gestation using a stadiometer to the nearest 0.1 cm and an electronic scale (Uniscale; SECA) with a precision

Enseignement Superieur (ABES) . Protected by copyright, including for uses related to text and data mining, Al training, and similar technologies.

1	
<sup>2</sup> 107 4	of 0.10 kg. In the third trimester, paramedics interviewed the participating women in privacy regarding their experiences of
5 6 108 7	domestic violence. A modified version of the WHO collaborative study questionnaire was used [27,28], based on the conflict tactic
8 9 109 10 11	scale covering physical, sexual and emotional violence and controlling behavior [29]. Household drinking water was analyzed for
12 13 110 14	arsenic concentration [30].
15 16 17 111 18 19	A birth notification system allowed birth anthropometry to be measured within 72 hours. In the few cases where the
<sup>20</sup> 112 21 22	newborns were reached after 72 hours, the measurements were adjusted to the time of birth using an SD score transformation,
23 24 113 25	assuming that the infants remained in the same relative position in the anthropometric distribution during this period [31]. At
26 27 114 28 29	birth, data on sex, birth weight, length, and breastfeeding practices were collected. During the subsequent two-year study period,
<sup>30</sup> <sub>31</sub> 115 <sub>32</sub>	the mother-and-child pairs were visited monthly in their homes during the first year, and every three months during the second
33 34 116 35	year. On these occasions, data on infant feeding practices, child morbidity and anthropometry were collected. The mothers were
36 <sup>37</sup> 117 38 39	interviewed about breastfeeding and complementary feeding practices. Breastfeeding practices were categorized into exclusive,
40 41 118 42	predominant, partial, or any breastfeeding for each month from one to twelve months. The total time for exclusive, predominant,
43 44 119 45 46	and any breastfeeding was calculated. The WHO recommendations guided the breastfeeding assessment [32] and results were
47 48 49	validated with a stable-isotope technique. The classification of exclusive breastfeeding was found to suffer from limited
50 51 121 52	misclassification in both directions and to be accurate at the group level [33]. The food given to the infant was categorized into
53 54 122 55 56	semi-solids and solids each month from one to twelve months. The data collection did not include full dietary assessments or
57 58 123 59 60	classification of dietary diversity and meal frequency.

1 2 3 4	The mothers were also asked whether the child had had any of the following symptoms during the last week; fever, cough,
5 6 125 7 8	difficult breathing, chest in-drawing, rapid breathing, diarrhea, bloody diarrhea and the duration of these symptoms [34].
9 126 10 11	Categories were created based on whether the child had suffered from fever, respiratory symptoms, suspected pneumonia, or
12 13 127 14	diarrhea, and the sum of days with each symptom and total morbidity calculated from birth to 24 months. To reduce the risk of
15 16 128 17 18	recall bias the mothers were visited monthly with an interview recall period of seven days for child morbidity. One week has been
19 20 21 22	found to be optimal for this kind of morbidity recall assessment [35].
23 24 130 25 26	Children's weight was measured by SECA beam and electronic scales (UNICEF Uniscale; SECA Gmbh & Co, Hamburg,
27 131 28 29	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
30 31 32	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
33 34 133 35 36	stadiometer. Head and chest circumference were measured with a measuring tape. Two measurements were recorded on each
<sup>37</sup> 134 38 39	occasion and the mean was calculated. The equipment was calibrated daily and refresher training on data collection methods,
40 41 42 43 44	including the standardization of anthropometric measurements, was conducted periodically.
45 46 47	Outcomes
48 137 49 50	Height-for-age z-scores (HAZ) were calculated from the measured length and height data using the program WHOAnthro, based
51 52 53	on the WHO growth reference for children [36]. Children with a HAZ below minus two SD-scores were classified as stunted. Two
54 55 139 56 57	outcomes were analyzed: stunting at 24 months and the change in HAZ from birth to 24 months, referred to as $\Delta$ HAZ and
<sup>58</sup> 140 59 60	calculated by subtracting HAZ at birth from HAZ at 24 months i.e. $\Delta$ HAZ = HAZ at 24 months - HAZ at birth.

Enseignement Superieur (ABES) . Protected by copyright, including for uses related to text and data mining, Al training, and similar technologies.

## Statistical analysis

5 6 7	142	A database was created with 309 variables characterizing mothers and children in the MINIMat cohort from enrolment in early
10	143	pregnancy up to the time when the children were 24 months of age. The sub-set of records that had height measurements at birth
11 12 13 14	144	and 24 months was selected ( <i>n</i> =2 723). The average percent of missing values among all the predictors were 4 %. The highest
15	145	percent missing were among maternal morbidity data during pregnancy (22%) and categorical monthly child morbidity data (ill
20	146	or not), ranging from 0% to 35% with the highest number of missing observations in the first months. The continuous child
21 22 23 24	147	morbidity data however (sum of days with different types illnesses), had no missing values. The most important variables identified
25	148	by the random forest analyses and the variables included by the conditional inference trees had less than 1% missing values. The
30	149	missing values of the predictor variables were imputed. To find the best method to impute the missing data we made a simulation
31 32 33 34	150	study of the performance of the following imputation methods: imputation by variable mean, K-nearest neighbor imputation [37],
35	151	and random forest imputation [38]. The design of the study followed a procedure similar to the strategy described in Jonsson et al.
40	152	[37], see S appendix. Accordingly, we imputed the data by use of the random forest as the simulation study revealed that this
41 42 43 44	153	method provided the most accurate imputations.
45 46 47 48	154	Decision trees [22] are data mining methods that allow for specifying an arbitrarily high number of predictor variables,
	155	handle variables of different types, automatically discover complex interactions between predictor variables, and include them in
52 53 54 55	156	the model. Traditional decision trees, such as Classification and Regression Trees (CART) have been shown to be biased [39]. This
56 57 58	157	motivated us to select the Conditional Inference Trees (CIT) framework, a method that embeds a statistical hypothesis-testing
59 60	158	framework into a recursive partitioning algorithm used for model building [39]. Conditional inference trees were used in order to
		10 For peer review only - http://bmjopen.bmj.com/site/about/guidelines.xhtml

1 2 3 159	identify sub-groups characterized by combinations of levels of certain predictors with distinct values of $m \Delta$ HAZ or prevalence of
4 5 6 160 7	stunting at 24 months. Cross-validation, a well-established model selection method that selects a tree with an optimal predictive
8 9 161 10	performance for new unseen data, was applied. Cross-validation splits the data set into different train and test sets repeatedly,
11 12 13 162 14	estimates the model in one set and validates the prediction on another set, followed by an aggregation of the predictions[40]. To
15 16 163 17 18	ensure public health relevance, the minimum number of observations in each terminal node (subgroup) was set to 250.
19 20 21 22	Conditional random forest (CRF) analyses were performed to assess and rank the importance of predictors with regard to
23 24 165 25	their ability to explain the variation of the continuous outcome of the change in HAZ from birth to 24 months and the presence of
26 27 166 28 29	stunting at 24 months of age. In conditional random forest analysis, an ensemble of conditional inference trees is created by means
<sup>30</sup> <sub>31</sub> 167 <sub>32</sub>	of drawing subsamples from the original data and fitting a unique randomized conditional inference tree to each sample. Possible
33 34 168 35	predictors at each split are selected randomly from the complete set of predictors, which leads to a better predictive performance of
36 37 38 39	the tree ensemble [40]. The importance of a variable is computed by comparing the predictive mean squared error (MSE) from the
40 41 42	original data and a dataset where the corresponding variable values are specified incorrectly, which makes the variable irrelevant
43 44 171 45 46	for the prediction. If the variable does not contribute to the prediction, the MSE is expected to be small when the values of the
47 48 49	variable are permutated. An aggregated difference between the MSE values over the given ensemble of trees makes up the relative
50 51 173 52	importance of a variable. The random forests analyses were created based on 3000 trees, and the 30 variables with the highest
53 54 55 56	importance measure are presented. The programming language R version 3.2.4 [41] and the 'party' package [42] were used for all
57 58 175 59 60	analyses.

1	
2 3 176 4	Patient and public involvement
5 6 177 7	No participants were involved in developing the hypothesis, the specific aims or the research questions, nor were they involved in
8 9 178 10 11	developing plans for design or implementation of the study. No participants were involved in the interpretation of study results or
<sup>12</sup> 179 13 14 15 16	write up of the manuscript. There are no plans to disseminate the results of the research to study participants.
17 180 18 19	Results
20 181 21 22	There were 4436 women enrolled into the MINIMat trial, of whom 845 were lost to follow-up before delivery, mainly due to fetal
<sup>23</sup> 182 24 25 26	death, outmigration, or because they withdrew their consent. Of the 3625 live-born children, including twins and triplets, 155 died
20 27 183 28 29	between birth and two years and 682 were excluded because of missing anthropometry, at birth or at two years, resulting in 2723
30 184 31 32	children available for analysis (Figure 2). In the non-analyzed group, there was a slightly higher percentage of mothers with more
<sup>33</sup> 34 185 35 36	than five years of education, younger than 20 years, and belonging to the lowest socioeconomic tertile, and preterm births of
37 186 38 39 40	children.
41 187 42 43	The characteristics of the households, mothers, fathers at eight weeks of gestation, and children at birth are given in Table
44 45 188 46 47	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
48 189 49 50	(SD 6·8) at recruitment. One-third of the women were underweight, with a BMI below 18·5 at pregnancy week eight. The average
51 190 52 190 53 54	number of years of education was similar for mothers and fathers (5 years). The sample of children comprised an equal proportion
55 191 56 57 58 59 60	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

1 2 3 192 4 5	low (mean	= -0·94), and declined further at up to two years	of age with a mean change	of -1 HAZ, resulting in a me	an HAZ at two
6 193 7 8 9	years of -2∙	0 (Figure 3) and 50% being stunted (girls 51-1%,	boys 48·5%)		
10 194 11 12 13					
14 15 15 16 17	<b>Table 1.</b> Bas	seline characteristics, prevalence of stunting at 22	4 months, and mean $\Delta$ HAZ	Z (change in height-for-age Z	Z-score) 0—24
18 18 19	months in	the MINIMat cohort, Bangladesh.			
20					
21 22 23 197 24		Characteristics	n/n (%)	Stunted at 24 months <i>n/n</i> (%)	Mean Δ HAZ 0-24 months
25		Mother's age (years)			
26		<20	395/2723 (14.5)	199/395 (50.4)	-0.74
27 198		20–29	1556/2723 (57.1)	753/1556 (48.4)	-1.05
28		>30	772/2723 (28.4)	417/772 (54.0)	-1.28
29		Mother's education			1.07
30 31 100		No education	913/2723 (33.5)	556/913 (60.9)	-1.27
<sup>31</sup> 199 32		Enrolled in primary school (1-5y)	624/2723 (22.9)	364/624 (58.3)	-1.24
33		Completed primary school (>5y)	1186/2723 (43.6)	449/1186 (37.9)	-0.83
34		Father's education	007/0702 (04.0)	E00/007 (04 4)	1.00
<sup>35</sup> 36 200		No education Enrolled in primary school (1-5y)	867/2723 (31.8)	532/867 (61.4)	-1.29 -1.12
36 <sup>200</sup>		Completed primary school (>5y)	670/2723 (24.6) 1186/2723 (43.6)	<u>369/670 (55.1)</u> 468/1186 (39.5)	-0.89
37 201		Parity	1100/21/20 (43.0)	0.80) 0011 00 <del>7</del>	-0.03
38		First child	791/2723 (29.0)	348/791 (44.0)	-0.76
39 202		Second child	774/2723 (28.4)	385/774 (49.7)	-1.09
<sup>40</sup> 41 203		Third or more child	1158/2723 (42.5)	636/1158 (54.9)	-1.28
		Number of saris mother owns			
<sup>42</sup> 204 43		<5	1078/2723 (39.6)	665/1078 (61.5)	-1.26
44 205		5–8	865/2723 (31.8)	427/865 (49.4)	-1.03
45		>8	780/2723 (28.6)	277/780 (35.5)	-0.87
46 206		Child at birth			ļ
47 48207		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
49 208		Low Birth Weight (LBW)	797/2723 (29.3)	546/797 (68.5)	-0.56
50 51 209		Normal birth weight	1926/2723 (70.7)	823/1926 (42.7)	-1.29
51 209 52		Preterm (<37 weeks of gestation)	190/2723 (7.0)	117/190 (61.6)	0.02
52 53		Term	2533/2723 (93)	1252/2533 (49.4)	-1.15

Enseignement Superieur (ABES) . Protected by copyright, including for uses related to text and data mining, Al training, and similar technologies.

1	
2 3 210 4	Relative importance of predictors for stunting at 24 months and change in height scores from
5 211 6	birth to 24 months
8 212 9 10	The relative importance of predictors with respect to their ability to explain the probability of stunting at 24 months and the
11 213 12 13	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
14 15 214 16	most important predictors of stunting at 24 months, followed by maternal height, Small for Gestational Age (SGA), maternal
17 18 215 19 20	weight at eight weeks of gestation, household asset score, and parental education. The most important factors for $\Delta$ HAZ were HAZ
<sup>21</sup> 216 22 23 24	and WAZ at birth, pregnancy duration, head and chest circumference at birth, and maternal education.
25 26 217 27	Subgroups with different levels of stunting at 24 months and levels of change in height scores from
<sup>28</sup> <sub>29</sub> 218 30	birth to 24 months
31 219 32 33	The conditional inference trees presented in Figure 6 and 7 display subgroups with different probability of stunting at 24 months
<sup>34</sup> 220 35 36	and levels of $\Delta$ HAZ 0-24 months due to distinctive combinations of levels of certain predictors. The conditional inference trees for
37 38 221 39 40	stunting and $\Delta$ HAZ were composed of subgroups defined by the same predictors, specifically; HAZ at birth, maternal height,
41 222 42 43	father's educational level, and the number of saris owned by the mother. The probability of stunting ranged from 14% to 84%.
44 45 223 46 47	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
48 224 49 50	highest probability of stunting at 24 months, at 84%. Children of a father with more than seven years of education, who had HAZ
51 52 53 54	at birth above -0·2, had the lowest probability of stunting at 24 months, at 14% (Figure 6). The difference in <b><math>\Delta</math></b> HAZ between the
54 55 226 56 57 58 59 60	identified subgroups of children with the most negative change and the subgroup with the most positive change was 2·22 HAZ.

1	
<sup>2</sup> <sub>3</sub> 227 4	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
5 6 228 7 8	HAZ), while children who were born with a HAZ above 0.19 had the most negative <b><math>\Delta</math></b> HAZ (-2·04 HAZ) (Figure 7).
9	
10 <u>229</u> 11	
12	
13 14 15 230	
16 17	
18 19 20 231	Discussion
21 22 232	In our analysis of 309 predictors characterizing household, environmental, biological, and socioeconomic factors, we found birth
23	in our analysis of 309 preactors characterizing nouschold, churonmental, biological, and socioeconomic factors, we found birth
24 25 aaa	
<sup>25</sup> 233 26 27	size, maternal anthropometry and parental education to be the most influential for linear growth up to and stunting at 24 months.
28 29 234 30 31	Conditions after birth, such as feeding practices and morbidity, were less important for linear growth trajectories and stunting at
32 235 33 34	two years. The difference between the identified subgroups of children with the highest and lowest probabilities of stunting was as
<sup>35</sup> 36236	high.
37 38	
39 40 237 41	The most important predictors of stunting at 24 months were different indicators of size at birth, maternal height, asset
42 43 238 44	score and maternal education. These findings are in line with a multi-country longitudinal study that found birth or enrollment
45 46 47 239 48	weight of the infant and maternal height to have the highest cumulative odds ratios for linear growth deficit up to two years of age
49 50 240 51	[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
52 53 54 55	a pooled analysis of 19 birth cohorts with longitudinal follow-up, 20% of stunting was attributable to small-for-gestational-age
56 57 242 58	weight at birth [17]. That study did not include any post-natal factors in the analysis. In a study in Indonesia, neonatal length and
59 60 243	weight were the strongest predictors of nutritional status and increases in weight and length during infancy [44]. Our study

1	
<sup>2</sup> 3 244	included both pre- and post-natal factors and, in contrast to most other studies, assessed not only the relative importance of
5 6 245 7 8	different potential predictors, but also the public health importance of each element.
9 10 246 11 12	In a study with pooled data from five Demographic and Health Surveys in South Asia, maternal height and underweight,
13 14 247 15	household wealth, maternal education, and minimum dietary diversity were found to be the most important factors among
16 17 248 18 19	children aged 6—23 months [16]. Similar results were reported from a study in India [45]. These studies were, however, cross-
20 249 21 22 23	sectional, without access to birth characteristics.
24 25 26	Maternal height is a strong determinant of fetal growth [46] that indirectly reflect the epigenetic heredity. Maternal
27 28 251 29 30	height is directly associated with the uterine volume [47], cephalo-pelvic disproportion and subsequent infant and childhood
<sup>31</sup> 252 32 33	stunting, and child mortality [48,49]. In a previous analysis of the MINIMat cohort, a short maternal height was strongly associated
34 35 253 36 37	with stunting all the way up to 10 years of age [49]. Thus, factors that well precede pregnancy generate a vicious intergenerational
38 254 39 40	cycle, where small mothers give birth to small children of whom a high proportion become and remain stunted. In the conditional
41 42 43 44	inference trees for stunting at 24 months, children who were born with a higher HAZ but who had shorter mothers were as likely to
45 256 46 47	be stunted as children with lower HAZ at birth but with a taller mother. This finding suggests that intergenerational improvements
<sup>48</sup> 257 49 50 51	in height are achievable and that interventions with a particular focus on adolescents and women of reproductive health are needed
51 52 258 53 54 55	to break the vicious intergenerational cycle.
56 259 57 58	A strong relationship between stunting and poverty has been reported from many low-middle income settings [50]. Asset
<sup>59</sup> 260 60	score and other socioeconomic markers, such as the number of shoes and saris the mother owned, were highly ranked in the

1 2 3 261 4	random forest analysis and categorized subgroups with a higher probability of stunting and undesirable linear growth trajectories.
5 6 262 7	Poverty is associated with unfavorable food and sanitation practices that can lead to poor nutrition and an increased occurrence of
8 9 263 10 11	infections during pregnancy, infancy, and childhood. Poverty increases the risk of maternal stress, depression [51] and weak
12 13 264 14	mother-to-child interaction and stimulation.
15 16 17 265 18 19	The number of shoes and saris the mother owns might also be markers of the woman's status in the household. During
<sup>20</sup> 266 21 22	the last few decades, the importance of women's position in household and society for child nutrition has been emphasized [52].
23 24 267 25	Maternal status is associated with food allocation to mother and child, and a higher level of maternal autonomy has been
26 27 268 28	associated with better child weight and lower levels of stunting [53]. The subordinate position of women in South Asia has been
29 30 31 269 32	suggested to be a contributor to the high prevalence of child undernutrition in the region, compared to other areas with equivalent
33 34 270 35 36	levels of economic growth and food security [52].
37 38 271 39 40	An acknowledged way of increasing women's position is through improved education. The remarkable health
41 42 43 272	achievements in Bangladesh over the past two decades can partly be attributed to the progress in access to education, especially at
44 45 273 46 47	primary level and for girls [54]. However, there is a considerable risk of not completing primary school for both girls and boys [55].
48 49 50	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
51 52 275 53	conditional decision trees models for stunting and change in HAZ, the cut-off values for paternal and maternal education in the
54 55 276 56 57	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,
58 59 60	furthering the importance of girls and boys not only enrolling in but also continuing at school.

<sup>2</sup> 278 4	It may seem contradictory that children who were born with a very short length had the smallest change in HAZ. This
5 6 279 7 8	finding most likely reflects a situation where linear growth had already been severely restricted in fetal life.
9 10 280 11 12	A multi-country pooled analysis of cohort studies showed that a higher cumulative burden of diarrhea increased the risk
<sup>13</sup> 281 14 281 15	of stunting [57]. In situations, where measles still occurred, its impact on growth and mortality risks were repeatedly documented
16 17 282 18 19	[58]. One explanation to the discrepancy between our results and previous findings could be Bangladesh's remarkable success in
20 283 21 22	achieving the globally highest coverage of oral rehydration therapy in diarrhea [59], which may have reduced the impact on linear
23 24 284 25	growth. Another factor is the almost universal immunization coverage [60,61] that has reduced or partly eliminated immunization-
26 27 <u>285</u> 28 29	preventable morbidity and the subsequent effect on growth. Our previous publications on the MINIMat prenatal nutrition
<sup>30</sup> 286 31 286 32	interventions' effects on child growth and mortality were not mediated through morbidity [23,62], further supporting the modest
33 34 287 35	impact of child morbidity on linear growth in our sample [34]. In other settings with lower coverage of diarrhea treatment and
36 37 288 38 39	immunization, the relative importance of these factors may be greater.
40 41 42 289 43	Suboptimal infant and early childhood feeding practices have, in earlier studies, been reported as significant risk factors
44 45 290 46 47	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–
47 <sup>48</sup> 291 49 50	24 months had been randomized to appropriate complementary foods [64]. The infant feeding variables included in our analysis
51 52 292 53	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
54 55 293 56 57	documented effects of complementary feeding programs on stunting, these interventions are often the priority in efforts to combat
57 58 59 294 60	stunting.

Enseignement Superieur (ABES) . Protected by copyright, including for uses related to text and data mining, Al training, and similar technologies.

1 2 3 4	.95	The nutrition interventions from pre-conception to two years of age currently recommended by the WHO include efforts
7	96	to ensure exclusive breastfeeding, adequate complementary feeding, appropriate nutritional care of sick and malnourished
8 9 2 10 11	97	children and proper intake of vitamin A, iron and iodine for women and children [19]. All of these, except micronutrient
12 13 2 14	98	supplementation to pregnant women, are focused on the postnatal period from birth up to two years. Our results strengthen the
15 16 2 17 18	.99	evidence that the process of becoming stunted already begins in utero, as well as the importance of intergenerational effects.
19 20 21	00	Although worthwhile, the present focus on postnatal interventions results in missed opportunities to intervene before or during the
22 23 3 24	01	first nine months when the process of stunting is established.
25 26 27 3 28 29	02	So, what possibilities do we have to improve the postnatal linear growth trajectories prenatally? Attained height is mainly
30 31 31 32	03	dependent on one's genetic potential for linear growth, in turn determined by DNA sequence polymorphism [65,66] and epigenetic
33 34 3 35	04	heredity [67], and to some extent the environment. The modulation of non-DNA sequence epigenetic heredity has been proposed to
36 37 3 38 39	05	be one of the leading factors explaining variations in height and height changes over generations[67], especially in more deprived
40 41 3 42	06	populations [68]. Postnatal interventions can influence factors in the environment that constrain the ability to increase linear
43 44 3 45 46	07	growth, while prenatal interventions also have the potential to modulate the actual growth potential through an epigenetic
47 48 49	08	modification that results from changes to gene expression in response to the fetal environment.
50 51 52 3 53	09	Established prenatal nutritional interventions include balanced energy-protein supplementation, multiple micronutrient
54 55 3 56 57	10	supplements, and nutritional counseling and education. Unfortunately, most studies evaluating these interventions report only
58 59 60	11	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

48 325

Page 21 of 47

$^{1}_{3}$ 328	The extensive database that was available for our analysis covered a wide range of household, family, and environmental factors,
4 5 6 329 7	child characteristics at birth, feeding, and morbidity. Infant and young child growth was carefully assessed from birth up to two
8 9 330 10	years. The MINIMat cohort was implemented in an excellent research infrastructure that fulfills the prerequisites for obtaining
11 12 13 331 14	high-quality longitudinal data. Experienced field workers and study nurses collected data on the 309 variables during pregnancy
15 16 332 17	and the following two years. They received repeated training, including standardization exercises, and were supervised by senior
18 19 20 333 21	medical doctors.
22 23 24 334 25	Some potential determinants were not present in the database. Household water, sanitation, and hygiene (WASH)
26 27 335 28	characteristics were limited to information on arsenic contamination of the drinking water, but diarrhea and other morbidity
29 30 31 336 32	information were included in our analyses. Further, the cohort did not include the collection of stools for the study of
33 34 337 35	enteropathogens in the child, which may be associated with the risk of stunting [11]. Paternal height, which may be related to fetal
36 37 338 38 39	growth, was not available [80]. The mothers' smoking habits were not represented in the data, as smoking was extremely rare
40 41 339 42	among women in the study area.
43 44 45 340 46 47	There were slight differences in basic characteristics of the analyzed and non-analyzed groups. These differences had
48 49 50	most likely no influence on the primary outcomes of this study. There were no or few missing values of the critical variables that
51 52 342 53	ranked high in the random forest and defined the sub-groups in the conditional inference trees. A sub-study was carried out to
54 55 343 56 57	ensure the most accurate method to impute missing data. Thus, it is also highly unlikely that missing data influenced the main
58 59 60	findings.

2 2/	15 Some of the includ	
4		ed variables like "household asset score" are composite variables, which depend on individual variables
5 6 34 7 8	16 like TV ownership, number o	f cows, etc. Presence of both composite and individual variables creates computational problems for
9 32 10 11	17 traditional models like linear	r regression and for some machine learning models due to a possible high correlation between the
12 13 34 14	18 individual and the composit	e variables. However, CIT methods perform automatic variable selection by choosing the most relevant
15 16 34 17 18	19 variable (with the strongest a	association to the response) at each decision tree split step [39]. Accordingly, these methods
19 20 35 21	50 automatically choose either a	a composite variable or an individual variable at each split step based on the relevance of this variable
22 23 35 24 25	51 to the response.	
26 27 35 28 29	52 Traditional metho	ds like linear regression often have lower predictive power than data mining methods. In some cases,
<sup>30</sup> 31 32	53 the traditional methods are	not even possible to compute due to a high number of predictor variables and complex interactions.
33 34 35 35 36	54 The method used in this wor	k, Conditional Inference Trees, belongs to the class of Interpretable Machine Learning models and
37 35 38 39	55 display precise information o	on the priority, size, and direction of the association of the predictors with the outcome. In addition, the
40 41 35 42	56 risk group identification, inc	luding the prioritization and relevant cut-offs of risk factors, can be of high public health relevance for
43 44 35 45 46	57 the design and targeting of $z$	appropriate interventions with the most significant benefit. Thus, we believe that the CIT framework
47 48 49 50	58 has a large potential in publ	ic health and medical applications.
51 52 35 53	59 It can be noted tha	t the CRF and the CIT models are not fully comparable. This can be explained by two factors. Firstly,
54 55 36 56 57	50 many predictors that were in	nportant in the CRF model are relatively highly correlated and thus have a similar relationship to the
58 59 36 60	51 response. Once one of these	variables is selected by the decision tree in a split, there is a high chance that the remaining correlated

22

1 2 3 362	variables (although also important according to the CRF) will not be picked up as the next splitting variable. Secondly, the CRF
4 5	
6 363 7	models and the CIT models cannot be matched directly. The CRF is a combination of many trees and is thus a more flexible model
8 9 364 10 11	than a CIT. However, CRFs are nearly black-box models: the only interpretable information that these models deliver is the variable
12 13 365 14	importance measure. On the contrary, CITs are "transparent" and interpretable models but have a smaller predictive power. This is
15 16 366 17 18	another reason why these models are not generally capable of efficiently embedding all the variables that are important in the
19 20 367 21	CRFs.
22 23 24 368 25	Another potential limitation is that decision trees do not deliver p-values or confidence intervals. The cross-validation
26 27 369 28 29	method, however, ensures that the selected tree is optimal. This validation method was chosen superior to other model validation
<sup>30</sup> <sub>31</sub> <sub>32</sub> 370	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
33 34 371 35 36	computational burden.
37 38 372 39 40	The study setting was a low socioeconomic area in rural Bangladesh, where maternal and child undernutrition in early
41 42 43 373	life still is widespread. The growth trajectories of our cohort were consistent with established growth trajectories in South Asia,
44 45 374 46 47	where children are born below the WHO growth reference and falter dramatically up to 24 months of age [5]. The sub-continents of
48 375 49 50	South Asia and Sub-Saharan Africa share similar proportions of stunted children and faltering patterns. The sub-Saharan African
51 52 376 53 54	children are however, on average born slightly bigger than children in South Asia [5], which makes our results mainly relevant for
55 377 56 57 58	the South Asian context.
59 60 378	Conclusion

1 2 3 379	This cohort study of determinants of young child stunting in a rural Bangladeshi setting included a wide range of high-quality pre-
4 5 6 380 7	and postnatal data, household and family information, environmental factors, child characteristics at birth, infant feeding, and
8 9 381 10 11	morbidity. Prenatal factors including birth size, the mother's anthropometry, and parental education were the most critical factors
12 13 382 14	for stunting at 24 months. These results should be seen in contrast to present practice and recommendations that mainly are
15 16 383 17	limited to child interventions. The findings emphasize the benefit of interventions before conception and during pregnancy to
18 19 20 384 21 22	reach a substantial reduction in stunting.
23 24 385 25 26	
27 28 386 29 30	
31 32 387 33	
34 35 36 37 388	
38 39 40 41 200	
41 389 42	Contributors
43 44 390 45 46	PS contributed to study design, data analysis and interpretation of the results and had the main responsibility of writing the paper.
47 48 49	LÅP and SEA were principal investigators of the MINIMat project. ECE, LÅP and KS contributed to the study design. ECE, RN, AR
50 51 392 52	and AIK took part in and supervised data collection. PS, OS, and KS analysed the data. All authors contributed to the preparation
53 54 393 55 56 57	of the database, interpretation of the results and reviewed and approved the final version of the manuscript.
58 59 394	Competing interests
60 395	The authors declare that they have no competing interests
	24 For peer review only - http://bmjopen.bmj.com/site/about/guidelines.xhtml

1	
2 3 396 4	Data sharing statement
5 397 6	Data are available from the authors upon reasonable request and with permission from the principal investigator of the MINIMat
7 8 398 9	study.
10 399 11 12 13	
14 400 15	Funding
16 401 17 18	The MINIMat research study was funded by the icddr,b, United Nations Children's Emergency Fund, Swedish International
<sup>19</sup> <sub>20</sub> 402 21	Development Cooperation Agency (Sida), UK Medical Research Council, Swedish Research Council, Department for International
22 23 403 24 25	Development (DFID), Japan Society for the Promotion of Science, Child Health and Nutrition Research Initiative, Uppsala
26 404 27 28	University, and US Agency for International Development. The funding agencies had no role in the design and conduct of the
29 30 405 31 32	study, in the collection, analysis, and interpretation of the data, or in the preparation, review, or approval of the manuscript.
33 34 406 35	Acknowledgements
<sup>36</sup> 407 37 38	We thank the participants and their families in Matlab for their continuing involvement in the MINIMat trial, and the field-team
<ol> <li>39</li> <li>40 408</li> <li>41</li> <li>42</li> <li>43</li> <li>44</li> <li>45</li> <li>46</li> <li>47</li> <li>48</li> <li>49</li> <li>50</li> <li>51</li> <li>52</li> <li>53</li> <li>54</li> <li>55</li> <li>56</li> <li>57</li> <li>58</li> <li>59</li> <li>60</li> </ol>	members and data management staff for their excellent work.

Page 26 of 47

1 2 3 410 4 5 411 6 7 8 9 9	
11	
12 13 14 15 16 17 18 19	
18         19         20         21         22         23         24         25         26         27         28         29         30         31         32         33         34         35         36         37	
26 27 28 29 30 31	
38 39 40 41 42 43 44 45 46 47 48 49	
45 46 47 48 49 50	
50 51 52 53 54 55 56 57 58 59 60	
57 58 59 60	

1 2 413 3	Legend to Figures
4 5 414 6 7	Figure 1. Factors, variables and outcomes included in the analysis of data from the MINIMat cohort, Bangladesh. Grouping
8 415 9 10 11	according to the WHO conceptual framework on childhood stunting [21]
12 13 14	Figure 2. Flow chart of pregnant women and their children included in the data mining analyses of the MINIMat cohort from
15 16 417 17 18	conception to two years of age.
19 20 418 21 22 23	Figure 3. Height-for-age Z-scores from birth to 24 months in the MINIMat cohort in rural Bangladesh.
<sup>24</sup> 419 25 26	Figure 4. Conditional random forest plot ranking the relative importance of 30 predictors with regard to their ability to explain the
27 28 420 29	presence of stunting at 24 months of age. The MINIMat cohort in rural Bangladesh. Colour coding according to the WHO
30 31 421 32 33 34	conceptual framework on causes of stunting.
<sup>35</sup> 422 36 37	Figure 5. Conditional random forest plot ranking the relative importance of 30 predictors with regard to their ability to explain the
<sup>38</sup> 39 423 40 41	variation in change in HAZ (🛆 HAZ) from birth to 24 months of age. The MINIMat cohort in rural Bangladesh. Colour coding
42 424 43 44 45	according to the WHO conceptual framework on causes of stunting.
46 47 48	Figure 6. Conditional inference tree identifying sub-groups with different probabilities of stunting at 24 months. The MINIMat
49 50 426 51 52 53	cohort in rural Bangladesh.
54 427 55 56	Figure 7. Conditional inference tree identifying sub-groups with different mean change in HAZ ( $\Delta$ HAZ=HAZ <sub>24</sub> –HAZ <sub>o</sub> ) o–24
57 428 58 59 60 429	months within the MINIMat cohort in rural Bangladesh.

1 2 420		
<sup>2</sup> 430 4		
5 6		
7 431 8		
9 10		
$^{11}_{12}432$	Reference	2S
$^{13}_{14}433$	1	Pradhan M, Sahn DE, Younger SD. Decomposing world health inequality. <i>Journal of Health Economics</i>
<sup>15</sup> 434 16		2003; <b>22</b> :271—93. doi:10.1016/S0167-6296(02)00123-6
17 18 435	2	Victora CG, Adair L, Fall C, <i>et al.</i> Maternal and child undernutrition: consequences for adult health and human
$^{19}_{20}436$		capital. <i>Lancet</i> 2008; <b>371</b> :340—57. doi:10.1016/S0140-6736(07)61692-4
21 22 23 437	3	Black RE, Victora CG, Walker SP, <i>et al.</i> Maternal and child undernutrition and overweight in low-income and
24 4 38	5	middle-income countries. <i>The Lancet</i> 2013.
25 26 27 439		
	4	General Assembly, UNICEF. Transforming our world: the 2030 Agenda for Sustainable Development. <i>United</i>
<sup>28</sup> 440 29 30		<i>Nations</i> 2015;:1–35.
31 441 32	5	Victora CG, de Onis M, Hallal PC, <i>et al.</i> Worldwide timing of growth faltering: revisiting implications for
<sub>33</sub> 442		interventions. <i>Pediatrics</i> 2010; <b>125</b> :e473—80. doi:10.1542/peds.2009-1519
34 35 36 443	6	Black RE, Allen LH, Bhutta ZA, <i>et al.</i> Maternal and child undernutrition: global and regional exposures and health
37 444		consequences. <i>The Lancet</i> 2008; <b>371</b> :243—60. doi:10.1016/S0140-6736(07)61690-0
38 39		
40 445 41	7	Bayer R, Galea S. Public Health in the Precision-Medicine Era PubMed - NCBI. <i>N Engl J Med</i> 2015; <b>373</b> :499—501.
41 42 446 43		doi:10.1056/NEJMp1506241
<sup>44</sup> 447 45	8	Martorell R, Ramakrishnan U, Schroeder DG, <i>et al.</i> Intrauterine growth retardation, body size, body composition
46 448 47		and physical performance in adolescence. <i>Eur J Clin Nutr</i> 1998; <b>52 Suppl 1</b> :S43–52–discussionS52–3.
48 49 449	9	Frongillo EA Jr. <i>Symposium: Causes and Etiology of Stunting. J Nutr</i> 1999; <b>129</b> :529S—530S. doi:10.1016/S0140-
<sup>50</sup> 450	-	6736(94)91337-4
52		
53 451 54	10	Cole TJ, Parkin JM. Infection and its effect on the growth of young children: A comparison of The Gambia and
55 452 56		Uganda. <i>Trans R Soc Trop Med Hyg</i> 1977; <b>71</b> :196—8. doi:10.1016/0035-9203(77)90005-0
<sup>57</sup> <sub>58</sub> 453	11	Investigators M-EN. Childhood stunting in relation to the pre- and postnatal environment during the first 2 years
59 454 60		of life: The MAL-ED longitudinal birth cohort study. <i>PLoS Medicine</i> 2017;:1—21. doi:10.1371/journal.pmed.1002408

28

Page 29 of 47

1		
$^2_3$ 455	12	Espo M, Kulmala T, Maleta K, <i>et al.</i> Determinants of linear growth and predictors of severe stunting during
4 456 5		infancy in rural Malawi. <i>Acta Paediatrica</i> 2002; <b>91</b> :1364—70. doi:10.1111/j.1651-2227.2002.tb02835.x
6 7 457	13	Rahman A, Chowdhury S. Determinants of Chronic Malnutrtion Among Preschool Children In Bangladesh. J
8 9 458		<i>Biosoc Sci</i> 2006; <b>39</b> :161. doi:10.1017/S0021932006001295
10 11 459 12	14	Amin R, Murumi MW, Carabaza AFM. Factors Associated with Stunting Among 0-23 Months-Old Children in
13 460 14		Rural Bangladesh. <i>Journal of Nutrition Education and Behavior</i> 2017; <b>49</b> :S21. doi:10.1016/j.jneb.2017.05.061
15 16 461	15	Hong R, Banta JE, Betancourt JA. Relationship between household wealth inequality and chronic childhood under-
<sup>17</sup> 462 18 19		nutrition in Bangladesh. <i>International Journal for Equity in Health</i> 2006; <b>5</b> :1—10. doi:10.1186/1475-9276-5-15
20 463	16	Kim R, Mejía-Guevara I, Corsi DJ, <i>et al.</i> Relative importance of 13 correlates of child stunting in South Asia: Insights
21 22 464		from nationally representative data from Afghanistan, Bangladesh, India, Nepal, and Pakistan. <i>Social Science &amp;</i>
<sup>23</sup> 465 24 25		<i>Medicine</i> 2017; <b>187</b> :144—54. doi:10.1016/j.socscimed.2017.06.017
26 466 27	17	Christian P, Lee SE, Donahue Angel M, <i>et al.</i> Risk of childhood undernutrition related to small-for-gestational age
28 467 29		and preterm birth in low- and middle-income countries. <i>Int J Epidemiol</i> 2013; <b>42</b> :1340—55. doi:10.1093/ije/dyt109
<sup>30</sup> 31 468	18	World Health Organization. WHO Guidline: Daily iron and folic acid supplementation in pregnant women. WHO
<sup>32</sup> 469 <sup>33</sup>		2012;:1–32.
34 35 470	19	World Health Organization. Essential Nutrition Actions: improving maternal, newborn, infant and young child
36 37 471 38		health and nutrition. <i>WHO</i> 2013;:1—144.
<sup>39</sup> 472 40	20	Bhutta ZA, Das JK, Arjumand R, <i>et al.</i> Maternal and Child Nutrition 2 Evidence-based interventions for
41 473		improvement of maternaland child nutrition: what can be done and at what cost? <i>The Lancet</i> 2013; <b>382</b> :452–77.
42 43 474 44		doi:10.1016/S0140-6736(13)60996-4
<sup>45</sup> <sub>46</sub> 475	21	Stewart CP, Iannotti L, Dewey KG, <i>et al.</i> Contextualising complementary feeding in a broader framework for
47 476 48		stunting prevention. <i>Matern Child Nutr</i> 2013; <b>9</b> :27-45. doi:10.1111/mcn.12088
49 50 477 51	22	Breiman L, Friedman J, Stone CJ, <i>et al. Classification and Regression Trees</i> . Chapman and Hall/CRC 1984.
<sup>52</sup> 53 478	23	Persson LÅ, Arifeen S, Ekström E-C, <i>et al.</i> Effects of Prenatal Micronutrient and Early Food Supplementation on
<sup>54</sup> 479 55		Maternal Hemoglobin, Birth Weight, and Infant Mortality Among Children in Bangladesh: The MINIMat
56 480 57		Randomized Trial. <i>JAMA</i> 2012; <b>307</b> :2050—9. doi:10.1001/jama.2012.4061
58 59 481	24	Khan AI, Hawkesworth S, Ekström E-C, <i>et al.</i> Effects of exclusive breastfeeding intervention on child growth and
<sup>60</sup> 482		body composition: the MINIMat trial, Bangladesh. <i>Acta Paediatrica</i> 2013; <b>102</b> :815—23. doi:10.1111/apa.12282

Page 30 of 47

## BMJ Open

1		
$^2_3$ 483	25	Gwatkin DR, Rustein S, Johnson K, et al. Socio-economic differences in health nutrition and population in
4 484		Bangladesh. Washington, DC: HNP/Poverty Thematic Group of the World Bank 2000.
5 6		
7 485 8 40 (	26	Frongillo EA, Chowdhury N, Ekström E-C, <i>et al.</i> Understanding the Experience of Household Food Insecurity in
<sub>9</sub> 486		Rural Bangladesh Leads to a Measure Different from That Used in Other Countries
10 11 12 487	.]Nutr2003;:	41=8-62
12 <sup>107</sup> 13	. 977447 2003,.	
14 488	27	Straus MA, Douglas EM. A Short Form of the Revised Conflict Tactics Scales, and Typologies for Severity and
15 16 489		Mutuality. <i>Violence and Victims</i> 2004; <b>19</b> :1—15.
17 19		
<sup>18</sup> 490 19	28	WHO. WHO Multi-country Study on Women's Health and Domestic Violence against Women. <i>WHO</i> 2005;:1–38.
20 21 491	29	Asling-Monemi K, Naved RT, Persson LA. Violence against women and the risk of fetal and early childhood growth
22 23 492	-9	impairment: a cohort study in rural Bangladesh. <i>Arch Dis Child</i> 2009; <b>94</b> :775–9. doi:10.1136/adc.2008.144444
24		impairment. a conort study in rurat bangtaucsn. 7 ben Dis Child 2009, <b>94</b> .775–9. uot.10.1130/auc.2000.144444
<sup>25</sup> 493	30	Rahman A, Vahter M, Smith AH, <i>et al.</i> Arsenic Exposure During Pregnancy and Size at Birth: A Prospective
27 494		Cohort Study in Bangladesh. <i>American Journal of Epidemiology</i> 2008; <b>169</b> :304—12. doi:10.1093/aje/kwn332
28 29		
30 495	31	Arifeen SE, Black RE, Caulfield LE, <i>et al.</i> Infant growth patterns in the slums of Dhaka in relation to birth weight,
<sup>31</sup> 32 496		intrauterine growth retardation, and prematurity. <i>Am J Clin Nutr</i> 2000; <b>72</b> :1010—7.
33 <sup>34</sup> 497	22	World Health Organization. Indicators for assessing breastfeeding practices. <i>WHO</i> 1991;:1—16.
35 36	32	worth realth Organization. Indicators for assessing breastfeeding practices. Write 1991;:1–16.
37 498	33	Moore SE, Prentice AM, Coward WA, <i>et al.</i> Use of stable-isotope techniques to validate infant feeding practices
38 39 499		reported by Bangladeshi women receiving breastfeeding counseling. <i>Am J Clin Nutr</i> 2007;:1075–82.
40		
41 500 42	34	Kallioinen M, Ekström E-C, Khan AI, <i>et al.</i> Prenatal early food and multiple micronutrient supplementation trial
43 501 44		reduced infant mortality in Bangladesh, but did not influence morbidity. <i>Acta Paediatrica</i> 2017; <b>382</b> :427–8.
45 502		doi:10.1111/apa.14009
46 47 503	27	Arnold BF, Galiani S, Ram PK, <i>et al.</i> Optimal Recall Period for Caregiver-reported Illness in Risk Factor and
47 48 503 49 504	35	
50		Intervention Studies: A Multicountry Study. <i>American Journal of Epidemiology</i> 2013; <b>177</b> :361—70.
51 505 52		doi:10.1093/aje/kws281
<sup>53</sup> 54 506	36	Onis M de, Onyango AW, Borghi E, <i>et al.</i> Development of a WHO growth reference for school-aged children and
<sup>55</sup> 507		adolescents. <i>Bulletin of the World Health Organization</i> 2007; <b>85</b> :660—7. doi:10.1590/S0042-96862007000900010
56 57		
58 508 59	37	Jönsson P, Wohlin C. An Evaluation of K-Nearest Neighbour Imputation Using Likert Data. <i>Proceedings of th</i>
<sub>60</sub> 509		International Symposium on Software Metrics 2004;:108–18.

30

Page 31 of 47

## BMJ Open

1		
${}^2_3$ 510	38	Stekhoven DJ, Buhlmann P. MissForest—non-parametric missing value imputation for mixed-type data.
4 511 5		Bioinformatics 2012;:112—8.
6 7 512	39	Torsten Hothorn KHAZ. ctree: Conditional Inference Trees. www.cran.r-project.org. 2016;:1—21.https://cran.r-
8 9 513 10		project.org/web/packages/partykit/vignettes/ctree.pdf (accessed 24 Mar2017) page 655.
<sup>11</sup> 514 12	40	Hastie T, Tibshirani R, Friedman J. <i>The Elements of Statistical Learning</i> . New York, NY: : Springer Science &
13 515 14		Business Media 2013.
15 16 516 17	41	R Core Team. R: A language and environment for stastical computing. 2017. httpa://www.R-project.org/
<sup>18</sup> 517 19 20	42	Achim Zeileis THKH. party: A Laboratory for Recursive Partytioning. 2017;:1–18.
21 518 22	43	Christian P, Lee SE, Donahue Angel M, <i>et al.</i> Risk of childhood undernutrition related to small-for-gestational age
23 519 24		and preterm birth in low- and middle-income countries. <i>Int J Epidemiol</i> 2013; <b>42</b> :1340—55. doi:10.1093/ije/dyt109
<sup>25</sup> <sub>26</sub> 520	44	Schmidt MK, Muslimatun S, West CE, <i>et al.</i> Nutritional Status and Linear Growth of Indonesian Infants in West
27 521 28 29		Java Are Determined More by Prenatal Environment than by Postnatal Factors. <i>J Nutr</i> 2002;:2202—7.
30 522	45	Corsi DJ, Mejía-Guevara I, Subramanian SV. Risk factors for chronic undernutrition among children in India:
<sup>31</sup> 32 523		Estimating relative importance, population attributable risk and fractions. Social Science & Medicine 2016; <b>157</b> :165-
<sup>33</sup> 524 34 35		85. doi:10.1016/j.socscimed.2015.11.014
36 525 37 38	46	Kramer MS. The Epidemiology of Adverse Pregnancy Outcomes: An Overview. <i>J Nutr</i> 2003;:1592S—1596S.
<sub>39</sub> 526	47	A WHO Collaborative Study of Maternal Anhropomehtry and Pregnancy Outcomes. International Journal of
40 527 41 42		Gynecology and Obstetrics 1996; <b>57</b> :1–15.
43 528 44	48	Özaltin E. Association of Maternal Stature With Offspring Mortality, Underweight, and Stunting in Low- to
45 529 46		Middle-Income Countries. <i>JAMA</i> 2010; <b>303</b> :1507—21. doi:10.1001/jama.2010.450
47 48 530	49	Svefors P, Rahman A, Ekström E-C, <i>et al.</i> Stunted at 10 Years. Linear Growth Trajectories and Stunting from Birth
49 531 50		to Pre-Adolescence in a Rural Bangladeshi Cohort. <i>PLoS ONE</i> 2016 <b>,11</b> :e0149700—18.
51 532 52		doi:10.1371/journal.pone.0149700
<sup>53</sup> 54 533	50	Van de Poel E, Hosseinpoor AR, Speybroeck N, <i>et al.</i> Socioeconomic inequality in malnutrition in developing
<sup>55</sup> 534 56 57		countries. <i>Bulletin of the World Health Organization</i> 2008; <b>86</b> :282–91. doi:10.1590/S0042-96862008000400013
58 535	51	Bradley RH, Corwyn RF. Socioeconomic Status and Child Development. Annual Review of Phsycology 2002; <b>53</b> :371–
<sup>59</sup> 60 536		99.

Enseignement Superieur (ABES) . Protected by copyright, including for uses related to text and data mining, Al training, and similar technologies.

1		
<sup>2</sup> 537 4	52	Ramalingaswami V, U J, J R. The Asian enigma. 1996.
5 538 6	53	Shroff M, Griffiths P, Adair L, <i>et al.</i> Maternal autonomy is inversely related to child stunting in Andhra Pradesh,
7 539 8		India. <i>Matern Child Nutr</i> 2009; <b>5</b> :64—74. doi:10.1111/j.1740-8709.2008.00161.x
9 10 540	54	Chowdhury AMR, Bhuiya A, Chowdhury ME, <i>et al.</i> The Bangladesh paradox: exceptional health achievement
11 541 12 13		despite economic poverty. <i>The Lancet</i> 2013; <b>382</b> :1734—45. doi:10.1016/S0140-6736(13)62148-0
14 542	55	Asadullah MN, Chaudhury N. Girls school enrolement figures are good in Bangladesh but drop out rates are
15 16 543 17		worrying. <i>Development and Cooperation</i> 2017; <b>59</b> :447–72. doi:10.1086/681929
<sup>18</sup> 544	56	UNESCO. EFA 2015 national review: Bangladesh; 2014. unesdoc.unesco.org.
20 545 21 22		http://unesdoc.unesco.org/images/0023/002305/230507e.pdf (accessed 13 Feb2018).
23 546	57	Checkley W, Buckley G, Gilman RH, <i>et al.</i> Multi-country analysis of the effects of diarrhoea on childhood stunting.
<sup>24</sup> 547 25 26		<i>Int J Epidemiol</i> 2008; <b>37</b> :816—30. doi:10.1093/ije/dyn099
27 548 28	58	Koster FT, Curlin GC, Aziz KMA, <i>et al.</i> Synergistic impact of measels and diarrhoea on nutrtion and mortality in
29 549 30		Bangladesh. <i>Bulletin of the World Health Organization</i> 1981;:901—8.
$\frac{31}{32}550$	59	Arifeen DrPH El S, MPH AC, ScD LR, <i>et al.</i> Community-based approaches and partnerships: innovations in
<sup>33</sup> 551 34 35		health-service delivery in Bangladesh. <i>The Lancet</i> 2013; <b>382</b> :2012—26. doi:10.1016/S0140-6736(13)62149-2
36 552 37	60	WHO, UNICEF. Bangladesh: WHO and UNICEF estimates of immunization coverage: 2016 revision. 2016;:1—30.
<sup>38</sup> 39 553		Chowdhury AMR, Bhuiya A, Chowdhury ME, <i>et al.</i> The Bangladesh paradox: exceptional health achievement
40 554 41 42		despite economic poverty. <i>The Lancet</i> 2013; <b>382</b> :1734—45. doi:10.1016/S0140-6736(13)62148-0
43 555	62	Khan AI, Kabir I, Ekstrom EC, <i>et al.</i> Effects of prenatal food and micronutrient supplementation on child growth
44 45 556 46		from birth to 54 months of age: a randomized trial in Bangladesh. <i>Nutr J</i> 2011.
47 48 557	63	Martorell R. Physical growth and development of the malnourished child: Contributions from 50 years of research
49 558 50 51		at INCAP. <i>Food Nutr Bull</i> 2010; <b>31</b> :68—82.
52 559		Imdad A, Yakoob MY, Bhutta ZA. Impact of maternal education about complementary feeding and provision of
$\frac{53}{54}560$		complementary foods on child growth in developing countries. <i>BMC Public Health</i> 2011 <b>;11</b> :S25. doi:10.1186/1471-
<sup>55</sup> 561 56 57		2458-11-S3-S25
58 562	65	Weedon MN, Frayling TM. Reaching new heights: insights into the genetics of human stature. <i>Trends in Genetics</i>
<sup>59</sup> 60 563		2008; <b>24</b> :595—603. doi:10.1016/j.tig.2008.09.006

32

Page 33 of 47

1			
$\frac{2}{3}$ 56	64	66	Silventoinen K, Sammalisto S, Perola M, <i>et al.</i> Heritability of Adult Body Height: A Comparative Study of Twin
5	65		Cohorts in Eight Countries. <i>Twin Research</i> 2003; <b>6</b> :399—408. doi:10.1375/twin.6.5.399
	66	67	Simeone P, Alberti S. Epigenetic heredity of human height. <i>Physiological Reports</i> 2014; <b>2</b> :e12047—7.
8 9 50	67		doi:10.14814/phy2.12047
10			
11 56 12		68	Silventoinen K, Kaprio J, Lahelma E, <i>et al.</i> Relative Effect of Genetic and Environmental Factors on Body Height:
13 5 <b>(</b> 14			Differences Across Birth Cohorts Among Finnish Men and Women. American Journal of Public Health
14 15 57 16	70		2000; <b>90</b> :627–30.
17 5-	71	69	Imdad A, Bhutta ZA. Effect of balanced protein energy supplementation during pregnancy on birth outcomes.
18 <sup>3</sup> 19 57			<i>BMC Public Health</i> 2011, <b>11</b> :S17. doi:10.1186/1471-2458-11-S3-S17
20			
21 22 57		70	Fall CHD, Fisher DJ, Osmond C, <i>et al.</i> Multiple Micronutrient Supplementation during Pregnancy in Low-Income
<sup>23</sup> <sub>24</sub> 57	74		Countries: A Meta-Analysis of Effects on Birth Size and Length of Gestation. <i>Food Nutr Bull</i> 2009; <b>30</b> :S533–46.
25 57 26	75		doi:10.1177/15648265090304S408
27			
28 57 29		71	Ramakrishnan U, Grant FK, Goldenberg T, <i>et al.</i> Effect of Multiple Micronutrient Supplementation on Pregnancy
29 30 57			and Infant Outcomes: A Systematic Review. <i>Paediatric and Perinatal Epidemiology</i> 2012; <b>26</b> :153—67.
31 57 32	78		doi:10.1111/j.1365-3016.2012.01276.x
33 34 57	79	72	West KP, Shamim AA, Mehra S, <i>et al.</i> Effect of Maternal Multiple Micronutrient vs Iron—Folic Acid
35 36 58	80		Supplementation on Infant Mortality and Adverse Birth Outcomes in Rural Bangladesh: The JiVitA-3 Randomized
<sup>37</sup> 58 38			Trial. JAMA 2014; <b>312</b> :2649—58. doi:10.1001/jama.2014.16819
38 39			
40 58 41	82	73	Johnson W, Darboe MK, Sosseh F, <i>et al.</i> Association of prenatal lipid-based nutritional supplementation with fetal
42 58	83		growth in rural Gambia. <i>Matern Child Nutr</i> 2017 <b>,13</b> :e12367. doi:10.1111/mcn.12367
43 44 45 58	84	74	Ashorn P, Alho L, Ashorn U, <i>et al.</i> Supplementation of Maternal Diets during Pregnancy and for 6 Months
45 46 58		/4	Postpartum and Infant Diets Thereafter with Small-Quantity Lipid-Based Nutrient Supplements Does Not
47 48 58			Promote Child Growth by 18 Months of Age in Rural Malawi: A Randomized Controlled Trial. <i>J Nutr</i>
49 50 58			2015, <b>145</b> :1345–53. doi:10.3945/jn.114.207225
50 ° °	07		20.0, <b>-40</b> .0340 - 03. 00010.3940/1001420/220
52 58 53	88	75	Adu-Afarwuah S, Lartey A, Okronipa H, <i>et al.</i> Lipid-based nutrient supplement increases the birth size of infants
54 58	89		of primiparous women in Ghana. <i>Am J Clin Nutr</i> 2015; <b>101</b> :ajcn.091546—846. doi:10.3945/ajcn.114.091546
55 56 57 59	00	C	
		76	Huybregts L, Roberfroid D, Lanou H, <i>et al.</i> Prenatal food supplementation fortified with multiple micronutrients
<sup>58</sup> 59			increases birth length: a randomized controlled trial in rural Burkina Faso. American Journal of Clinical Nutrition
60 59	92		2009; <b>90</b> :1593—600. doi:10.3945/ajcn.2009.28253

1		
$^2_3$ 593	77	Gluckman PD, Hanson MA, Cooper C, <i>et al.</i> Effect of In Utero and Early-Life Conditions on Adult Health and
4 594 5		Disease. <i>http://dxdoiorg/101056/NEJMra0708473</i> 2009; <b>359</b> :61—73. doi:10.1056/NEJMra0708473
6	2	
7 595 8 506	78	Potdar RD, Sahariah SA, Gandhi M, <i>et al.</i> Improving women's diet quality preconceptionally and during gestation:
<sub>9</sub> 596		effects on birth weight and prevalence of low birth weighta randomized controlled efficacy trial in India (Mumbai
10 597 11		Maternal Nutrition Project). American Journal of Clinical Nutrition 2014; <b>100</b> :1257—68. doi:10.3945/ajcn.114.084921
12 13 598 14	79	The World Bank. World Development Indicators: Reproductive health. 2016;:1—6.
15 16 599	80	Wills AK, Chinchwadkar MC, Joglekar CV, <i>et al.</i> Maternal and paternal height and BMI and patterns of fetal
$\frac{17}{18}600$		growth: The Pune Maternal Nutrition Study. <i>Early Human Development</i> 2010; <b>86</b> :535—40.
19 601		
20 21		doi:10.1016/j.earlhumdev.2010.07.002
22 602		
23 24		
25 26		
27		
28 29		
30 31		
32		
33 34		
35 36		
37		
38 39		
40 41		
42		
43 44		
45		
46 47		
48 49		
50		
51 52		
53		
54 55		
56 57		
58		
59 60		

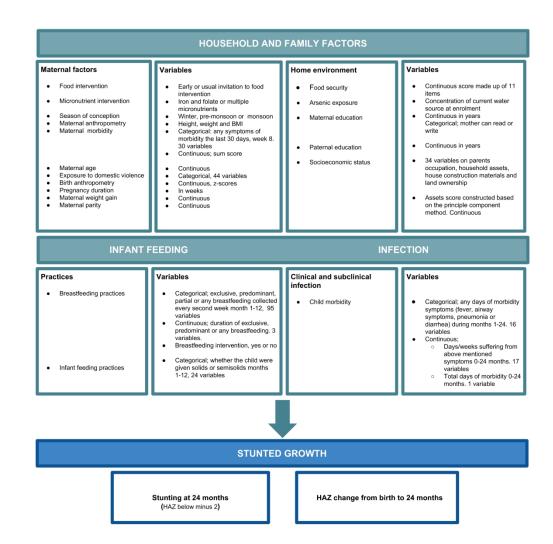
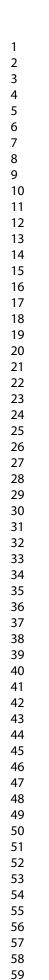


Figure 1. Factors, variables and outcomes included in the analysis of data from the MINIMat cohort, Bangladesh. Grouping according to the WHO conceptual framework on childhood stunting [20]

190x190mm (300 x 300 DPI)

BMJ Open: first published as 10.1136/bmjopen-2018-025154 on 5 August 2019. Downloaded from http://bmjopen.bmj.com/ on June 12, 2025 at Agence Bibliographique de l Enseignement Superieur (ABES) . Protected by copyright, including for uses related to text and data mining, Al training, and similar technologies.



60

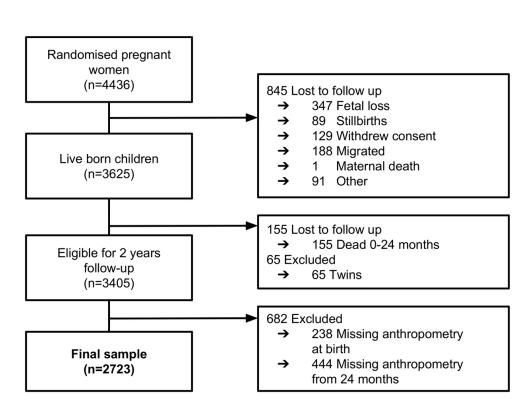
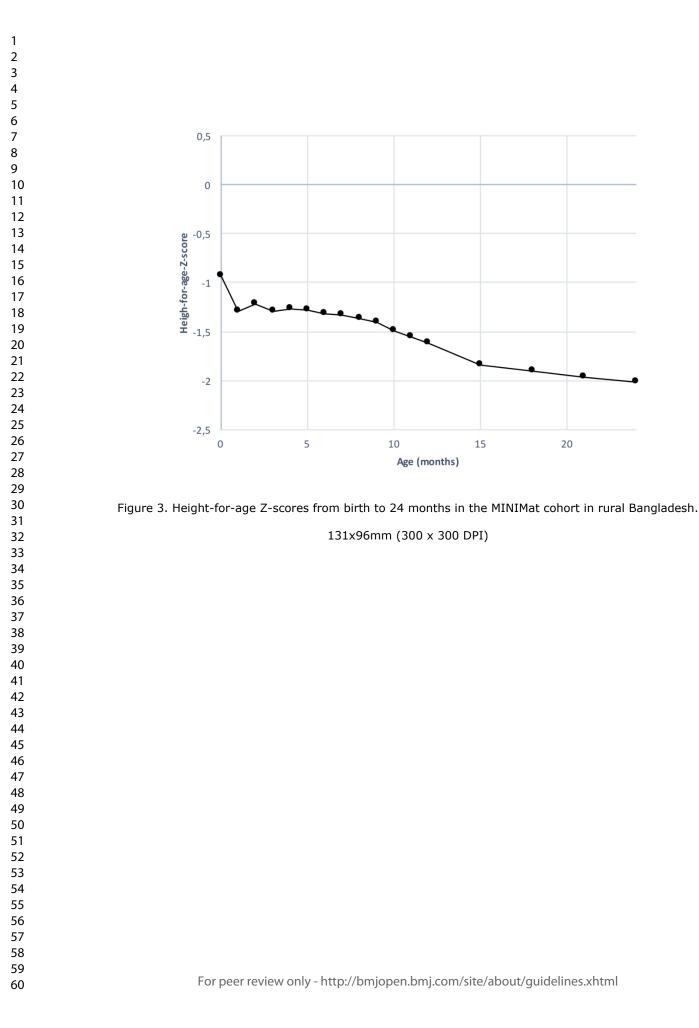


Figure 2. Flow chart of pregnant women and their children included in the data mining analyses of the MINIMat cohort from conception to two years of age.

106x80mm (300 x 300 DPI)

For peer review only - http://bmjopen.bmj.com/site/about/guidelines.xhtml

Page 37 of 47



Page 38 of 47

BMJ Open: first published as 10.1136/bmjopen-2018-025154 on 5 August 2019. Downloaded from http://bmjopen.bmj.com/ on June 12, 2025 at Agence Bibliographique de I Enseignement Superieur (ABES) . Protected by copyright, including for uses related to text and data mining, Al training, and similar technologies.

their

1				
2				
3				
4				
5				
6				
7				
8		[		
9	Height for age Z-score at birth Weight for age Z-score at birth		•	
10	Maternal height		•	
11	Small for gestational age Maternal weight at week 8	•		
12	Asset score Maternal education in years	•		
13	Chest circumference at birth Father's education in years			
14	Number of saris mother owns Head circumference at birth	•		
15	Household owns tashak or mattress	•		
16	Mother is literate Household owns clock or watch	•		
17	Number of pairs of shoes mother owns Household owns a television			
18	Household owns almirah	•		
19	Maternal BMI at 8 weeks Number of shalwar kameez mother owns	•		
20	Household owns cows Household member works on daily basis			
20	Household owns lep or quilt Food security score			
	Household owns chair or table	•	Mahamad Kashana	
22	Household owns an electric fan Child received semisolids at 12 months	•	<ul> <li>Maternal factors</li> <li>Home environment</li> </ul>	
23	Household has a stable source of income Sum of days with illness 0-24 months		<ul> <li>Infant feeding practices</li> <li>Infection</li> </ul>	
24	Number of days with chest indrawings 0-24m		• Intection	
25	Number of days with rapid breathing 0-24 month			
26		0.000 0.002 0.004	0.006 0.008	
27		Relative importance		
28				
29	Figure 4. Conditional random forest plo	at ranking the relative importa	nce of 30 predictors with rev	ard to thei
30	ability to explain the presence of stu			
31		to the WHO conceptual framev		ngiaacom
32	5 5	•	5	
33		190x134mm (300 x 300 DPI)	)	
34				
35				
36				
37				
38				
39				
40				
41				
42				
42 43				
43				
43 44				
43 44 45				
43 44 45 46				
43 44 45 46 47				
43 44 45 46 47 48				
43 44 45 46 47 48 49				
43 44 45 46 47 48 49 50				
43 44 45 46 47 48 49 50 51				
43 44 45 46 47 48 49 50 51 52				
43 44 45 46 47 48 49 50 51 52 53				
43 44 45 46 47 48 49 50 51 52 53 54				
43 44 45 46 47 48 49 50 51 52 53 54 55				
43 44 45 46 47 48 49 50 51 52 53 54 55 56				
43 44 45 46 47 48 49 50 51 52 53 54 55 56 57				
43 44 45 46 47 48 49 50 51 52 53 54 55 56 57 58				
43 44 45 46 47 48 49 50 51 52 53 54 55 56 57 58 59				
43 44 45 46 47 48 49 50 51 52 53 54 55 56 57 58	For peer review only -	http://bmjopen.bmj.com/site/	about/guidelines.xhtml	

60

-										
1										
2										
3										
4										
5										
6										
7										
8										
9	Weight	t for age Z-score at birth t for age Z-score at birth				•			•	
10		ancy duration in weeks circumference at birth			•					
11	Head o Preterr	circumference at birth m		•	•					
12		nal education in years for gestational age								
13	Parity									
14 15		er of pairs of shoes mother owns								
15	House	ehold owns cows	•							
16 17		er of saris mother owns								
17 18	Numbe	hold owns a television er of shalwar kamiz mother owns	•							
18		r is literate hold owns an electric fan	•							
20		sive breastfeeding at age 4 months nal height	•							
20	Materia	ial of floor or brother-in-law in the same household	•							
22	Arsenio	ic concentration in drinking water shold owns clock or watch					Materna	al factors		
23	Exposu	ure to lifetime physical violence	•				Home e	environment		
24	House	I breastfeeding at 4 months shold owns chair or table	• • • • • • • • • • • • • • • • • • • •				<ul> <li>Infant f</li> <li>Infectio</li> </ul>	eeding practi n	ces	
25		received semisolids at 10 months stic violence; ever been slapped	•							
23					1	1	1	1		
26			0.00	0.01	0.02	0.03	0.04	0.05	0.06	
26 27			0.00	0.01	0.02	0.03 Relative importanc	0.04	0.05	0.06	
27			0.00	0.01		0.03 Relative importanc		0.05	0.06	
					R	Relative importanc	e			
27 28	Figure 5. Conditio	onal random forest plo	t ranking	the rela	ative imp	Relative importance	∞ of 30 pr	edictors	with re	
27 28 29 30 31	Figure 5. Conditio ability to expla	ain the variation in cha	t ranking Inge in H <i>i</i>	the rela λΖ (Δ H	⊧ ative imp IAZ) fror	elative importance portance n birth to	of 30 pr 24 mor	edictors oths of a	with re ge. The	MINIMat
27 28 29 30 31 32	Figure 5. Conditio ability to expla		t ranking Inge in H <i>i</i>	the rela λΖ (Δ H	ہ ative imp IAZ) fror ر to the ۱	elative importance portance n birth to	of 30 pr 24 mor	edictors oths of a	with re ge. The	MINIMat
27 28 29 30 31 32 33	Figure 5. Conditio ability to expla	ain the variation in cha al Bangladesh. Colour	t ranking inge in HA coding ac	the rela \Z (Δ H cording stuntio	native imp AZ) from to the N ng.	oortance n birth tc WHO con	of 30 pr 24 mor	edictors oths of a	with re ge. The	MINIMat
27 28 29 30 31 32 33 34	Figure 5. Conditio ability to expla	ain the variation in cha al Bangladesh. Colour	t ranking Inge in H <i>i</i>	the rela \Z (Δ H cording stuntio	native imp AZ) from to the N ng.	oortance n birth tc WHO con	of 30 pr 24 mor	edictors oths of a	with re ge. The	MINIMat
27 28 29 30 31 32 33 34 35	Figure 5. Conditio ability to expla	ain the variation in cha al Bangladesh. Colour	t ranking inge in HA coding ac	the rela \Z (Δ H cording stuntio	native imp AZ) from to the N ng.	oortance n birth tc WHO con	of 30 pr 24 mor	edictors oths of a	with re ge. The	MINIMat
27 28 29 30 31 32 33 34 35 36	Figure 5. Conditio ability to expla	ain the variation in cha al Bangladesh. Colour	t ranking inge in HA coding ac	the rela \Z (Δ H cording stuntio	native imp AZ) from to the N ng.	oortance n birth tc WHO con	of 30 pr 24 mor	edictors oths of a	with re ge. The	MINIMat
27 28 29 30 31 32 33 34 35 36 37	Figure 5. Conditio ability to expla	ain the variation in cha al Bangladesh. Colour	t ranking inge in HA coding ac	the rela \Z (Δ H cording stuntio	native imp AZ) from to the N ng.	oortance n birth tc WHO con	of 30 pr 24 mor	edictors oths of a	with re ge. The	MINIMat
27 28 29 30 31 32 33 34 35 36 37 38	Figure 5. Conditio ability to expla	ain the variation in cha al Bangladesh. Colour	t ranking inge in HA coding ac	the rela \Z (Δ H cording stuntio	native imp AZ) from to the N ng.	oortance n birth tc WHO con	of 30 pr 24 mor	edictors oths of a	with re ge. The	MINIMat
27 28 29 30 31 32 33 34 35 36 37 38 39	Figure 5. Conditio ability to expla	ain the variation in cha al Bangladesh. Colour	t ranking inge in HA coding ac	the rela \Z (Δ H cording stuntio	native imp AZ) from to the N ng.	oortance n birth tc WHO con	of 30 pr 24 mor	edictors oths of a	with re ge. The	MINIMat
27 28 29 30 31 32 33 34 35 36 37 38 39 40	Figure 5. Conditio ability to expla	ain the variation in cha al Bangladesh. Colour	t ranking inge in HA coding ac	the rela \Z (Δ H cording stuntio	native imp AZ) from to the N ng.	oortance n birth tc WHO con	of 30 pr 24 mor	edictors oths of a	with re ge. The	MINIMat
27 28 29 30 31 32 33 34 35 36 37 38 39 40 41	Figure 5. Conditio ability to expla	ain the variation in cha al Bangladesh. Colour	t ranking inge in HA coding ac	the rela \Z (Δ H cording stuntio	native imp AZ) from to the N ng.	oortance n birth tc WHO con	of 30 pr 24 mor	edictors oths of a	with re ge. The	MINIMat
27 28 29 30 31 32 33 34 35 36 37 38 39 40 41 42	Figure 5. Conditio ability to expla	ain the variation in cha al Bangladesh. Colour	t ranking inge in HA coding ac	the rela \Z (Δ H cording stuntio	native imp AZ) from to the N ng.	oortance n birth tc WHO con	of 30 pr 24 mor	edictors oths of a	with re ge. The	MINIMat
27 28 29 30 31 32 33 34 35 36 37 38 39 40 41 42 43	Figure 5. Conditio ability to expla	ain the variation in cha al Bangladesh. Colour	t ranking inge in HA coding ac	the rela \Z (Δ H cording stuntio	native imp AZ) from to the N ng.	oortance n birth tc WHO con	of 30 pr 24 mor	edictors oths of a	with re ge. The	MINIMat
27 28 29 30 31 32 33 34 35 36 37 38 39 40 41 42 43 44	Figure 5. Conditio ability to expla	ain the variation in cha al Bangladesh. Colour	t ranking inge in HA coding ac	the rela \Z (Δ H cording stuntio	native imp AZ) from to the N ng.	oortance n birth tc WHO con	of 30 pr 24 mor	edictors oths of a	with re ge. The	MINIMat
27 28 29 30 31 32 33 34 35 36 37 38 39 40 41 42 43 44	Figure 5. Conditio ability to expla	ain the variation in cha al Bangladesh. Colour	t ranking inge in HA coding ac	the rela \Z (Δ H cording stuntio	native imp AZ) from to the N ng.	oortance n birth tc WHO con	of 30 pr 24 mor	edictors oths of a	with re ge. The	MINIMat
27 28 29 30 31 32 33 34 35 36 37 38 39 40 41 41 42 43 44 45 46	Figure 5. Conditio ability to expla	ain the variation in cha al Bangladesh. Colour	t ranking inge in HA coding ac	the rela \Z (Δ H cording stuntio	native imp AZ) from to the N ng.	oortance n birth tc WHO con	of 30 pr 24 mor	edictors oths of a	with re ge. The	MINIMat
27 28 29 30 31 32 33 34 35 36 37 38 39 40 41 41 42 43 44 45 46 47	Figure 5. Conditio ability to expla	ain the variation in cha al Bangladesh. Colour	t ranking inge in HA coding ac	the rela \Z (Δ H cording stuntio	ative imp IAZ) from I to the N ng.	oortance n birth tc WHO con	of 30 pr 24 mor	edictors oths of a	with re ge. The	MINIMat
27 28 29 30 31 32 33 34 35 36 37 38 39 40 41 42 43 44 45 46 47 48	Figure 5. Conditio ability to expla	ain the variation in cha al Bangladesh. Colour	t ranking inge in HA coding ac	the rela \Z (Δ H cording stuntio	ative imp IAZ) from I to the N ng.	oortance n birth tc WHO con	of 30 pr 24 mor	edictors oths of a	with re ge. The	MINIMat
27 28 29 30 31 32 33 34 35 36 37 38 39 40 41 42 43 44 45 46 47 48 49	Figure 5. Conditio ability to expla	ain the variation in cha al Bangladesh. Colour	t ranking inge in HA coding ac	the rela \Z (Δ H cording stuntio	ative imp IAZ) from I to the N ng.	oortance n birth tc WHO con	of 30 pr 24 mor	edictors oths of a	with re ge. The	MINIMat
27 28 29 30 31 32 33 34 35 36 37 38 39 40 41 42 43 44 45 46 47 48 49 50	Figure 5. Conditio ability to expla	ain the variation in cha al Bangladesh. Colour	t ranking inge in HA coding ac	the rela \Z (Δ H cording stuntio	ative imp IAZ) from I to the N ng.	oortance n birth tc WHO con	of 30 pr 24 mor	edictors oths of a	with re ge. The	MINIMat
27 28 29 30 31 32 33 34 35 36 37 38 39 40 41 42 43 44 45 46 47 48 49 50 51	Figure 5. Conditio ability to expla	ain the variation in cha al Bangladesh. Colour	t ranking inge in HA coding ac	the rela \Z (Δ H cording stuntio	ative imp IAZ) from I to the N ng.	oortance n birth tc WHO con	of 30 pr 24 mor	edictors oths of a	with re ge. The	MINIMat
27 28 29 30 31 32 33 34 35 36 37 38 39 40 41 42 43 44 45 46 47 48 49 50 51 52	Figure 5. Conditio ability to expla	ain the variation in cha al Bangladesh. Colour	t ranking inge in HA coding ac	the rela \Z (Δ H cording stuntio	ative imp IAZ) from I to the N ng.	oortance n birth tc WHO con	of 30 pr 24 mor	edictors oths of a	with re ge. The	MINIMat
27 28 29 30 31 32 33 34 35 36 37 38 39 40 41 41 42 43 44 45 46 47 48 49 50 51 52 53	Figure 5. Conditio ability to expla	ain the variation in cha al Bangladesh. Colour	t ranking inge in HA coding ac	the rela \Z (Δ H cording stuntio	ative imp IAZ) from I to the N ng.	oortance n birth tc WHO con	of 30 pr 24 mor	edictors oths of a	with re ge. The	MINIMat
27 28 29 30 31 32 33 34 35 36 37 38 39 40 41 42 43 44 45 46 47 48 49 50 51 52 53 54	Figure 5. Conditio ability to expla	ain the variation in cha al Bangladesh. Colour	t ranking inge in HA coding ac	the rela \Z (Δ H cording stuntio	ative imp IAZ) from I to the N ng.	oortance n birth tc WHO con	of 30 pr 24 mor	edictors oths of a	with re ge. The	MINIMat
27 28 29 30 31 32 33 34 35 36 37 38 39 40 41 42 43 44 45 46 47 48 49 50 51 52 53 54 55	Figure 5. Conditio ability to expla	ain the variation in cha al Bangladesh. Colour	t ranking inge in HA coding ac	the rela \Z (Δ H cording stuntio	ative imp IAZ) from I to the N ng.	oortance n birth tc WHO con	of 30 pr 24 mor	edictors oths of a	with re ge. The	MINIMat
27 28 29 30 31 32 33 34 35 36 37 38 39 40 41 42 43 44 45 46 47 48 49 50 51 52 53 54 55 56	Figure 5. Conditio ability to expla	ain the variation in cha al Bangladesh. Colour	t ranking inge in HA coding ac	the rela \Z (Δ H cording stuntio	ative imp IAZ) from I to the N ng.	oortance n birth tc WHO con	of 30 pr 24 mor	edictors oths of a	with re ge. The	MINIMat
27 28 29 30 31 32 33 34 35 36 37 38 39 40 41 42 43 44 45 46 47 48 49 50 51 52 53 54 55 56 57	Figure 5. Conditio ability to expla	ain the variation in cha al Bangladesh. Colour	t ranking inge in HA coding ac	the rela \Z (Δ H cording stuntio	ative imp IAZ) from I to the N ng.	oortance n birth tc WHO con	of 30 pr 24 mor	edictors oths of a	with re ge. The	MINIMat
27 28 29 30 31 32 33 34 35 36 37 38 39 40 41 42 43 44 45 46 47 48 49 50 51 52 53 54 55 56	Figure 5. Conditio ability to expla cohort in rura	ain the variation in cha al Bangladesh. Colour	t ranking nge in H/ coding ac 190x134	the rela AZ (Δ H cording stuntii mm (30	ative imp AZ) from to the M ng. 00 x 300	bortance n birth to WHO con DPI)	of 30 pr o 24 mor ceptual f	edictors hths of a framewo	with re ge. The rk on c	MINIMat



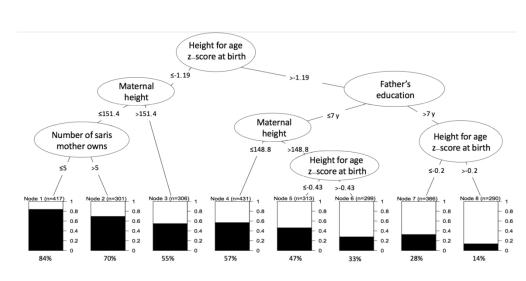


Figure 6. Conditional inference tree identifying sub-groups with different probabilities of stunting at 24 months. The MINIMat cohort in rural Bangladesh.

190x88mm (300 x 300 DPI)

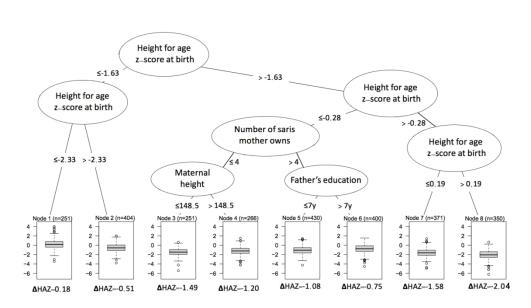


Figure 7. Conditional inference tree identifying sub-groups with different mean change in HAZ ( $\Delta$  HAZ=HAZ24-HAZ0) 0-24 months within the MINIMat cohort in rural Bangladesh.

193x101mm (300 x 300 DPI)

BMJ Open: first published as 10.1136/bmjopen-2018-025154 on 5 August 2019. Downloaded from http://bmjopen.bmj.com/ on June 12, 2025 at Agence Bibliographique de l Enseignement Superieur (ABES) . Protected by copyright, including for uses related to text and data mining, Al training, and similar technologies.

Enseignement Superieur (ABES) . Protected by copyright, including for uses related to text and data mining, Al training, and similar technologies.

## 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 ( $\alpha$ ) and deleted a proportion ( $\beta$ ) 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<sup>2</sup> and the MSE<sup>3</sup> for different imputation methods, computed from m=100 samples,  $\alpha = 0.05$ ,  $\beta = 0.05$ 

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

<sup>1</sup> K-nearest neighbour <sup>2</sup>Missclassification rate

<sup>3</sup>Mean square error

iviean square error

 $\alpha$  = proportion of the non-missing values deleted

 $\beta$  = proportion of the original data sampled

1

י ר	
2	
3	
4	
3 4 5 6 7 8 9	
6	
0	
7	
8	
9	
1	Λ
1	0 1 2
	I
1	2
1	3
1	4
1	5
1	2 3 4 5 6 7 8 9 0
	6
1	7
1	8
1	9
י ר	<u>ہ</u>
2	0
2	1
- 2	2
2	3
2	1
2	-
2	5
2	6
2	7
2	8
2	4 5 6 7 8 9
2	9
	0
3	1
3	2
2	2 3
נ ר	3 4 5 6 7 8
3	4
3	5
3	6
3	7
2	, 0
2	0
	9
	0
4	1
4	
4	-
	4
4	5
4	6
4	
	8
4	9
5	0
5	
5	
5	
5	
5	
5	
2	0
5	/
	8
5	9
	0

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

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

<sup>1</sup> K-nearest neighbour

<sup>2</sup>Missclassification rate

<sup>3</sup>Mean square error

 $\alpha$  = proportion of the non-missing values deleted

 $\beta$  = proportion of the original data sampled

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

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

<sup>1</sup> K-nearest neighbour

<sup>2</sup>Missclassification rate

<sup>3</sup>Mean square error

- $\alpha$  = proportion of the non-missing values deleted
- $\beta$  = 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 <sup>1</sup>	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

<sup>1</sup> K-nearest neighbour

<sup>2</sup>Missclassification rate

<sup>3</sup>Mean square error

 $\alpha$  = proportion of the non-missing values deleted

 $\beta$  = proportion of the original data sampled

Enseignement Superieur (ABES) . Protected by copyright, including for uses related to text and data mining, AI training, and similar technologies

#### 

# References

- 1. Stekhoven DJ, Buhlmann P. MissForest—non-parametric missing value imputation for mixed-type data. Bioinformatics. 2012;: 112–118.
- 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	( <i>a</i> ) Indicate the study's design with a commonly used term in the	1
		title or the abstract	
		(b) Provide in the abstract an informative and balanced summary	2
		of what was done and what was found	
Introduction			
Background/rationale	2	Explain the scientific background and rationale for the	4-5
-		investigation being reported	
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	6
Setting	9	periods of recruitment, exposure, follow-up, and data collection	0
Participants	6	(a) Cohort study—Give the eligibility criteria, and the sources	6
i unicipanto	Ū	and methods of selection of participants. Describe methods of	0
		follow-up	
		<i>Case-control study</i> —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	
		<i>Cross-sectional study</i> —Give the eligibility criteria, and the	
		sources and methods of selection of participants	
		(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	9 Figure
		confounders, and effect modifiers. Give diagnostic criteria, if	1
		applicable	
Data sources/ measurement	8*	For each variable of interest, give sources of data and details of	7-9
		methods of assessment (measurement). Describe comparability	
		of assessment methods if there is more than one group	
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
			applicabl
Quantitative variables	11	Explain how quantitative variables were handled in the analyses.	7-9
		If applicable, describe which groupings were chosen and why	
Statistical methods	12	( <i>a</i> ) Describe all statistical methods, including those used to	10-11
		control for confounding	
		(b) Describe any methods used to examine subgroups and	10-11
		interactions	
		(c) Explain how missing data were addressed	10-11
		(d) Cohort study—If applicable, explain how loss to follow-up	10
		(a) Conort study—It applicable, explain now loss to follow-up was addressed	10
		Case-control study—If applicable, explain how matching of	
		cases and controls was addressed	

Continued on next page10

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



Participants 13*		(a) Report numbers of individuals at each stage of study—eg numbers potentially	Fig 2
		eligible, examined for eligibility, confirmed eligible, included in the study,	12
		completing follow-up, and analysed	
		(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	Table 1,
		information on exposures and potential confounders	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	13
		time	
		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	16	(a) Give unadjusted estimates and, if applicable, confounder-adjusted estimates	14
		and their precision (eg, 95% confidence interval). Make clear which confounders	
		were adjusted for and why they were included	
		(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	17	Report other analyses done eg analyses of subgroups and interactions, and	14
		sensitivity analyses	
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	15,16,17
		imprecision. Discuss both direction and magnitude of any potential bias	
Interpretation 20	20	Give a cautious overall interpretation of results considering objectives, limitations,	18-22
		multiplicity of analyses, results from similar studies, and other relevant evidence	
Generalisability	21	Discuss the generalisability (external validity) of the study results	17
Other informatio	on		
Funding	22	Give the source of funding and the role of the funders for the present study and, if	23
0		applicable, for the original study on which the present article is based	

BMJ Open: first published as 10.1136/bmjopen-2018-025154 on 5 August 2019. Downloaded from http://bmjopen.bmj.com/ on June 12, 2025 at Agence Bibliographique de l Enseignement Superieur (ABES)

, Al training, and similar technologies.

Protected by copyright, including for uses related to text and data mining,

\*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.

For peer review only - http://bmjopen?bmj.com/site/about/guidelines.xhtml