

# BMJ Open

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 (<http://bmjopen.bmj.com>).

If you have any questions on BMJ Open's open peer review process please email [info.bmjopen@bmj.com](mailto:info.bmjopen@bmj.com)

# 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:	<i>BMJ Open</i>
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™  
Manuscripts

1  
2 **Relative importance of pre- and postnatal determinants of**  
3  
4  
5 **stunting. Data mining approaches to the Maternal and Infant**  
6  
7  
8 **Nutrition Interventions in Matlab (MINIMat) cohort,**  
9  
10  
11 **Bangladesh**  
12  
13  
14  
15  
16

17 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>,  
18  
19 Ruchira Naved<sup>4</sup>, Anisur Rahman<sup>4</sup>, Ashraful Islam Khan<sup>4</sup>, Katarina Ekholm Selling<sup>1</sup>  
20  
21

22  
23 <sup>1</sup>International Maternal and Child Health, Department of Women's and Children's Health,  
24  
25 Uppsala University, Uppsala, Sweden. <sup>2</sup>Department of Computer and Information Sciences,  
26  
27 Linköping University, Linköping, Sweden, <sup>3</sup>Department of Disease Control, London School of  
28  
29 Hygiene and Tropical Medicine, London, UK, <sup>4</sup>International Centre for Diarrhoeal Disease  
30  
31 Research, Bangladesh (iccdr,b), Dhaka, Bangladesh.  
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

## Abstract

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

**Methods** Conditional inference-tree-based methods were applied to the extensive Maternal and Infant Nutrition Interventions in Matlab (MINIMat) trial database with 309 variables of 2,723 children, their parents, and living conditions, including socioeconomic, nutritional and other biological characteristics of the parents; maternal exposure to violence; household food security; breast and complementary feeding; and measurements of morbidity of the mothers during pregnancy and repeatedly of their children up to 24 months of age. Child anthropometry was measured monthly from birth to 12 months, thereafter quarterly to 24 months.

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

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

### Strengths and limitations of this study

- Assesses the relative public health importance of pre- and post-natal risk factors.
- The extensive database with over 300 variables available for the analysis covers a wide range of pre and postnatal household, family, and environmental factors, child characteristics at birth, infant feeding, and morbidity. However, some potential determinants were not present in the database.
- Includes high-quality longitudinal data with low rates of missing data.
- Employs decision-tree-based methods that permit the inclusion of a high number of predictor variables, variables of different types and automatically discover complex interactions between predictor variables and include them in the model. They do not however, deliver *p*-values or confidence intervals to the results.

# Introduction

Linear growth is considered to be the best overall indicator of children's present and future health [1, 2] and the reduction of growth failure is one of the targets within the sustainable development agenda. Stunted growth is associated with short-term morbidity and mortality, impaired cognitive development, lower future productivity, and increased risk of adult chronic diseases [3]. In 2012, the WHO adopted a resolution on maternal and child undernutrition, targeting a reduction of stunting by 40% by 2025 [4]. Linear growth is most susceptible to environmentally modifiable factors from conception up to two years of age, i.e., the first 1000 days when most of the growth faltering takes place [5, 6]. To develop and deliver appropriate interventions, it is imperative to establish the relative importance of stunting risk factors. In addition, the sustainable development health goal has emphasized the personalized perspective under the universal coverage of health care. Identifying and targeting high-risk subgroups have thus been highlighted as one of the strategies to reach this goal.

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

1  
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

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

1 stunting at two years, and to identify risk groups with negative growth trajectories and high  
2 prevalence of stunting at two years.  
3  
4  
5

## 6 **Methods**

### 7 **Study setting, participants and study design**

8  
9  
10  
11 The MINIMat trial (Maternal and Infant Nutrition Interventions in Matlab, isrctn.org  
12 identifier: ISRCTN16581394) was carried out in Matlab, Bangladesh, a rural delta region  
13 located 57 km southeast of the capital Dhaka. In this area, a health and demographic  
14 surveillance system enables early pregnancy identification and longitudinal follow-up.  
15 Pregnant women were enrolled in the MINIMat trial and the follow-up included their  
16 offspring. MINIMat was a factorial randomized trial primarily evaluating the effect of an early  
17 invitation to prenatal food supplementation (versus usual timing) combined with multiple  
18 micronutrient supplementation (versus usual program iron-folate) to pregnant women on  
19 maternal hemoglobin, birth weight, gestational age at birth, and infant mortality [22]. Further,  
20 the participating women were randomly assigned to either counselling for exclusive  
21 breastfeeding or a different health education message of equivalent intensity [23]. The  
22 MINIMat trial recruited pregnant women from November 2001 to October 2003. When a  
23 woman reported to a community health worker that her menstruation was delayed by more  
24 than 14 days, she was offered a pregnancy test and her date for the last menstrual period  
25 (LMP) was recorded. If LMP date was missing, the gestational age assessment was based on  
26 ultrasound examination. In total, 4436 pregnant women participated, giving birth to 3625 live  
27 born infants from April 2002 to June 2004. The pregnant women were enrolled at around  
28 gestational week 8. In this analysis, the mothers and children were followed through  
29 pregnancy, birth, and up to two years of age.  
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  
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

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

## Data collection

Predictor and outcome variables are presented in Figure 1, grouped according to the WHO conceptual framework of stunting [20]. Data were collected using questionnaires, physical examinations, and laboratory analyses. At enrolment, well-trained field workers collected information on women's age, parity, marital status, educational level, occupation, maternal morbidity, socioeconomic characteristics, and household food security. Socioeconomic status was assessed based on a range of household assets, and a continuous household asset score, with a mean value of zero, was constructed based on a principal component analysis [24]. A validated household food security scale was created from eleven items with data on frequency of food purchased, cooked, borrowed or lent (food and money), and whether there was ready access to adequate meals and snacks [25]. The participating women were also asked whether they had suffered any of thirty morbidity symptoms from twelve different categories, including airway, urinary tract, fever, circulation, bowel, or pain symptoms during the last month. A sum score ranging from zero to twelve was created based on absence of symptoms or those not recorded for each category.

Home visits were followed by clinic visits at local health sub-centers. Maternal height and weight were measured at around eight weeks of gestation using a stadiometer to the nearest 0.1 cm and an electronic scale (Uniscale; SECA) with a precision of 0.10 kg. In the third trimester, paramedics interviewed the participating women in privacy regarding their experiences of domestic violence. A modified version of the WHO collaborative study

1 questionnaire was used [26,27], based on the conflict tactic scale covering physical, sexual  
2 and emotional violence and controlling behavior [28]. Household drinking water was  
3  
4 analyzed for arsenic concentration [29].  
5  
6  
7

8  
9 A birth notification system allowed birth anthropometry to be measured within 72  
10 hours. In the few cases where the newborns were reached after 72 hours, the measurements  
11 were adjusted to the time of birth using an SD score transformation, assuming that the infants  
12 remained in the same relative position in the anthropometric distribution during this period  
13 [30]. At birth, data on sex, birth weight, length, and breastfeeding practices were collected.  
14  
15 During the subsequent two-year study period, the mother-and-child pairs were visited  
16 monthly in their homes during the first year, and every three months during the second year.  
17  
18 On these occasions, data on infant feeding practices, child morbidity and anthropometry were  
19 collected. The mothers were interviewed about breastfeeding and complementary feeding  
20 practices. Breastfeeding practices were categorized into exclusive, predominant, partial, or  
21 any breastfeeding for each month from one to twelve months. The total time for exclusive,  
22 predominant, and any breastfeeding was calculated. The WHO recommendations guided the  
23 breastfeeding assessment [31] and results were validated with a stable-isotope technique.  
24  
25 The classification of exclusive breastfeeding was found to suffer from limited misclassification  
26 in both directions and to be accurate at the group level [32]. The food given to the infant was  
27 categorized into semi-solids and solids each month from one to twelve months. The data  
28 collection did not include full dietary assessments or classification of dietary diversity and  
29 meal frequency.  
30  
31  
32  
33  
34  
35  
36  
37  
38  
39  
40  
41  
42  
43  
44  
45  
46  
47  
48

49 The mothers were also asked whether the child had had any of the following  
50 symptoms during the last week; fever, cough, difficult breathing, chest in-drawing, rapid  
51 breathing, diarrhea, bloody diarrhea and the duration of these symptoms [33]. Categories  
52 were created based on whether the child had suffered from fever, respiratory symptoms,  
53  
54  
55  
56  
57  
58

1 suspected pneumonia, or diarrhea, and the sum of days with each symptom and total  
2 morbidity calculated from birth to 24 months. To reduce the risk of recall bias the mothers  
3 were visited monthly with an interview recall period of seven days for child morbidity. One  
4 week has been found to be optimal for this kind of morbidity recall assessment [34].  
5  
6  
7  
8  
9

10 Children's weight was measured by SECA beam and electronic scales (UNICEF  
11 Uniscale; SECA GmbH & Co, Hamburg, Germany) with a precision of 0.01 kg. The length at  
12 birth and up to 1.5 years was measured with a collapsible, locally manufactured length board  
13 with a precision of 0.1 cm. From 1.5 to two years, height was measured to the nearest 0.1 cm,  
14 using a freestanding stadiometer. Head and chest circumference were measured with a  
15 measuring tape. Two measurements were recorded on each occasion and the mean was  
16 calculated. The equipment was calibrated daily and refresher training on data collection  
17 methods, including the standardization of anthropometric measurements, was conducted  
18 periodically.  
19  
20  
21  
22  
23  
24  
25  
26  
27  
28  
29  
30

## 31 Outcomes

32 Height-for-age z-scores (HAZ) were calculated from the measured length and height data  
33 using the program WHOAnthro, based on the WHO growth reference for children [35].  
34  
35

36 Children with a HAZ below minus two SD-scores were classified as stunted. Two outcomes  
37 were analyzed: stunting at 24 months and the change in HAZ from birth to 24 months,  
38 referred to as  $\Delta$  HAZ.  
39  
40  
41  
42  
43  
44  
45  
46  
47

## 48 Statistical analysis

49 A database was created with 309 variables characterizing mothers and children in the  
50 MINIMat cohort from enrolment in early pregnancy up to the time when the children were 24  
51 months of age. The sub-set of records that had height measurements at birth and 24 months  
52 was selected ( $n=2\ 723$ ). The average percent of missing values among all the predictors were  
53  
54  
55  
56  
57  
58  
59

1 4 %. The highest percent missing were among maternal morbidity data during pregnancy  
2 (22%) and categorical monthly child morbidity data (ill or not), ranging from 0% to 35% with  
3  
4 the highest number of missing observations in the first months. The continuous child  
5  
6 morbidity data however (sum of days with different types illnesses), had no missing values.  
7  
8 The most important variables identified by the random forest analyses and the variables  
9  
10 included by the conditional inference trees had less than 1% missing values. The missing  
11  
12 values of the predictor variables were imputed. To find the best method to impute the missing  
13  
14 data we made a simulation study of the performance of the following imputation methods:  
15  
16 imputation by variable mean, K-nearest neighbor imputation [36], and random forest  
17  
18 imputation [37]. The design of the study followed a procedure similar to the strategy  
19  
20 described in Jonsson et al. [36], see S appendix. Accordingly, we imputed the data by use of the  
21  
22 random forest as the simulation study revealed that this method provided the most accurate  
23  
24 imputations.  
25  
26  
27  
28  
29  
30

31 Decision trees [21] are data mining methods that allow for specifying an arbitrarily  
32  
33 high number of predictor variables, handle variables of different types, automatically discover  
34  
35 complex interactions between predictor variables, and include them in the model. Traditional  
36  
37 decision trees, such as Classification and Regression Trees (CART) have been shown to be  
38  
39 biased [38]. This motivated us to select the Conditional Inference Trees (CIT) framework, a  
40  
41 method that embeds a statistical hypothesis-testing framework into a recursive partitioning  
42  
43 algorithm used for model building [38]. Conditional inference trees were used in order to  
44  
45 identify sub-groups characterized by combinations of levels of certain predictors with distinct  
46  
47 values of  $\Delta$  HAZ or prevalence of stunting at 24 months. Cross-validation, a well-established  
48  
49 model selection method that selects a tree with an optimal predictive performance for new  
50  
51 unseen data, was applied. Cross-validation splits the data set into different train and test  
52  
53 sets repeatedly, estimates the model in one set and validates the prediction on another  
54  
55  
56  
57  
58  
59  
60

1 set, followed by an aggregation of the predictions[39]. To ensure public health relevance,  
2  
3 the minimum number of observations in each terminal node (subgroup) was set to 250.  
4  
5

6  
7 Conditional random forest (CRF) analyses were performed to assess and rank the  
8 importance of predictors with regard to their ability to explain the variation of the continuous  
9 outcome of the change in HAZ from birth to 24 months and the presence of stunting at 24  
10 months of age. In conditional random forest analysis, an ensemble of conditional inference  
11 trees is created by means of drawing subsamples from the original data and fitting a unique  
12 randomized conditional inference tree to each sample. Possible predictors at each split are  
13 selected randomly from the complete set of predictors, which leads to a better predictive  
14 performance of the tree ensemble [39]. The importance of a variable is computed by  
15 comparing the predictive mean squared error (MSE) from the original data and a dataset  
16 where the corresponding variable values are specified incorrectly, which makes the variable  
17 irrelevant for the prediction. If the variable does not contribute to the prediction, the MSE is  
18 expected to be small when the values of the variable are permuted. An aggregated  
19 difference between the MSE values over the given ensemble of trees makes up the relative  
20 importance of a variable. The random forests analyses were created based on 3000 trees, and  
21 the 30 variables with the highest importance measure are presented. The exact parameters of  
22 the reported trees are shown in STable 1. The programming language R version 3.2.4 [40]  
23 and the 'party' package [41] were used for all analyses.  
24  
25  
26  
27  
28  
29  
30  
31  
32  
33  
34  
35  
36  
37  
38  
39  
40  
41  
42  
43  
44

## 45 **Patient and public involvement**

46 No participants were involved in developing the hypothesis, the specific aims or the  
47 research questions, nor were they involved in developing plans for design or  
48 implementation of the study. No participants were involved in the interpretation of study  
49 results or write up of the manuscript. There are no plans to disseminate the results of the  
50 research to study participants.  
51  
52  
53  
54  
55  
56  
57  
58  
59  
60

## 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.

Characteristics	<i>n/n (%)</i>	Stunted at 24 months <i>n/n (%)</i>	$\Delta$ HAZ 0-24 months
<b>Mother's age (years)</b>			
<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
<b>Mother's education</b>			
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
<b>Father's education</b>			
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
<b>Parity</b>			
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
<b>Number of saris mother owns</b>			
<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
<b>Child at birth</b>			
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.

## 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



1 saris), was 84%, while a child of better birth length with an educated father only had a  
2 probability of 14% to be stunted at two years.  
3  
4

5  
6  
7 The extensive database that was available for our analysis covered a wide range of  
8 household, family, and environmental factors, child characteristics at birth, feeding, and  
9 morbidity. Infant and young child growth was carefully assessed from birth up to two years.  
10  
11 The MINIMat cohort was implemented in an excellent research infrastructure that fulfills the  
12 prerequisites for obtaining high-quality longitudinal data. Experienced field workers and  
13 study nurses collected data on the 309 variables during pregnancy and the following two  
14 years. They received repeated training, including standardization exercises, and were  
15 supervised by senior medical doctors.  
16  
17  
18  
19  
20  
21  
22  
23

24  
25 Some potential determinants were not present in the database. Household water,  
26 sanitation, and hygiene (WASH) characteristics have been shown to be associated with the  
27 risk of growth restriction by increasing the risk of infections, primarily diarrheal diseases  
28 [42]. WASH data in the MINIMat database were limited to information on arsenic  
29 contamination of the drinking water, but diarrhea and other morbidity information were  
30 included in our analyses. Further, the cohort did not include the collection of stools for the  
31 study of enteropathogens in the child, which may be associated with the risk of stunting [10].  
32 Paternal height, which may be related to fetal growth, was not available [43]. The mothers'  
33 smoking habits were not represented in the data, as smoking was extremely rare among  
34 women in the study area.  
35  
36  
37  
38  
39  
40  
41  
42  
43  
44  
45  
46  
47

48 There were slight differences in basic characteristics of the analyzed and non-analyzed  
49 groups. These differences had most likely no influence on the primary outcomes of this study.  
50 There were no or few missing values of the critical variables that ranked high in the random  
51 forest and defined the sub-groups in the conditional inference trees. A sub-study was carried  
52  
53  
54  
55  
56  
57  
58  
59  
60

1 out to ensure the most accurate method to impute missing data. Thus, it is also highly unlikely  
2 that missing data influenced the main findings.  
3  
4

5  
6  
7 Decision-tree-based methods permit the inclusion of a large number of predictor  
8 variables of different types. Complex dependencies between predictor and response variables  
9 may be modeled without any need to specify the form of dependence or consider issues  
10 regarding multicollinearity. Also, the methods automatically identify interactions and include  
11 these in the models. In classical regression models, the inclusion of this large number of  
12 predictor variables and their interactions is not computationally possible. A benefit of  
13 applying random forest modelling compared to using conventional models with relative risks  
14 or odds ratios is that it ranks the predictors according to how important these are for the  
15 explaining the outcome. The random forest analysis does not provide information on whether  
16 the predictors have a positive or negative relation to the outcome. The conditional inference  
17 trees, on the other hand, display precise information on the priority, size, and direction of the  
18 association of the predictors with the outcome. The risk group identification, including the  
19 prioritization and relevant cut-offs of risk factors, is of high public health relevance for the  
20 design and targeting of appropriate interventions with the most significant benefit.  
21  
22  
23  
24  
25  
26  
27  
28  
29  
30  
31  
32  
33  
34  
35  
36

37  
38 If the data contain two essential and highly correlated predictors, the conditional  
39 inference tree method may select only one of them in the analysis, although the other  
40 predictor might be as important. Further, decision trees do not deliver  $p$ -values or confidence  
41 intervals to the results. The cross-validation method, however, ensures that the selected tree  
42 is optimal. This validation method was chosen superior to other model validation methods,  
43 e.g., the training-test approach, as it uses the potential of the data to a greater extent at  
44 the cost of a greater computational burden.  
45  
46  
47  
48  
49  
50  
51  
52

53  
54 The study setting was a low socioeconomic area in rural Bangladesh, where maternal  
55 and child undernutrition in early life still is widespread. The growth trajectories of our cohort  
56  
57  
58

1 were consistent with established growth trajectories in South Asia, where children are born  
2 below the WHO growth reference and falter dramatically up to 24 months of age [5]. In South  
3 Asia, 39% or 64 million children under five years are reportedly stunted, which accounts for  
4 40% of the global burden. Sub-Saharan Africa is the region with the second highest frequency  
5 of stunting. Although these sub-continent share a similar proportion of stunted children and  
6 faltering patterns from 3 to 24 months, the sub-Saharan African children are on average born  
7 slightly bigger than children in South Asia [5]. This dissimilarity in growth patterns across the  
8 continents makes our results mainly relevant for the South Asian context.  
9

10  
11  
12  
13  
14  
15  
16  
17  
18  
19  
20 The most important predictors of stunting at 24 months were different indicators of  
21 size at birth, maternal height, asset score and maternal education. These findings are in line  
22 with a multi-country longitudinal study that found birth or enrollment weight of the infant  
23 and maternal height to have the highest cumulative odds ratios for linear growth deficit up to  
24 two years of age [10]. These results add to the growing evidence that a large part of linear  
25 growth faltering already originates in fetal life [10,44,45]. In a pooled analysis of 19 birth  
26 cohorts with longitudinal follow-up, 20% of stunting was attributable to small-for-gestational-  
27 age weight at birth [16]. That study did not include any post-natal factors in the analysis. In a  
28 study in Indonesia, neonatal length and weight were the strongest predictors of nutritional  
29 status and increases in weight and length during infancy [45]. Our study included both pre-  
30 and post-natal factors and, in contrast to most other studies, assessed not only the relative  
31 importance of different potential predictors, but also the public health importance of each  
32 element.  
33  
34  
35  
36  
37  
38  
39  
40  
41  
42  
43  
44  
45  
46  
47  
48

49 In a study with pooled data from five Demographic and Health Surveys in South Asia,  
50 maternal height and underweight, household wealth, maternal education, and minimum  
51 dietary diversity were found to be the most important factors among children aged 6–23  
52  
53  
54  
55  
56  
57  
58  
59  
60

1 months [15]. Similar results were reported from a study in India [46]. These studies were,  
2  
3  
4 however, cross-sectional, without access to birth characteristics.  
5

6  
7 Maternal height is a strong determinant of fetal growth [47] that indirectly reflect the  
8  
9 epigenetic heredity. Maternal height is directly associated with the uterine volume [48],  
10  
11 cephalo-pelvic disproportion and subsequent infant and childhood stunting, and child  
12  
13 mortality [49,50]. In a previous analysis of the MINIMat cohort, a short maternal height was  
14  
15 strongly associated with stunting all the way up to 10 years of age [50]. Thus, factors that well  
16  
17 precede pregnancy generate a vicious intergenerational cycle, where small mothers give birth  
18  
19 to small children of whom a high proportion become and remain stunted. In the conditional  
20  
21 inference trees for stunting at 24 months, children who were born with a higher HAZ but who  
22  
23 had shorter mothers were as likely to be stunted as children with lower HAZ at birth but with  
24  
25 a taller mother. This finding suggests that intergenerational improvements in height are  
26  
27 achievable and that interventions with a particular focus on adolescents and women of  
28  
29 reproductive health are needed to break the vicious intergenerational cycle.  
30  
31  
32  
33

34 A strong relationship between stunting and poverty has been reported from many low-  
35  
36 middle income settings [51]. Asset score and other socioeconomic markers, such as the  
37  
38 number of shoes and saris the mother owned, were highly ranked in the random forest  
39  
40 analysis and categorized subgroups with a higher probability of stunting and undesirable  
41  
42 linear growth trajectories. Poverty is associated with unfavorable food and sanitation  
43  
44 practices that can lead to poor nutrition and an increased occurrence of infections during  
45  
46 pregnancy, infancy, and childhood. Poverty increases the risk of maternal stress, depression  
47  
48 [52] and weak mother-to-child interaction and stimulation.  
49  
50  
51

52 The number of shoes and saris the mother owns might also be markers of the woman's  
53  
54 status in the household. During the last few decades, the importance of women's position in  
55  
56 household and society for child nutrition has been emphasized [53]. Maternal status is  
57  
58

1 associated with food allocation to mother and child, and a higher level of maternal autonomy  
2 has been associated with better child weight and lower levels of stunting [54]. The  
3 subordinate position of women in South Asia has been suggested to be a contributor to the  
4 high prevalence of child undernutrition in the region, compared to other areas with  
5 equivalent levels of economic growth and food security [53].  
6  
7  
8  
9  
10  
11

12  
13 An acknowledged way of increasing women's position is through improved education.  
14 The remarkable health achievements in Bangladesh over the past two decades can partly be  
15 attributed to the progress in access to education, especially at primary level and for girls [55]  
16 However, there is a considerable risk of not completing primary school for both girls and boys  
17 [56]. In 2013, the continuation to the last grade of primary school (5 years) was 75% [57]  
18 and, in our study, less than 50%. In the conditional decision trees models for stunting and  
19 change in HAZ, the cut-off values for paternal and maternal education in the groups with a  
20 lower prevalence of stunting and a more positive change in HAZ from birth to 24 months  
21 ranged from 6 to 8 years, furthering the importance of girls and boys not only enrolling in but  
22 also continuing at school.  
23  
24  
25  
26  
27  
28  
29  
30  
31  
32  
33  
34  
35

36 It may seem contradictory that children who were born with a very short length had  
37 the smallest change in HAZ. This finding most likely reflects a situation where linear growth  
38 had already been severely restricted in fetal life.  
39  
40  
41  
42

43 A multi-country pooled analysis of cohort studies showed that a higher cumulative  
44 burden of diarrhea increased the risk of stunting [58]. In situations, where measles still  
45 occurred, its impact on growth and mortality risks were repeatedly documented [59]. One  
46 explanation to the discrepancy between our results and previous findings could be  
47 Bangladesh's remarkable success in achieving the globally highest coverage of oral  
48 rehydration therapy in diarrhea [60], which may have reduced the impact on linear growth.  
49 Another factor is the almost universal immunization coverage [61,62] that has reduced or  
50  
51  
52  
53  
54  
55  
56  
57  
58  
59

1 partly eliminated immunization-preventable morbidity and the subsequent effect on growth.  
2  
3 Our previous publications on the MINIMat prenatal nutrition interventions' effects on child  
4 growth and mortality were not mediated through morbidity [22,63], further supporting the  
5 modest impact of child morbidity on linear growth in our sample [33]. In other settings with  
6 lower coverage of diarrhea treatment and immunization, the relative importance of these  
7 factors may be greater.  
8  
9  
10  
11  
12  
13

14  
15 Suboptimal infant and early childhood feeding practices have, in earlier studies, been  
16 reported as significant risk factors for stunting [64]. A systematic review and meta-analysis of  
17 17 trials showed an average effect of 0.5 cm in height when children 6–24 months had been  
18 randomized to appropriate complementary foods [65]. The infant feeding variables included  
19 in our analysis ranked low in the random forest analysis and did not show up in any of the  
20 conditional inference trees. In spite of the relatively few documented effects of  
21 complementary feeding programs on stunting, these interventions are often the priority in  
22 efforts to combat stunting.  
23  
24  
25  
26  
27  
28  
29  
30  
31  
32  
33

34 The nutrition interventions from pre-conception to two years of age currently  
35 recommended by the WHO include efforts to ensure exclusive breastfeeding, adequate  
36 complementary feeding, appropriate nutritional care of sick and malnourished children and  
37 proper intake of vitamin A, iron and iodine for women and children [18]. All of these, except  
38 micronutrient supplementation to pregnant women, are focused on the postnatal period from  
39 birth up to two years. Our results strengthen the evidence that the process of becoming  
40 stunted already begins in utero, as well as the importance of intergenerational effects.  
41 Although worthwhile, the present focus on postnatal interventions results in missed  
42 opportunities to intervene before or during the first nine months when the process of  
43 stunting is established.  
44  
45  
46  
47  
48  
49  
50  
51  
52  
53  
54  
55  
56  
57  
58  
59  
60

1 So, what possibilities do we have to improve the postnatal linear growth trajectories  
2 prenatally? Attained height is mainly dependent on one's genetic potential for linear growth,  
3 in turn determined by DNA sequence polymorphism [66,67] and epigenetic heredity [68], and  
4 to some extent the environment. The modulation of non-DNA sequence epigenetic heredity  
5 has been proposed to be one of the leading factors explaining variations in height and height  
6 changes over generations[68], especially in more deprived populations [69]. Postnatal  
7 interventions can influence factors in the environment that constrain the ability to increase  
8 linear growth, while prenatal interventions also have the potential to modulate the actual  
9 growth potential through an epigenetic modification that results from changes to gene  
10 expression in response to the fetal environment.  
11  
12  
13  
14  
15  
16  
17  
18  
19  
20  
21  
22  
23

24 Established prenatal nutritional interventions include balanced energy-protein  
25 supplementation, multiple micronutrient supplements, and nutritional counseling and  
26 education. Unfortunately, most studies evaluating these interventions report only birth  
27 weight, not length, which is why evidence to directly assess the effect on fetal linear growth is  
28 limited. Meta-analyses and randomized trials evaluating these interventions report their  
29 positive impact on birth weight and a reduced risk of LBW [70-77]. Effect sizes vary from  
30 increases in birth weight of 20–200g, with the smallest effects seen in studies of multiple  
31 micronutrients and bigger effects seen by balanced energy-protein and lipid-based nutrient  
32 supplements. Considerable heterogeneity in growth response is common, and is related to the  
33 mother's nutritional status when entering pregnancy and possibly also to the genetic  
34 potential to benefit. In the MINIMat food and micronutrient interventions, all women received  
35 food supplementation, but they were randomized to an early invitation to supplementation  
36 (week 9) or the usual program start of supplementation (week 20). Children of mothers who  
37 participated in food supplementation from early pregnancy (versus the usual start) had a  
38 13% reduction in stunting up to five years [63].  
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  
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

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



1 the collection, analysis, and interpretation of the data, or in the preparation, review, or  
2  
3 approval of the manuscript.  
4  
5

## 6 **Acknowledgements**

7  
8  
9  
10 We thank the participants and their families in Matlab for their continuing involvement in  
11  
12 the MINIMat trial, and the field-team members and data management staff for their  
13  
14 excellent work.  
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  
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

For peer review only

## 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 ( $\Delta$  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 ( $\Delta$  HAZ=HAZ<sub>24</sub>-HAZ<sub>0</sub>) 0–24 months within the MINIMat cohort in rural Bangladesh.

## References

- 1 Pradhan M, Sahn DE, Younger SD. Decomposing world health inequality. *Journal of Health Economics* 2003;**22**:271–93. doi:10.1016/S0167-6296(02)00123-6
- 2 Victora CG, Adair L, Fall C, *et al*. Maternal and child undernutrition: consequences for adult health and human capital. *Lancet* 2008;**371**:340–57. doi:10.1016/S0140-6736(07)61692-4
- 3 Black RE, Victora CG, Walker SP, *et al*. Maternal and child undernutrition and overweight in low-income and middle-income countries. *The Lancet* 2013.
- 4 General Assembly, UNICEF. Transforming our world: the 2030 Agenda for Sustainable Development. *United Nations* 2015;:1–35.
- 5 Victora CG, de Onis M, Hallal PC, *et al*. Worldwide timing of growth faltering: revisiting implications for interventions. *Pediatrics* 2010;**125**:e473–80. doi:10.1542/peds.2009-1519
- 6 Black RE, Allen LH, Bhutta ZA, *et al*. Maternal and child undernutrition: global and regional exposures and health consequences. *The Lancet* 2008;**371**:243–60. doi:10.1016/S0140-6736(07)61690-0
- 7 Martorell R, Ramakrishnan U, Schroeder DG, *et al*. Intrauterine growth retardation, body size, body composition and physical performance in adolescence. *Eur J Clin Nutr* 1998;**52 Suppl 1**:S43–52–discussionS52–3.
- 8 Frongillo EA Jr. *Symposium: Causes and Etiology of Stunting*. *J Nutr* 1999;**129**:529S–530S. doi:10.1016/S0140-6736(94)91337-4
- 9 Cole TJ, Parkin JM. Infection and its effect on the growth of young children: A comparison of The Gambia and Uganda. *Trans R Soc Trop Med Hyg* 1977;**71**:196–8. doi:10.1016/0035-9203(77)90005-0
- 10 Investigators M-EN. Childhood stunting in relation to the pre- and postnatal environment during the first 2 years of life: The MAL-ED longitudinal birth cohort study. *PLoS Medicine* 2017;:1–21. doi:10.1371/journal.pmed.1002408
- 11 Espo M, Kulmala T, Maleta K, *et al*. Determinants of linear growth and predictors of severe stunting during infancy in rural Malawi. *Acta Paediatrica* 2002;**91**:1364–70. doi:10.1111/j.1651-2227.2002.tb02835.x
- 12 Rahman A, Chowdhury S. Determinants of Chronic Malnutrition Among Preschool Children In Bangladesh. *J Biosoc Sci* 2006;**39**:161. doi:10.1017/S0021932006001295
- 13 MS MRAM, RN MWMP, BS AFMC. Factors Associated with Stunting Among 0-23 Months-Old Children in Rural Bangladesh. *Journal of Nutrition Education and Behavior* 2017;**49**:S21. doi:10.1016/j.jneb.2017.05.061

- 1  
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
- 14 Hong R, Banta JE, Betancourt JA. Relationship between household wealth inequality and chronic childhood under-nutrition in Bangladesh. *International Journal for Equity in Health* 2006;**5**:1–10. doi:10.1186/1475-9276-5-15
- 15 Kim R, Mejía-Guevara I, Corsi DJ, *et al.* Relative importance of 13 correlates of child stunting in South Asia: Insights from nationally representative data from Afghanistan, Bangladesh, India, Nepal, and Pakistan. *Social Science & Medicine* 2017;**187**:144–54. doi:10.1016/j.socscimed.2017.06.017
- 16 Christian P, Lee SE, Donahue Angel M, *et al.* Risk of childhood undernutrition related to small-for-gestational age and preterm birth in low- and middle-income countries. *Int J Epidemiol* 2013;**42**:1340–55. doi:10.1093/ije/dyt109
- 17 World Health Organization. WHO Guidline: Daily iron and folic acid supplementation in pregnant women. *WHO* 2012;;1–32.
- 18 World Health Organization. Essential Nutrition Actions: improving maternal, newborn, infant and young child health and nutrition. *WHO* 2013;;1–144.
- 19 Bhutta ZA, Das JK, Arjumand R, *et al.* Maternal and Child Nutrition 2 Evidence-based interventions for improvement of maternal and child nutrition: what can be done and at what cost? *The Lancet* 2013;**382**:452–77. doi:10.1016/S0140-6736(13)60996-4
- 20 Stewart CP, Iannotti L, Dewey KG, *et al.* Contextualising complementary feeding in a broader framework for stunting prevention. *Matern Child Nutr* 2013;**9**:27–45. doi:10.1111/mcn.12088
- 21 Breiman L, Friedman J, Stone CJ, *et al.* *Classification and Regression Trees*. Chapman and Hall/CRC 1984.
- 22 Persson LÅ, Arifeen S, Ekström E-C, *et al.* Effects of Prenatal Micronutrient and Early Food Supplementation on Maternal Hemoglobin, Birth Weight, and Infant Mortality Among Children in Bangladesh: The MINIMat Randomized Trial. *JAMA* 2012;**307**:2050–9. doi:10.1001/jama.2012.4061
- 23 Khan AI, Hawkesworth S, Ekström E-C, *et al.* Effects of exclusive breastfeeding intervention on child growth and body composition: the MINIMat trial, Bangladesh. *Acta Paediatrica* 2013;**102**:815–23. doi:10.1111/apa.12282
- 24 Gwatkin DR, Rustein S, Johnson K, *et al.* *Socio-economic differences in health nutrition and population in Bangladesh*. Washington, DC: HNP/Poverty Thematic Group of the World Bank 2000.
- 25 Frongillo EA, Chowdhury N, Ekström E-C, *et al.* Understanding the Experience of Household Food Insecurity in Rural Bangladesh Leads to a Measure Different from That Used in Other Countries  
. *J Nutr* 2003;;4158–62.
- 26 Straus MA, Douglas EM. A Short Form of the Revised Conflict Tactics Scales, and Typologies for Severity and Mutuality. *Violence and Victims* 2004;**19**:1–15.

- 1  
2 27 WHO. WHO Multi-country Study on Women's Health and Domestic Violence  
3 against Women. *WHO* 2005;:1-38.
- 4  
5 28 Asling-Monemi K, Naved RT, Persson LA. Violence against women and the risk of  
6 fetal and early childhood growth impairment: a cohort study in rural Bangladesh.  
7 *Arch Dis Child* 2009;**94**:775-9. doi:10.1136/adc.2008.144444
- 8  
9 29 Rahman A, Vahter M, Smith AH, *et al*. Arsenic Exposure During Pregnancy and Size  
10 at Birth: A Prospective Cohort Study in Bangladesh. *American Journal of*  
11 *Epidemiology* 2008;**169**:304-12. doi:10.1093/aje/kwn332
- 12  
13 30 Arifeen SE, Black RE, Caulfield LE, *et al*. Infant growth patterns in the slums of  
14 Dhaka in relation to birth weight, intrauterine growth retardation, and  
15 prematurity. *Am J Clin Nutr* 2000;**72**:1010-7.
- 16  
17 31 World Health Organization. Indicators for assessing breastfeeding practices. *WHO*  
18 1991;:1-16.
- 19  
20 32 Moore SE, Prentice AM, Coward WA, *et al*. Use of stable-isotope techniques to  
21 validate infant feeding practices reported by Bangladeshi women receiving  
22 breastfeeding counseling. *Am J Clin Nutr* 2007;:1075-82.
- 23  
24 33 Kallioinen M, Ekström E-C, Khan AI, *et al*. Prenatal early food and multiple  
25 micronutrient supplementation trial reduced infant mortality in Bangladesh, but  
26 did not influence morbidity. *Acta Paediatrica* 2017;**382**:427-8.  
27 doi:10.1111/apa.14009
- 28  
29 34 Arnold BF, Galiani S, Ram PK, *et al*. Optimal Recall Period for Caregiver-reported  
30 Illness in Risk Factor and Intervention Studies: A Multicountry Study. *American*  
31 *Journal of Epidemiology* 2013;**177**:361-70. doi:10.1093/aje/kws281
- 32  
33 35 Onis M de, Onyango AW, Borghi E, *et al*. Development of a WHO growth reference  
34 for school-aged children and adolescents. *Bulletin of the World Health*  
35 *Organization* 2007;**85**:660-7. doi:10.1590/S0042-96862007000900010
- 36  
37 36 Jönsson P, Wohlin C. An Evaluation of K-Nearest Neighbour Imputation Using  
38 Likert Data. *Proceedings of th International Symposium on Software Metrics*  
39 2004;:108-18.
- 40  
41 37 Stekhoven DJ, Buhlmann P. MissForest—non-parametric missing value  
42 imputation for mixed-type data. *Bioinformatics* 2012;:112-8.
- 43  
44 38 Torsten Hothorn KHAZ. ctree: Conditional Inference Trees. www.cran.r-  
45 project.org. 2016;:1-21.https://cran.r-  
46 project.org/web/packages/partykit/vignettes/ctree.pdf (accessed 24 Mar2017).
- 47  
48 39 Hastie T, Tibshirani R, Friedman J. *The Elements of Statistical Learning*. New York,  
49 NY: : Springer Science & Business Media 2013.
- 50  
51 40 R Core Team. R: A language and environment for stastical computing. 2017.  
52 httpa://www.R-project.org/
- 53  
54 41 Achim Zeileis THKH. party: A Laboratory for Recursive Partytioning. 2017;:1-18.
- 55  
56  
57  
58  
59  
60

- 1  
2 42 Cumming O, Cairncross S. Can water, sanitation and hygiene help eliminate  
3 stunting? Current evidence and policy implications. *Matern Child Nutr*  
4 2016;**12**:91–105. doi:10.1111/mcn.12258  
5
- 6 43 Wills AK, Chinchwadkar MC, Joglekar CV, *et al.* Maternal and paternal height and  
7 BMI and patterns of fetal growth: The Pune Maternal Nutrition Study. *Early*  
8 *Human Development* 2010;**86**:535–40. doi:10.1016/j.earlhumdev.2010.07.002  
9
- 10 44 Christian P, Lee SE, Donahue Angel M, *et al.* Risk of childhood undernutrition  
11 related to small-for-gestational age and preterm birth in low- and middle-income  
12 countries. *Int J Epidemiol* 2013;**42**:1340–55. doi:10.1093/ije/dyt109  
13
- 14 45 Schmidt MK, Muslimatun S, West CE, *et al.* Nutritional Status and Linear Growth of  
15 Indonesian Infants in West Java Are Determined More by Prenatal Environment  
16 than by Postnatal Factors. *J Nutr* 2002;**132**:2202–7.  
17
- 18 46 Corsi DJ, Mejía-Guevara I, Subramanian SV. Risk factors for chronic undernutrition  
19 among children in India: Estimating relative importance, population attributable  
20 risk and fractions. *Social Science & Medicine* 2016;**157**:165–85.  
21 doi:10.1016/j.socscimed.2015.11.014  
22
- 23 47 Kramer MS. The Epidemiology of Adverse Pregnancy Outcomes: An Overview. *J*  
24 *Nutr* 2003;**133**:1592S–1596S.  
25
- 26 48 A WHO Collaborative Study of Maternal Anthropometry and Pregnancy  
27 Outcomes. *International Journal of Gynecology and Obstetrics* 1996;**57**:1–15.  
28
- 29 49 Özaltın E. Association of Maternal Stature With Offspring Mortality, Underweight,  
30 and Stunting in Low- to Middle-Income Countries. *JAMA* 2010;**303**:1507–21.  
31 doi:10.1001/jama.2010.450  
32
- 33 50 Svefors P, Rahman A, Ekström E-C, *et al.* Stunted at 10 Years. Linear Growth  
34 Trajectories and Stunting from Birth to Pre-Adolescence in a Rural Bangladeshi  
35 Cohort. *PLoS ONE* 2016;**11**:e0149700–18. doi:10.1371/journal.pone.0149700  
36
- 37 51 Van de Poel E, Hosseinpoor AR, Speybroeck N, *et al.* Socioeconomic inequality in  
38 malnutrition in developing countries. *Bulletin of the World Health Organization*  
39 2008;**86**:282–91. doi:10.1590/S0042-96862008000400013  
40
- 41 52 Bradley RH, Corwyn RF. Socioeconomic Status and Child Development. *Annual*  
42 *Review of Psychology* 2002;**53**:371–99.  
43
- 44 53 Ramalingaswami V, U J, J R. The Asian enigma. 1996.  
45
- 46 54 Shroff M, Griffiths P, Adair L, *et al.* Maternal autonomy is inversely related to child  
47 stunting in Andhra Pradesh, India. *Matern Child Nutr* 2009;**5**:64–74.  
48 doi:10.1111/j.1740-8709.2008.00161.x  
49
- 50 55 Chowdhury AMR, Bhuiya A, Chowdhury ME, *et al.* The Bangladesh paradox:  
51 exceptional health achievement despite economic poverty. *The Lancet*  
52 2013;**382**:1734–45. doi:10.1016/S0140-6736(13)62148-0  
53  
54  
55  
56  
57  
58  
59

- 1  
2 56 Asadullah MN, Chaudhury N. Girls school enrolment figures are good in  
3 Bangladesh but drop out rates are worrying. *Development and Cooperation*  
4 2017;**59**:447–72. doi:10.1086/681929  
5
- 6 57 UNESCO. EFA 2015 national review: Bangladesh; 2014. unesdoc.unesco.org.  
7 <http://unesdoc.unesco.org/images/0023/002305/230507e.pdf> (accessed 13  
8 Feb2018).  
9
- 10 58 Checkley W, Buckley G, Gilman RH, *et al.* Multi-country analysis of the effects of  
11 diarrhoea on childhood stunting. *Int J Epidemiol* 2008;**37**:816–30.  
12 doi:10.1093/ije/dyn099  
13
- 14 59 Koster FT, Curlin GC, Aziz KMA, *et al.* Synergistic impact of measles and diarrhoea  
15 on nutrition and mortality in Bangladesh. *Bulletin of the World Health*  
16 *Organization* 1981;:901–8.  
17
- 18 60 Arifeen DrPH El S, MPH AC, ScD LR, *et al.* Community-based approaches and  
19 partnerships: innovations in health-service delivery in Bangladesh. *The Lancet*  
20 2013;**382**:2012–26. doi:10.1016/S0140-6736(13)62149-2  
21  
22
- 23 61 WHO, UNICEF. Bangladesh: WHO and UNICEF estimates of immunization  
24 coverage: 2016 revision. 2016;:1–30.  
25
- 26 62 Chowdhury AMR, Bhuiya A, Chowdhury ME, *et al.* The Bangladesh paradox:  
27 exceptional health achievement despite economic poverty. *The Lancet*  
28 2013;**382**:1734–45. doi:10.1016/S0140-6736(13)62148-0  
29
- 30 63 Khan AI, Kabir I, Ekstrom EC, *et al.* Effects of prenatal food and micronutrient  
31 supplementation on child growth from birth to 54 months of age: a randomized  
32 trial in Bangladesh. *Nutr J* 2011.  
33
- 34 64 Martorell R. Physical growth and development of the malnourished child:  
35 Contributions from 50 years of research at INCAP. *Food Nutr Bull* 2010;**31**:68–82.  
36  
37
- 38 65 Imdad A, Yakoob MY, Bhutta ZA. Impact of maternal education about  
39 complementary feeding and provision of complementary foods on child growth in  
40 developing countries. *BMC Public Health* 2011;**11**:S25. doi:10.1186/1471-2458-  
41 11-S3-S25  
42
- 43 66 Weedon MN, Frayling TM. Reaching new heights: insights into the genetics of  
44 human stature. *Trends in Genetics* 2008;**24**:595–603.  
45 doi:10.1016/j.tig.2008.09.006  
46
- 47 67 Silventoinen K, Sammalisto S, Perola M, *et al.* Heritability of Adult Body Height: A  
48 Comparative Study of Twin Cohorts in Eight Countries. *Twin Research*  
49 2003;**6**:399–408. doi:10.1375/twin.6.5.399  
50
- 51 68 Simeone P, Alberti S. Epigenetic heredity of human height. *Physiological Reports*  
52 2014;**2**:e12047–7. doi:10.14814/phy2.12047  
53
- 54 69 Silventoinen K, Kaprio J, Lahelma E, *et al.* Relative Effect of Genetic and  
55 Environmental Factors on Body Height: Differences Across Birth Cohorts Among  
56 Finnish Men and Women. *American Journal of Public Health* 2000;**90**:627–30.  
57  
58



- 1  
2 70 Imdad A, Bhutta ZA. Effect of balanced protein energy supplementation during  
3 pregnancy on birth outcomes. *BMC Public Health* 2011;**11**:S17. doi:10.1186/1471-  
4 2458-11-S3-S17  
5
- 6 71 Fall CHD, Fisher DJ, Osmond C, *et al.* Multiple Micronutrient Supplementation  
7 during Pregnancy in Low-Income Countries: A Meta-Analysis of Effects on Birth  
8 Size and Length of Gestation. *Food Nutr Bull* 2009;**30**:S533–46.  
9 doi:10.1177/15648265090304S408  
10
- 11 72 Ramakrishnan U, Grant FK, Goldenberg T, *et al.* Effect of Multiple Micronutrient  
12 Supplementation on Pregnancy and Infant Outcomes: A Systematic Review.  
13 *Paediatric and Perinatal Epidemiology* 2012;**26**:153–67. doi:10.1111/j.1365-  
14 3016.2012.01276.x  
15
- 16 73 West KP, Shamim AA, Mehra S, *et al.* Effect of Maternal Multiple Micronutrient vs  
17 Iron–Folic Acid Supplementation on Infant Mortality and Adverse Birth Outcomes  
18 in Rural Bangladesh: The JiVitA-3 Randomized Trial. *JAMA* 2014;**312**:2649–58.  
19 doi:10.1001/jama.2014.16819  
20
- 21 74 Johnson W, Darboe MK, Sosseh F, *et al.* Association of prenatal lipid-based  
22 nutritional supplementation with fetal growth in rural Gambia. *Matern Child Nutr*  
23 2017;**13**:e12367. doi:10.1111/mcn.12367  
24
- 25 75 Ashorn P, Alho L, Ashorn U, *et al.* Supplementation of Maternal Diets during  
26 Pregnancy and for 6 Months Postpartum and Infant Diets Thereafter with Small-  
27 Quantity Lipid-Based Nutrient Supplements Does Not Promote Child Growth by  
28 18 Months of Age in Rural Malawi: A Randomized Controlled Trial. *J Nutr*  
29 2015;**145**:1345–53. doi:10.3945/jn.114.207225  
30
- 31 76 Adu-Afarwuah S, Lartey A, Okronipa H, *et al.* Lipid-based nutrient supplement  
32 increases the birth size of infants of primiparous women in Ghana. *Am J Clin Nutr*  
33 2015;**101**:ajcn.091546–846. doi:10.3945/ajcn.114.091546  
34
- 35 77 Huybregts L, Roberfroid D, Lanou H, *et al.* Prenatal food supplementation fortified  
36 with multiple micronutrients increases birth length: a randomized controlled trial  
37 in rural Burkina Faso. *American Journal of Clinical Nutrition* 2009;**90**:1593–600.  
38 doi:10.3945/ajcn.2009.28253  
39
- 40 78 Gluckman PD, Hanson MA, Cooper C, *et al.* Effect of In Utero and Early-Life  
41 Conditions on Adult Health and Disease.  
42 <http://dxdoiorg/101056/NEJMra0708473> 2009;**359**:61–73.  
43 doi:10.1056/NEJMra0708473  
44
- 45 79 Potdar RD, Sahariah SA, Gandhi M, *et al.* Improving women's diet quality  
46 preconceptionally and during gestation: effects on birth weight and prevalence of  
47 low birth weight--a randomized controlled efficacy trial in India (Mumbai  
48 Maternal Nutrition Project). *American Journal of Clinical Nutrition*  
49 2014;**100**:1257–68. doi:10.3945/ajcn.114.084921  
50
- 51 80 The World Bank. World Development Indicators: Reproductive health. 2016;:1–6.  
52  
53  
54  
55  
56  
57  
58  
59  
60

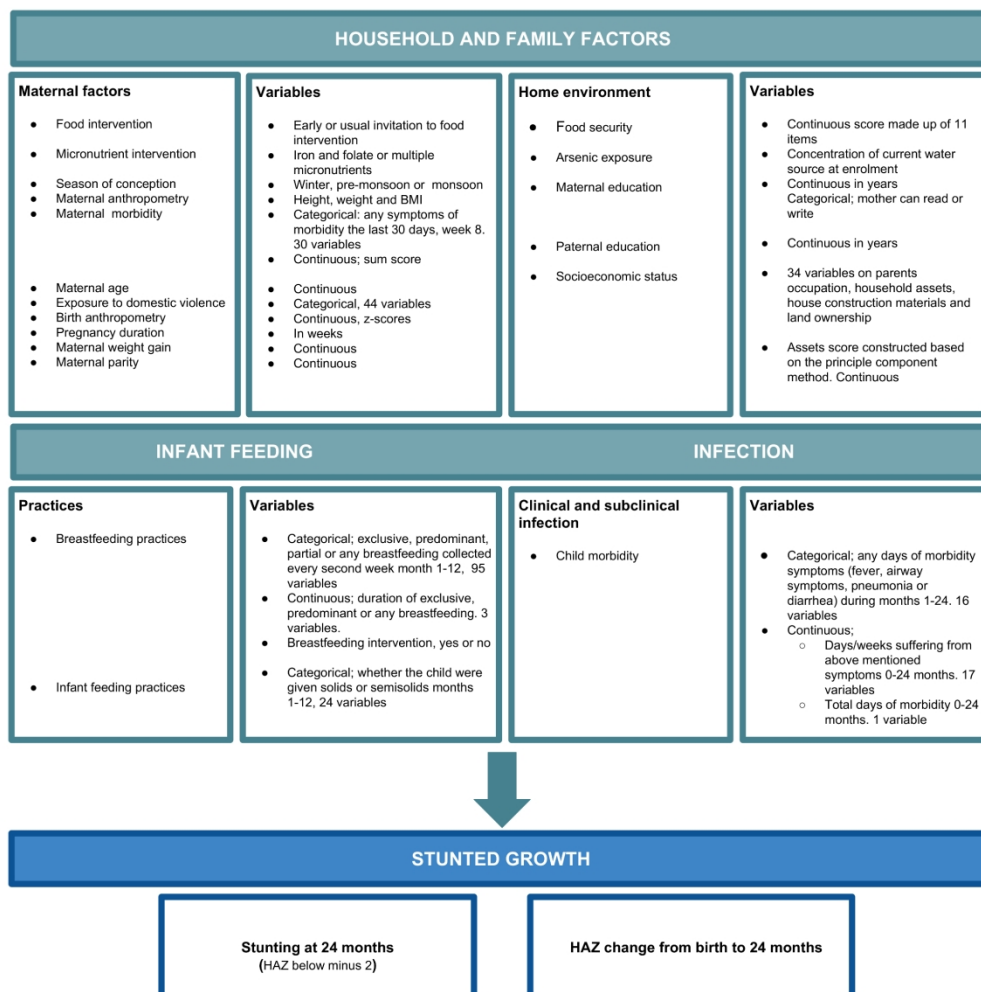


Figure 1. Factors, variables and outcomes included in the analysis of data from the MINIMat cohort, Bangladesh. Grouping according to the WHO conceptual framework on childhood stunting [20]

190x190mm (300 x 300 DPI)

1  
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

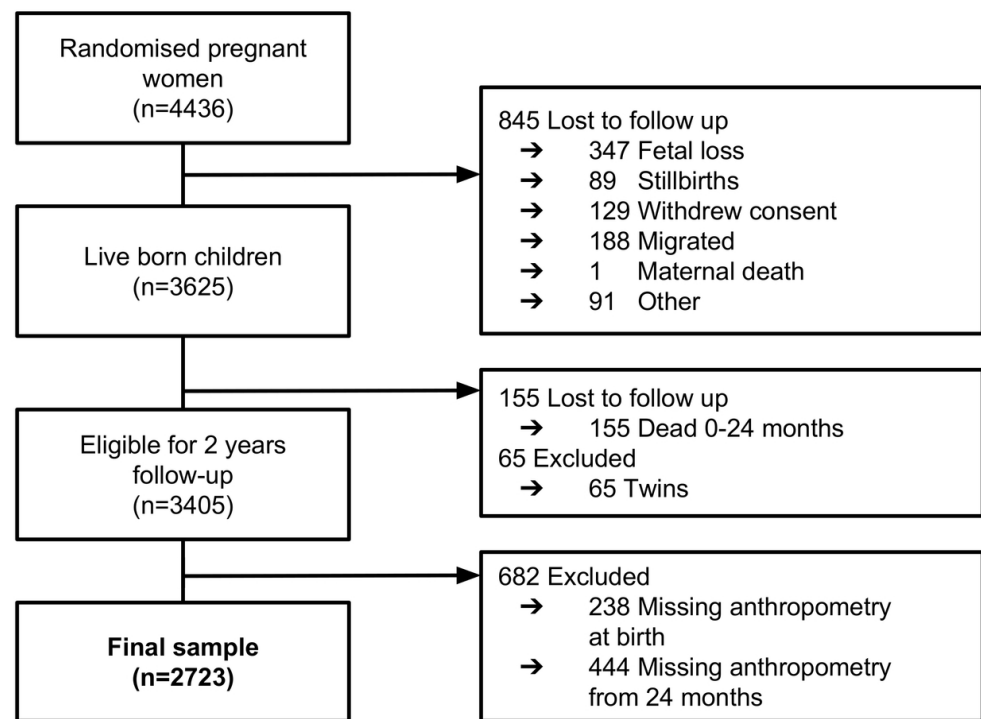


Figure 2. Flow chart of pregnant women and their children included in the data mining analyses of the MINIMat cohort from conception to two years of age.

106x80mm (300 x 300 DPI)

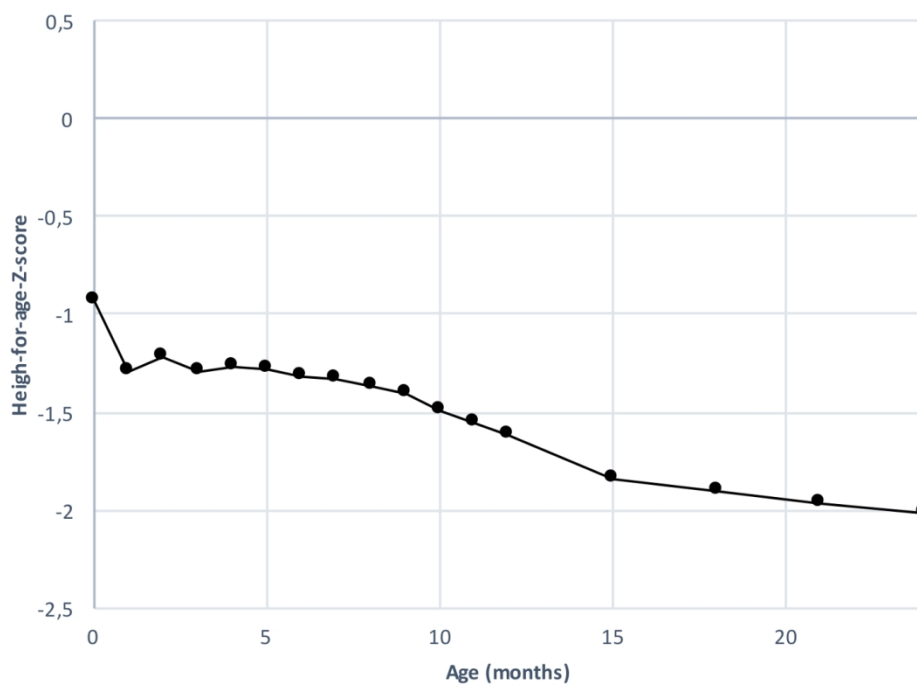


Figure 3. Height-for-age Z-scores from birth to 24 months in the MINIMat cohort in rural Bangladesh.

131x96mm (300 x 300 DPI)

1  
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

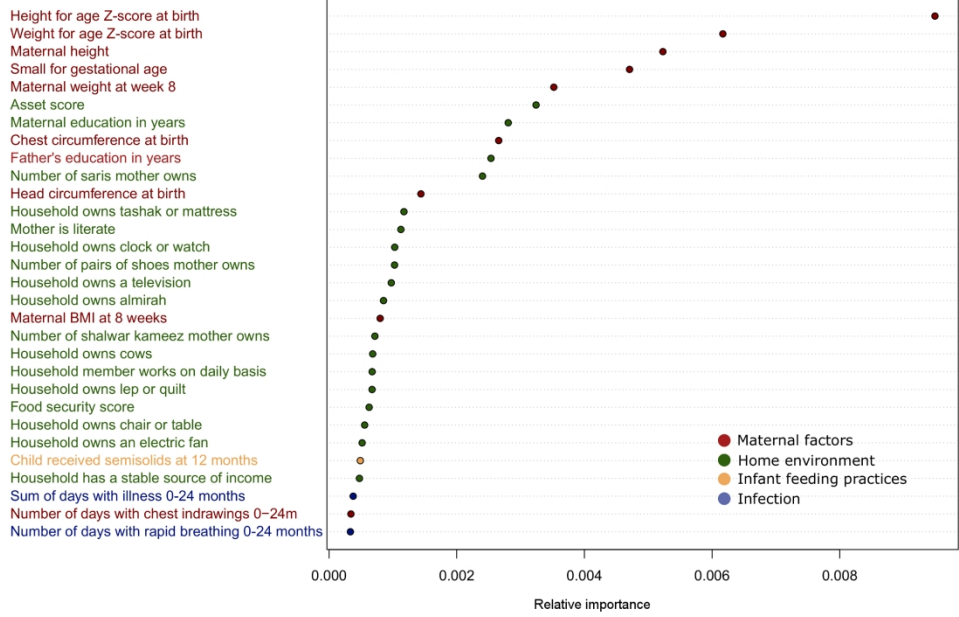


Figure 4. Conditional random forest plot ranking the relative importance of 30 predictors with regard to their ability to explain the presence of stunting at 24 months of age. The MINIMat cohort in rural Bangladesh. Colour coding according to the WHO conceptual framework on causes of stunting.

190x134mm (300 x 300 DPI)

1  
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

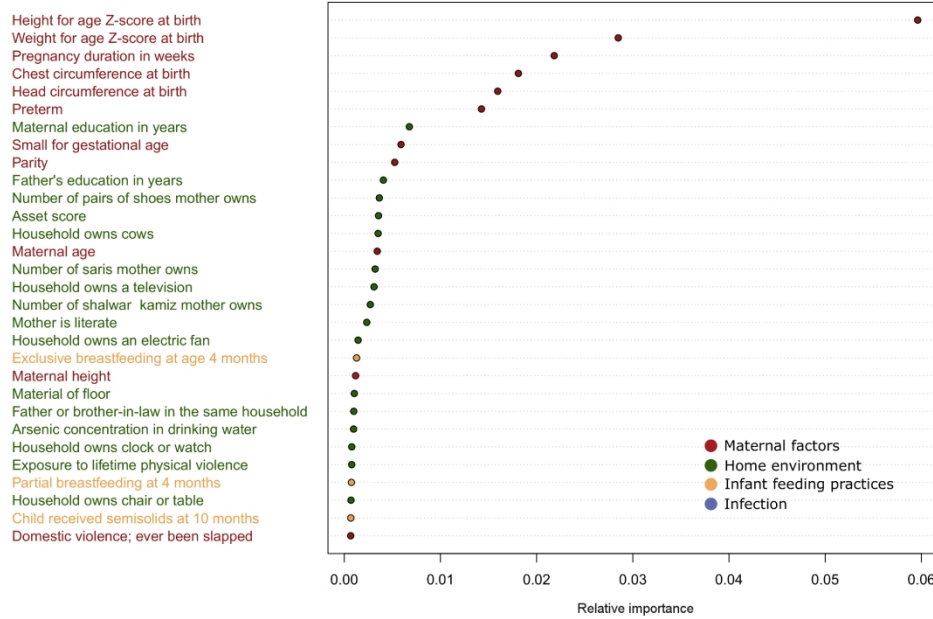


Figure 5. Conditional random forest plot ranking the relative importance of 30 predictors with regard to their ability to explain the variation in change in HAZ ( $\Delta$  HAZ) from birth to 24 months of age. The MINIMat cohort in rural Bangladesh. Colour coding according to the WHO conceptual framework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

190x134mm (300 x 300 DPI)

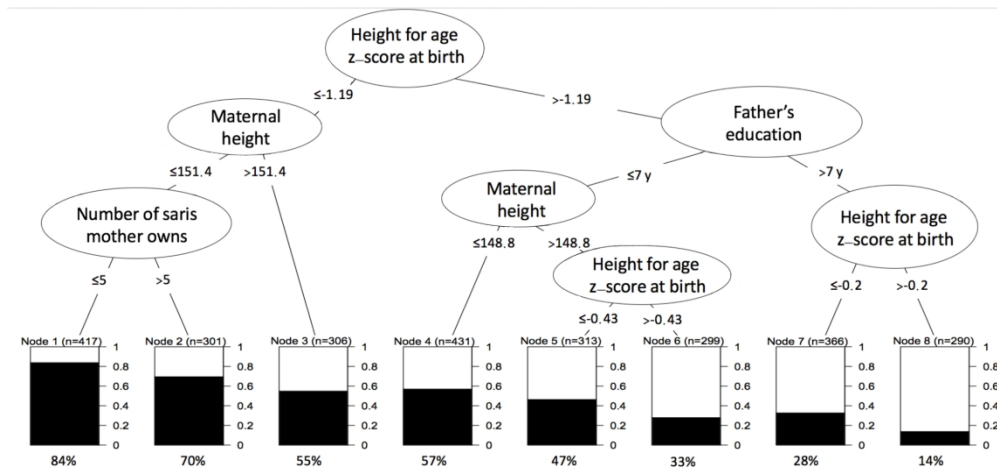


Figure 6. Conditional inference tree identifying sub-groups with different probabilities of stunting at 24 months. The MINIMat cohort in rural Bangladesh.

190x88mm (300 x 300 DPI)

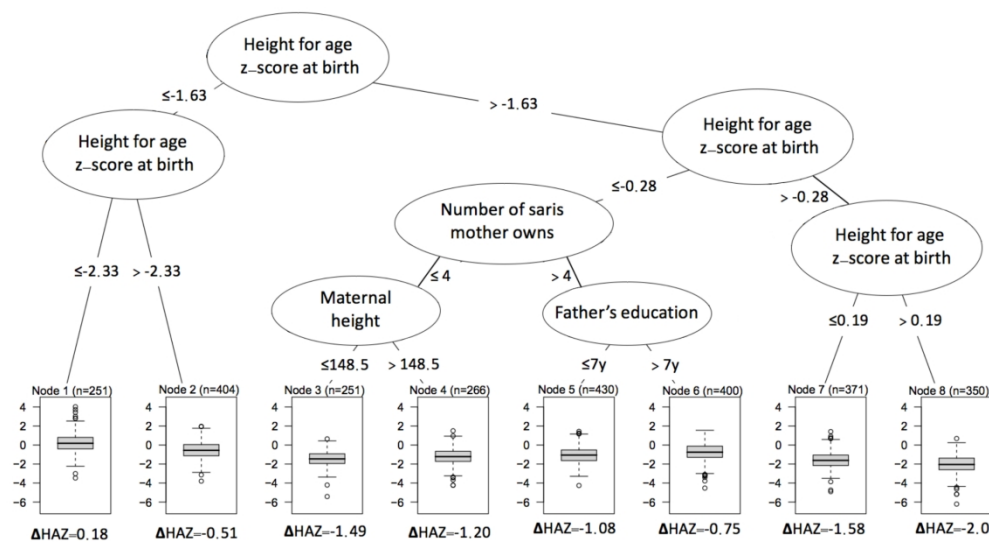


Figure 7. Conditional inference tree identifying sub-groups with different mean change in HAZ ( $\Delta$  HAZ=HAZ<sub>24</sub>-HAZ<sub>0</sub>) 0-24 months within the MINIMat cohort in rural Bangladesh.

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].

**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> Misclassification rate

<sup>3</sup> Mean square error

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

$\beta$  = proportion of the original data sampled

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

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

<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  $MR^2$  and the  $MSE^3$  for different imputation methods, computed from  $m=100$  samples,  $\alpha = 0.2$ ,  $\beta = 0.05$

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

<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, Bühlmann P. MissForest—non-parametric missing value imputation for mixed-type data. *Bioinformatics*. 2012;: 112–118.
2. Jönsson P, Wohlin C. An Evaluation of K-Nearest Neighbour Imputation Using Likert Data. *Proceedings of the 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	
<b>Title and abstract</b>	1	(a) Indicate the study's design with a commonly used term in the title or the abstract	1
		(b) Provide in the abstract an informative and balanced summary of what was done and what was found	2
<b>Introduction</b>			
Background/rationale	2	Explain the scientific background and rationale for the investigation being reported	4-5
Objectives	3	State specific objectives, including any prespecified hypotheses	6
<b>Methods</b>			
Study design	4	Present key elements of study design early in the paper	5
Setting	5	Describe the setting, locations, and relevant dates, including periods of recruitment, exposure, follow-up, and data collection	6
Participants	6	(a) <i>Cohort study</i> —Give the eligibility criteria, and the sources and methods of selection of participants. Describe methods of follow-up	6
		<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) <i>Cohort study</i> —For matched studies, give matching criteria and number of exposed and unexposed <i>Case-control study</i> —For matched studies, give matching criteria and the number of controls per case	
Variables	7	Clearly define all outcomes, exposures, predictors, potential confounders, and effect modifiers. Give diagnostic criteria, if applicable	9 Figure 1
Data sources/ measurement	8*	For each variable of interest, give sources of data and details of methods of assessment (measurement). Describe comparability of assessment methods if there is more than one group	7-9
Bias	9	Describe any efforts to address potential sources of bias	7-9
Study size	10	Explain how the study size was arrived at	Not applicable
Quantitative variables	11	Explain how quantitative variables were handled in the analyses. If applicable, describe which groupings were chosen and why	7-9
Statistical methods	12	(a) Describe all statistical methods, including those used to control for confounding	10-11
		(b) Describe any methods used to examine subgroups and interactions	10-11
		(c) Explain how missing data were addressed	10-11
		(d) <i>Cohort study</i> —If applicable, explain how loss to follow-up was addressed <i>Case-control study</i> —If applicable, explain how matching of cases and controls was addressed	10

1  
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

*Cross-sectional study*—If applicable, describe analytical methods taking account of sampling strategy

(g) Describe any sensitivity analyses

10

Continued on next page10

For peer review only

**Results**

Participants	13*	(a) Report numbers of individuals at each stage of study—eg numbers potentially eligible, examined for eligibility, confirmed eligible, included in the study, completing follow-up, and analysed	Fig 2 12
		(b) Give reasons for non-participation at each stage	Fig 2
		(c) Consider use of a flow diagram	Figure 2
Descriptive data	14*	(a) Give characteristics of study participants (eg demographic, clinical, social) and information on exposures and potential confounders	Table 1, 13
		(b) Indicate number of participants with missing data for each variable of interest	13
		(c) <i>Cohort study</i> —Summarise follow-up time (eg, average and total amount)	13
Outcome data	15*	<i>Cohort study</i> —Report numbers of outcome events or summary measures over time	13
		<i>Case-control study</i> —Report numbers in each exposure category, or summary measures of exposure	
		<i>Cross-sectional study</i> —Report numbers of outcome events or summary measures	
Main results	16	(a) Give unadjusted estimates and, if applicable, confounder-adjusted estimates and their precision (eg, 95% confidence interval). Make clear which confounders were adjusted for and why they were included	14
		(b) Report category boundaries when continuous variables were categorized	14
		(c) If relevant, consider translating estimates of relative risk into absolute risk for a meaningful time period	
Other analyses	17	Report other analyses done—eg analyses of subgroups and interactions, and sensitivity analyses	14

**Discussion**

Key results	18	Summarise key results with reference to study objectives	15,18
Limitations	19	Discuss limitations of the study, taking into account sources of potential bias or imprecision. Discuss both direction and magnitude of any potential bias	15,16,17
Interpretation	20	Give a cautious overall interpretation of results considering objectives, limitations, multiplicity of analyses, results from similar studies, and other relevant evidence	18-22
Generalisability	21	Discuss the generalisability (external validity) of the study results	17

**Other information**

Funding	22	Give the source of funding and the role of the funders for the present study and, if applicable, for the original study on which the present article is based	23
---------	----	---	----

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

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

# BMJ Open

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

Journal:	<i>BMJ Open</i>
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

SCHOLARONE™  
Manuscripts

1  
2  
3 1 **Relative importance of pre- and postnatal determinants of stunting; data mining**

4  
5  
6  
7  
8 2 **approaches to the MINIMat cohort, Bangladesh**

9  
10  
11  
12  
13 3  
14  
15  
16 4 Pernilla Svefors<sup>\*</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>,

17  
18  
19  
20 5 Ashraful Islam Khan<sup>4</sup>, Katarina Ekholm Selling<sup>†</sup>

21  
22  
23  
24 6 <sup>1</sup>International Maternal and Child Health, Department of Women's and Children's Health, Uppsala University, Uppsala, Sweden. <sup>2</sup>

25  
26  
27 7 Department of Computer and Information Sciences, Linköping University, Linköping, Sweden, <sup>3</sup>Department of Disease Control,

28  
29  
30 8 London School of Hygiene and Tropical Medicine, London, UK, <sup>4</sup> International Centre for Diarrhoeal Disease Research, Bangladesh

31  
32  
33  
34 9 (iccd,r,b), Dhaka, Bangladesh.

35  
36  
37  
38  
39 10 <sup>\*</sup>Corresponding author [pernilla.svefors@kbh.uu.se](mailto:pernilla.svefors@kbh.uu.se)



## Abstract

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

**Methods** Conditional inference-tree-based methods were applied to the extensive Maternal and Infant Nutrition Interventions in Matlab (MINIMat) trial database with 309 variables of 2,723 children, their parents, and living conditions, including socioeconomic, nutritional and other biological characteristics of the parents; maternal exposure to violence; household food security; breast and complementary feeding; and measurements of morbidity of the mothers during pregnancy and repeatedly of their children up to 24 months of age. Child anthropometry was measured monthly from birth to 12 months, thereafter quarterly to 24 months.

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

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

1  
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

28 **Strengths and limitations of this study**

- 29 • *Assesses the relative public health importance of pre- and post-natal risk factors.*
- 30 • *Includes high-quality longitudinal data with low rates of missing data on child growth and a wide range of pre and*  
31 *postnatal household, family, and environmental factors, child characteristics at birth, infant feeding, and morbidity.*
- 32 • *Some potential important determinants of linear growth were not present in the database.*
- 33 • *Employs decision-tree-based methods that permit the inclusion of a high number of predictor variables, variables of*  
34 *different types and automatically discover complex interactions between predictor variables and include them in the*  
35 *model.*

## 37 Introduction

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  
39 failure is one of the targets within the sustainable development agenda. Stunted growth is associated with short-term morbidity  
40 and mortality, impaired cognitive development, lower future productivity, and increased risk of adult chronic diseases [3]. In 2012,  
41 the WHO adopted a resolution on maternal and child undernutrition, targeting a reduction of stunting by 40% by 2025 [4]. Linear  
42 growth is most susceptible to environmentally modifiable factors from conception up to two years of age, i.e., the first 1000 days  
43 when most of the growth faltering takes place [5] [6]. To develop and deliver appropriate interventions, it is imperative to establish  
44 the relative importance of stunting risk factors. In addition, the sustainable development health goal has emphasized the  
45 personalized perspective under the universal coverage of health care. Identifying and targeting high-risk subgroups have thus been  
46 highlighted as one of the strategies to reach this goal.

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  
50 of fetal growth restriction for later stunted growth, but rarely is the relative importance of pre- and post-natal factors assessed [16].  
51 Despite these findings, policy documents and recommendations emphasize interventions especially after birth, and pre-natal  
52 recommendations are usually limited to routine micronutrient supplementation for pregnant women [17-19].

1  
2 53           Despite a wealth of literature relating to the determinants of stunting, studies with a holistic approach, which  
3  
4  
5  
6 54           concurrently account for household, environmental, nutritional, biological, and socioeconomic influences are few. Moreover,  
7  
8  
9 55           individuals and groups may be stunted for various reasons and thus respond differently to interventions. Studies that identify risk  
10  
11  
12 56           groups with different probabilities of stunting are, to the best of our knowledge, not yet available. The available studies with a  
13  
14  
15  
16 57           multifactorial approach have frequently had a cross-sectional design and have applied traditional statistical methods. As visualized  
17  
18  
19 58           in the WHO's conceptual framework for childhood stunting [20], the causes of stunted linear growth are complex. The number of  
20  
21  
22  
23 59           risk factors and the complexity of the associations of these risk factors with linear growth restriction make traditional statistical  
24  
25  
26 60           models ineffective from a predictive perspective. Moreover, classical statistical methods do not have the capacity to identify groups  
27  
28  
29  
30 61           with different risks based on combinations of predictors. Decision trees are popular data mining (DM) methods, which allows for  
31  
32  
33 62           the inclusion of a high number of predictor variables, handling variables of different types, automatically discovering complex  
34  
35  
36 63           interactions between predictor variables and including them in the model [21]. Decision-tree-based algorithms can be used to rank  
37  
38  
39  
40 64           a high number of predictors according to their relative importance for the outcome and to identify subgroups with different risk  
41  
42  
43 65           patterns.

44  
45  
46  
47 66           The Maternal and Infant Nutrition Interventions in Matlab (MINIMat) was a randomized prenatal food and multiple  
48  
49  
50  
51 67           micronutrient trial carried out in rural Bangladesh. The frequent follow-up of mothers and children participating in this trial  
52  
53  
54 68           resulted in an extensive database, including frequent pre- and post-natal anthropometric assessments, socioeconomic and  
55  
56  
57  
58 69           biological characteristics of the mother and father, information on maternal exposure to violence, household food security, breast-  
59  
60  
70           and infant-feeding practices, and measurement of morbidity of the mothers during pregnancy and repeatedly of children up to 24

1  
2 71 months of age. The aim of this study is to, within this Bangladeshi cohort, assess the relative importance of determinants of linear  
3  
4  
5  
6 72 growth from 0–24 months and risk factors for stunting at two years, and to identify risk groups with negative growth trajectories  
7  
8  
9 73 and high prevalence of stunting at two years.  
10  
11  
12

## 14 74 **Methods**

### 19 75 **Study setting, participants and study design**

22 76 The MINIMat trial (Maternal and Infant Nutrition Interventions in Matlab, isrctn.org identifier: ISRCTN16581394) was carried out  
23  
24  
25  
26 77 in Matlab, Bangladesh, a rural delta region located 57 km southeast of the capital Dhaka. In this area, a health and demographic  
27  
28  
29 78 surveillance system enables early pregnancy identification and longitudinal follow-up. Pregnant women were enrolled in the  
30  
31  
32 79 MINIMat trial and the follow-up included their offspring. MINIMat was a factorial randomized trial primarily evaluating the effect  
33  
34  
35  
36 80 of an early invitation to prenatal food supplementation (versus usual timing) combined with multiple micronutrient  
37  
38  
39 81 supplementation (versus usual program iron-folate) to pregnant women on maternal hemoglobin, birth weight, gestational age at  
40  
41  
42 82 birth, and infant mortality [22]. Further, the participating women were randomly assigned to either counselling for exclusive  
43  
44  
45  
46 83 breastfeeding or a different health education message of equivalent intensity [23]. The MINIMat trial recruited pregnant women  
47  
48  
49 84 from November 2001 to October 2003. When a woman reported to a community health worker that her menstruation was delayed  
50  
51  
52  
53 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  
54  
55  
56 86 was missing, the gestational age assessment was based on ultrasound examination. In total, 4436 pregnant women participated,  
57  
58  
59  
60

1  
2 87 giving birth to 3625 live born infants from April 2002 to June 2004. The pregnant women were enrolled at around gestational week  
3  
4  
5  
6 88 8. In this analysis, the mothers and children were followed through pregnancy, birth, and up to two years of age.  
7  
8  
9

10 89 Written and oral informed consent was obtained from all participating women and from the parents of the participating  
11  
12  
13 90 children. The Ethical Review Committee at the International Centre for Diarrhoeal Disease Research, Bangladesh, approved the  
14  
15  
16  
17 91 study (approval registration numbers 2000-025; 2002-031; 2005-004)  
18  
19  
20  
21 92

## 25 93 **Data collection**

26  
27  
28 94 Predictor and outcome variables are presented in Figure 1, grouped according to the WHO conceptual framework of stunting [20].  
29  
30  
31  
32 95 Data were collected using questionnaires, physical examinations, and laboratory analyses. At enrolment, well-trained field workers  
33  
34  
35 96 collected information on women's age, parity, marital status, educational level, occupation, maternal morbidity, socioeconomic  
36  
37  
38  
39 97 characteristics, and household food security. Socioeconomic status was assessed based on a range of household assets, and a  
40  
41  
42 98 continuous household asset score, with a mean value of zero, was constructed based on a principal component analysis [24]. A  
43  
44  
45 99 validated household food security scale was created from eleven items with data on frequency of food purchased, cooked, borrowed  
46  
47  
48  
49 100 or lent (food and money), and whether there was ready access to adequate meals and snacks [25]. The participating women were  
50  
51  
52 101 also asked whether they had suffered any of thirty morbidity symptoms from twelve different categories, including airway, urinary  
53  
54  
55  
56 102 tract, fever, circulation, bowel, or pain symptoms during the last month. A sum score ranging from zero to twelve was created based  
57  
58  
59 103 on absence of symptoms or those not recorded for each category.  
60

1  
2 104 Home visits were followed by clinic visits at local health sub-centers. Maternal height and weight were measured at  
3  
4  
5  
6 105 around eight weeks of gestation using a stadiometer to the nearest 0.1 cm and an electronic scale (Uniscale; SECA) with a precision  
7  
8  
9 106 of 0.10 kg. In the third trimester, paramedics interviewed the participating women in privacy regarding their experiences of  
10  
11  
12 107 domestic violence. A modified version of the WHO collaborative study questionnaire was used [26,27], based on the conflict tactic  
13  
14  
15  
16 108 scale covering physical, sexual and emotional violence and controlling behavior [28]. Household drinking water was analyzed for  
17  
18  
19 109 arsenic concentration [29].  
20  
21  
22

23 110 A birth notification system allowed birth anthropometry to be measured within 72 hours. In the few cases where the  
24  
25  
26  
27 111 newborns were reached after 72 hours, the measurements were adjusted to the time of birth using an SD score transformation,  
28  
29  
30 112 assuming that the infants remained in the same relative position in the anthropometric distribution during this period [30]. At  
31  
32  
33  
34 113 birth, data on sex, birth weight, length, and breastfeeding practices were collected. During the subsequent two-year study period,  
35  
36  
37 114 the mother-and-child pairs were visited monthly in their homes during the first year, and every three months during the second  
38  
39  
40  
41 115 year. On these occasions, data on infant feeding practices, child morbidity and anthropometry were collected. The mothers were  
42  
43  
44 116 interviewed about breastfeeding and complementary feeding practices. Breastfeeding practices were categorized into exclusive,  
45  
46  
47 117 predominant, partial, or any breastfeeding for each month from one to twelve months. The total time for exclusive, predominant,  
48  
49  
50  
51 118 and any breastfeeding was calculated. The WHO recommendations guided the breastfeeding assessment [31] and results were  
52  
53  
54 119 validated with a stable-isotope technique. The classification of exclusive breastfeeding was found to suffer from limited  
55  
56  
57 120 misclassification in both directions and to be accurate at the group level [32]. The food given to the infant was categorized into  
58  
59  
60

1  
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

121 semi-solids and solids each month from one to twelve months. The data collection did not include full dietary assessments or  
122 classification of dietary diversity and meal frequency.

123 The mothers were also asked whether the child had had any of the following symptoms during the last week; fever, cough,  
124 difficult breathing, chest in-drawing, rapid breathing, diarrhea, bloody diarrhea and the duration of these symptoms [33].  
125 Categories were created based on whether the child had suffered from fever, respiratory symptoms, suspected pneumonia, or  
126 diarrhea, and the sum of days with each symptom and total morbidity calculated from birth to 24 months. To reduce the risk of  
127 recall bias the mothers were visited monthly with an interview recall period of seven days for child morbidity. One week has been  
128 found to be optimal for this kind of morbidity recall assessment [34].

129 Children's weight was measured by SECA beam and electronic scales (UNICEF Uniscale; SECA GmbH & Co, Hamburg,  
130 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  
131 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  
132 stadiometer. Head and chest circumference were measured with a measuring tape. Two measurements were recorded on each  
133 occasion and the mean was calculated. The equipment was calibrated daily and refresher training on data collection methods,  
134 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 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.

1

2

3 153

Decision trees [21] are data mining methods that allow for specifying an arbitrarily high number of predictor variables,

4

5

6 154

handle variables of different types, automatically discover complex interactions between predictor variables, and include them in

7

8

9 155

the model. Traditional decision trees, such as Classification and Regression Trees (CART) have been shown to be biased [38]. This

10

11

12

13 156

motivated us to select the Conditional Inference Trees (CIT) framework, a method that embeds a statistical hypothesis-testing

14

15

16 157

framework into a recursive partitioning algorithm used for model building [38]. Conditional inference trees were used in order to

17

18

19

20 158

identify sub-groups characterized by combinations of levels of certain predictors with distinct values of  $\Delta$  HAZ or prevalence of

21

22

23 159

stunting at 24 months. Cross-validation, a well-established model selection method that selects a tree with an optimal predictive

24

25

26 160

performance for new unseen data, was applied. Cross-validation splits the data set into different train and test sets repeatedly,

27

28

29

30 161

estimates the model in one set and validates the prediction on another set, followed by an aggregation of the predictions [39]. To

31

32

33 162

ensure public health relevance, the minimum number of observations in each terminal node (subgroup) was set to 250.

34

35

36

37 163

Conditional random forest (CRF) analyses were performed to assess and rank the importance of predictors with regard to

38

39

40

41 164

their ability to explain the variation of the continuous outcome of the change in HAZ from birth to 24 months and the presence of

42

43

44 165

stunting at 24 months of age. In conditional random forest analysis, an ensemble of conditional inference trees is created by means

45

46

47

48 166

of drawing subsamples from the original data and fitting a unique randomized conditional inference tree to each sample. Possible

49

50

51 167

predictors at each split are selected randomly from the complete set of predictors, which leads to a better predictive performance of

52

53

54 168

the tree ensemble [39]. The importance of a variable is computed by comparing the predictive mean squared error (MSE) from the

55

56

57

58 169

original data and a dataset where the corresponding variable values are specified incorrectly, which makes the variable irrelevant

59

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 variable are permuted. An aggregated difference between the MSE values over the given ensemble of trees makes up the relative  
4  
5  
6 172 importance of a variable. The random forests analyses were created based on 3000 trees, and the 30 variables with the highest  
7  
8  
9 173 importance measure are presented. The programming language R version 3.2.4 [40] and the 'party' package [41] were used for all  
10  
11  
12  
13 174 analyses.

## 17 175 **Patient and public involvement**

18  
19  
20 176 No participants were involved in developing the hypothesis, the specific aims or the research questions, nor were they involved in  
21  
22  
23 177 developing plans for design or implementation of the study. No participants were involved in the interpretation of study results or  
24  
25  
26  
27 178 write up of the manuscript. There are no plans to disseminate the results of the research to study participants.  
28  
29  
30

## 31 179 32 33 34 35 180 **Results**

36  
37  
38  
39  
40  
41 181 There were 4436 women enrolled into the MINIMat trial, of whom 845 were lost to follow-up before delivery, mainly due to fetal  
42  
43  
44  
45 182 loss, outmigration, or because they withdrew their consent. Of the 3625 live born children, 155 died between birth and two years and  
46  
47  
48 183 682 were excluded because of missing anthropometry, at birth or at two years, resulting in 2723 children available for analysis  
49  
50  
51 184 (Figure 2). In the non-analyzed group there was a slightly higher percentage of mothers with more than five years of education,  
52  
53  
54  
55 185 younger than 20 years, and belonging to the lowest socioeconomic tertile, and preterm births of children (data not shown).  
56  
57  
58  
59  
60

1  
2  
3 186 The characteristics of the households, mothers, fathers at eight weeks of gestation, and children at birth are given in Table  
4  
5  
6 187 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  
7  
8  
9 188 (SD 6.8) at recruitment. One-third of the women were underweight, with a BMI below 18.5 at pregnancy week eight. The average  
10  
11  
12 189 number of years of education was similar for mothers and fathers (5 years). The sample of children comprised an equal proportion  
13  
14  
15  
16 190 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  
17  
18  
19 191 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  
20  
21  
22  
23 192 years of -2.0 (Figure 3) and 50% being stunted (girls 51.1%, boys 48.5%)  
24  
25  
26  
27 193  
28  
29  
30  
31 194 **Table 1.** Baseline characteristics, prevalence of stunting at 24 months, and mean  $\Delta$  HAZ (change in height-for-age Z-score) 0–24  
32  
33  
34 195 months in the MINIMat cohort, Bangladesh.  
35  
36  
37  
38

Characteristics	<i>n/n</i> (%)	Stunted at 24 months <i>n/n</i> (%)	Mean $\Delta$ HAZ 0-24 months
<b>Mother's age (years)</b>			
<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
<b>Mother's education</b>			
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
<b>Father's education</b>			
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
<b>Parity</b>			
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
<b>Number of saris mother owns</b>			
<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
<b>Child at birth</b>			
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.

1  
2  
3 216 **Subgroups with different levels of stunting at 24 months and levels of change in height scores from**  
4  
5 217 **birth to 24 months**

6  
7  
8 218 The conditional inference trees presented in Figure 6 and 7 display subgroups with different probability of stunting at 24 months

9  
10  
11 219 and levels of  $\Delta$  HAZ 0-24 months due to distinctive combinations of levels of certain predictors. The conditional inference trees for

12  
13  
14  
15 220 stunting and  $\Delta$ HAZ were composed of subgroups defined by the same predictors, specifically; HAZ at birth, maternal height,

16  
17  
18 221 father's educational level, and the number of saris owned by the mother. The probability of stunting ranged from 14% to 84%.

19  
20  
21 222 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

22  
23  
24  
25 223 highest probability of stunting at 24 months, at 84%. Children of a father with more than seven years of education, who had HAZ

26  
27  
28 224 at birth above -0.2, had the lowest probability of stunting at 24 months, at 14%. The difference in  $\Delta$  HAZ between the identified

29  
30  
31  
32 225 subgroups of children with the most negative change and the subgroup with the most positive change was 2.22 HAZ. Children who

33  
34  
35 226 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

36  
37  
38 227 children who were born with a HAZ above 0.19 had the most negative  $\Delta$  HAZ (-2.04 HAZ).

39  
40  
41  
42  
43 228

44  
45  
46  
47 229 **Discussion**

48  
49  
50  
51  
52  
53 230 In our analysis of 309 predictors characterizing household, environmental, biological, and socioeconomic factors, we found birth

54  
55  
56 231 size, maternal anthropometry and parental education to be the most influential for linear growth up to and stunting at 24 months.

57  
58  
59  
60 232 Conditions after birth, such as feeding practices and morbidity, were less important for linear growth trajectories and stunting at

1  
2 233 two years. The difference between the identified subgroups of children with the highest and lowest probabilities of stunting was as  
3  
4  
5  
6 234 high.

7  
8  
9  
10 235 The extensive database that was available for our analysis covered a wide range of household, family, and environmental  
11  
12  
13 236 factors, child characteristics at birth, feeding, and morbidity. Infant and young child growth was carefully assessed from birth up  
14  
15  
16  
17 237 to two years. The MINIMat cohort was implemented in an excellent research infrastructure that fulfills the prerequisites for  
18  
19  
20 238 obtaining high-quality longitudinal data. Experienced field workers and study nurses collected data on the 309 variables during  
21  
22  
23  
24 239 pregnancy and the following two years. They received repeated training, including standardization exercises, and were supervised  
25  
26  
27 240 by senior medical doctors.

28  
29  
30  
31 241 Some potential determinants were not present in the database. Household water, sanitation, and hygiene (WASH)  
32  
33  
34  
35 242 characteristics were limited to information on arsenic contamination of the drinking water, but diarrhea and other morbidity  
36  
37  
38 243 information were included in our analyses. Further, the cohort did not include the collection of stools for the study of  
39  
40  
41 244 enteropathogens in the child, which may be associated with the risk of stunting [10]. Paternal height, which may be related to fetal  
42  
43  
44  
45 245 growth, was not available [42]. The mothers' smoking habits were not represented in the data, as smoking was extremely rare  
46  
47  
48 246 among women in the study area.

49  
50  
51  
52 247 There were slight differences in basic characteristics of the analyzed and non-analyzed groups. These differences had  
53  
54  
55  
56 248 most likely no influence on the primary outcomes of this study. There were no or few missing values of the critical variables that  
57  
58  
59 249 ranked high in the random forest and defined the sub-groups in the conditional inference trees. A sub-study was carried out to  
60

1  
2  
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 findings.

7  
8  
9  
10 252 A benefit of applying random forest modelling compared to using conventional models with relative risks or odds ratios  
11  
12  
13 253 is that it ranks the predictors according to how important these are for the explaining the outcome. The random forest analysis,  
14  
15  
16  
17 254 however, does not provide information on whether the predictors have a positive or negative relation to the outcome. The  
18  
19  
20 255 conditional inference trees, on the other hand, display precise information on the priority, size, and direction of the association of  
21  
22  
23  
24 256 the predictors with the outcome. The risk group identification, including the prioritization and relevant cut-offs of risk factors, can  
25  
26  
27 257 be of high public health relevance for the design and targeting of appropriate interventions with the most significant benefit.  
28  
29  
30

31 258 A potential limitation of the conditional inference tree method is that if the data contain two essential and highly  
32  
33  
34  
35 259 correlated predictors, the conditional inference tree method may select only one of them in the analysis, although the other  
36  
37  
38 260 predictor might be as important. Further, decision trees do not deliver  $p$ -values or confidence intervals to the results. The cross-  
39  
40  
41 261 validation method, however, ensures that the selected tree is optimal. This validation method was chosen superior to other model  
42  
43  
44  
45 262 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  
46  
47  
48 263 computational burden.  
49  
50  
51

52 264 The study setting was a low socioeconomic area in rural Bangladesh, where maternal and child undernutrition in early  
53  
54  
55  
56 265 life still is widespread. The growth trajectories of our cohort were consistent with established growth trajectories in South Asia,  
57  
58  
59 266 where children are born below the WHO growth reference and falter dramatically up to 24 months of age [5]. The sub-continent of  
60



1  
2 267 South Asia and Sub-Saharan Africa share similar proportions of stunted children and faltering patterns. The sub-Saharan African  
3  
4  
5  
6 268 children are however, on average born slightly bigger than children in South Asia [5], which makes our results mainly relevant for  
7  
8  
9 269 the South Asian context.

10  
11  
12  
13 270 The most important predictors of stunting at 24 months were different indicators of size at birth, maternal height, asset  
14  
15  
16  
17 271 score and maternal education. These findings are in line with a multi-country longitudinal study that found birth or enrollment  
18  
19  
20 272 weight of the infant and maternal height to have the highest cumulative odds ratios for linear growth deficit up to two years of age  
21  
22  
23  
24 273 [10]. These results add to the growing evidence that a large part of linear growth faltering already originates in fetal life [10,43,44].  
25  
26  
27 274 In a pooled analysis of 19 birth cohorts with longitudinal follow-up, 20% of stunting was attributable to small-for-gestational-age  
28  
29  
30  
31 275 weight at birth [16]. That study did not include any post-natal factors in the analysis. In a study in Indonesia, neonatal length and  
32  
33  
34 276 weight were the strongest predictors of nutritional status and increases in weight and length during infancy [44]. Our study  
35  
36  
37 277 included both pre- and post-natal factors and, in contrast to most other studies, assessed not only the relative importance of  
38  
39  
40  
41 278 different potential predictors, but also the public health importance of each element.

42  
43  
44  
45 279 In a study with pooled data from five Demographic and Health Surveys in South Asia, maternal height and underweight,  
46  
47  
48 280 household wealth, maternal education, and minimum dietary diversity were found to be the most important factors among  
49  
50  
51  
52 281 children aged 6–23 months [15]. Similar results were reported from a study in India [45]. These studies were, however, cross-  
53  
54  
55 282 sectional, without access to birth characteristics.

1  
2 283 Maternal height is a strong determinant of fetal growth [46] that indirectly reflect the epigenetic heredity. Maternal  
3  
4  
5  
6 284 height is directly associated with the uterine volume [47], cephalo-pelvic disproportion and subsequent infant and childhood  
7  
8  
9 285 stunting, and child mortality [48,49]. In a previous analysis of the MINIMat cohort, a short maternal height was strongly associated  
10  
11  
12 286 with stunting all the way up to 10 years of age [49]. Thus, factors that well precede pregnancy generate a vicious intergenerational  
13  
14  
15  
16 287 cycle, where small mothers give birth to small children of whom a high proportion become and remain stunted. In the conditional  
17  
18  
19 288 inference trees for stunting at 24 months, children who were born with a higher HAZ but who had shorter mothers were as likely to  
20  
21  
22  
23 289 be stunted as children with lower HAZ at birth but with a taller mother. This finding suggests that intergenerational improvements  
24  
25  
26 290 in height are achievable and that interventions with a particular focus on adolescents and women of reproductive health are needed  
27  
28  
29  
30 291 to break the vicious intergenerational cycle.

31  
32  
33  
34 292 A strong relationship between stunting and poverty has been reported from many low-middle income settings [50]. Asset  
35  
36  
37 293 score and other socioeconomic markers, such as the number of shoes and saris the mother owned, were highly ranked in the  
38  
39  
40  
41 294 random forest analysis and categorized subgroups with a higher probability of stunting and undesirable linear growth trajectories.  
42  
43  
44 295 Poverty is associated with unfavorable food and sanitation practices that can lead to poor nutrition and an increased occurrence of  
45  
46  
47 296 infections during pregnancy, infancy, and childhood. Poverty increases the risk of maternal stress, depression [51] and weak  
48  
49  
50  
51 297 mother-to-child interaction and stimulation.

52  
53  
54  
55 298 The number of shoes and saris the mother owns might also be markers of the woman's status in the household. During  
56  
57  
58 299 the last few decades, the importance of women's position in household and society for child nutrition has been emphasized [52].  
59  
60

1

2 300 Maternal status is associated with food allocation to mother and child, and a higher level of maternal autonomy has been

4

5 301 associated with better child weight and lower levels of stunting [53]. The subordinate position of women in South Asia has been

7

8 302 suggested to be a contributor to the high prevalence of child undernutrition in the region, compared to other areas with equivalent

10

11 303 levels of economic growth and food security [52].

14

15

16

17 304 An acknowledged way of increasing women's position is through improved education. The remarkable health

18

19

20 305 achievements in Bangladesh over the past two decades can partly be attributed to the progress in access to education, especially at

22

23

24 306 primary level and for girls [54]. However, there is a considerable risk of not completing primary school for both girls and boys [55].

25

26

27 307 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

28

29

30 308 conditional decision trees models for stunting and change in HAZ, the cut-off values for paternal and maternal education in the

32

33

34 309 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,

35

36

37 310 furthering the importance of girls and boys not only enrolling in but also continuing at school.

38

39

40

41 311 It may seem contradictory that children who were born with a very short length had the smallest change in HAZ. This

43

44

45 312 finding most likely reflects a situation where linear growth had already been severely restricted in fetal life.

46

47

48

49 313 A multi-country pooled analysis of cohort studies showed that a higher cumulative burden of diarrhea increased the risk

50

51

52 314 of stunting [57]. In situations, where measles still occurred, its impact on growth and mortality risks were repeatedly documented

54

55

56 315 [58]. One explanation to the discrepancy between our results and previous findings could be Bangladesh's remarkable success in

57

58

59 316 achieving the globally highest coverage of oral rehydration therapy in diarrhea [59], which may have reduced the impact on linear

60

1  
2 317 growth. Another factor is the almost universal immunization coverage [60,61] that has reduced or partly eliminated immunization-  
3  
4  
5  
6 318 preventable morbidity and the subsequent effect on growth. Our previous publications on the MINIMat prenatal nutrition  
7  
8  
9 319 interventions' effects on child growth and mortality were not mediated through morbidity [22,62], further supporting the modest  
10  
11  
12  
13 320 impact of child morbidity on linear growth in our sample [33]. In other settings with lower coverage of diarrhea treatment and  
14  
15  
16 321 immunization, the relative importance of these factors may be greater.

17  
18  
19  
20 322 Suboptimal infant and early childhood feeding practices have, in earlier studies, been reported as significant risk factors  
21  
22  
23  
24 323 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–  
25  
26  
27 324 24 months had been randomized to appropriate complementary foods [64]. The infant feeding variables included in our analysis  
28  
29  
30 325 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  
31  
32  
33  
34 326 documented effects of complementary feeding programs on stunting, these interventions are often the priority in efforts to combat  
35  
36  
37 327 stunting.

38  
39  
40  
41 328 The nutrition interventions from pre-conception to two years of age currently recommended by the WHO include efforts  
42  
43  
44  
45 329 to ensure exclusive breastfeeding, adequate complementary feeding, appropriate nutritional care of sick and malnourished  
46  
47  
48 330 children and proper intake of vitamin A, iron and iodine for women and children [18]. All of these, except micronutrient  
49  
50  
51  
52 331 supplementation to pregnant women, are focused on the postnatal period from birth up to two years. Our results strengthen the  
53  
54  
55 332 evidence that the process of becoming stunted already begins in utero, as well as the importance of intergenerational effects.

56  
57  
58  
59  
60

1

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

333 Although worthwhile, the present focus on postnatal interventions results in missed opportunities to intervene before or during the  
334 first nine months when the process of stunting is established.

335 So, what possibilities do we have to improve the postnatal linear growth trajectories prenatally? Attained height is mainly  
336 dependent on one's genetic potential for linear growth, in turn determined by DNA sequence polymorphism [65,66] and epigenetic  
337 heredity [67], and to some extent the environment. The modulation of non-DNA sequence epigenetic heredity has been proposed to  
338 be one of the leading factors explaining variations in height and height changes over generations[67], especially in more deprived  
339 populations [68]. Postnatal interventions can influence factors in the environment that constrain the ability to increase linear  
340 growth, while prenatal interventions also have the potential to modulate the actual growth potential through an epigenetic  
341 modification that results from changes to gene expression in response to the fetal environment.

342 Established prenatal nutritional interventions include balanced energy-protein supplementation, multiple micronutrient  
343 supplements, and nutritional counseling and education. Unfortunately, most studies evaluating these interventions report only  
344 birth weight, not length, which is why evidence to directly assess the effect on fetal linear growth is limited. Meta-analyses and  
345 randomized trials evaluating these interventions report their positive impact on birth weight and a reduced risk of LBW [69-76].  
346 Effect sizes vary from increases in birth weight of 20–200g, with the smallest effects seen in studies of multiple micronutrients and  
347 bigger effects seen by balanced energy-protein and lipid-based nutrient supplements. Considerable heterogeneity in growth  
348 response is common, and is related to the mother's nutritional status when entering pregnancy and possibly also to the genetic  
349 potential to benefit. In the MINIMat food and micronutrient interventions, all women received food supplementation, but they were

1  
2 350 randomized to an early invitation to supplementation (week 9) or the usual program start of supplementation (week 20). Children  
3  
4  
5  
6 351 of mothers who participated in food supplementation from early pregnancy (versus the usual start) had a 13% reduction in  
7  
8  
9 352 stunting up to five years [62].  
10  
11  
12

13 353 There is increasing evidence that preconception interventions may be even more appropriate[77]. A few trials examining  
14  
15  
16  
17 354 the effect of interventions initiated before pregnancy are underway, but few results have so far been published [78]. Preconception  
18  
19  
20 355 interventions have the potential to bring about epigenetic modulation and improved growth in present and future generations.  
21  
22  
23  
24 356 Thus, the launch and evaluation of interventions targeting adolescent and women of reproductive age that focus on adequate  
25  
26  
27 357 health, education, and nutrition before and during pregnancy is needed, especially in South Asia with its high burden of maternal  
28  
29  
30 358 undernutrition and young age at first pregnancy [79]. Targeting high-risk subgroups, in this setting characterized by short, poor,  
31  
32  
33  
34 359 women with low education, can be another strategy to address the intractable problem of stunting.  
35  
36  
37

## 38 360 **Contributors**

39  
40  
41  
42  
43

44 361 PS contributed to study design, data analysis and interpretation of the results and had the main responsibility of writing the paper.  
45  
46  
47 362 LÅP and SEA were principal investigators of the MINIMat project. ECE, LÅP and KES contributed to the study design. ECE, RN, AR  
48  
49  
50  
51 363 and AIK took part in and supervised data collection. PS, OS, and KES analysed the data. All authors contributed to the preparation  
52  
53  
54 364 of the database, interpretation of the results and reviewed and approved the final version of the manuscript.  
55  
56  
57  
58

## 59 365 **Competing interests**

60

1  
2 366 The authors declare that they have no competing interests  
3  
4  
5

## 6 367 **Data sharing statement**

8 368 Data are available from the authors upon reasonable request and with permission from the principal investigator of the MINIMat  
9  
10  
11 369 study.

12  
13 370  
14  
15

## 17 371 **Funding**

18  
19  
20  
21  
22 372 The MINIMat research study was funded by the icddr,b, United Nations Children's Emergency Fund, Swedish International  
23  
24

25 373 Development Cooperation Agency (Sida), UK Medical Research Council, Swedish Research Council, Department for International  
26  
27

28  
29 374 Development (DFID), Japan Society for the Promotion of Science, Child Health and Nutrition Research Initiative, Uppsala  
30  
31

32 375 University, and US Agency for International Development. The funding agencies had no role in the design and conduct of the  
33  
34

35 376 study, in the collection, analysis, and interpretation of the data, or in the preparation, review, or approval of the manuscript.  
36  
37  
38  
39

## 40 377 **Acknowledgements**

41  
42  
43  
44  
45 378 We thank the participants and their families in Matlab for their continuing involvement in the MINIMat trial, and the field-team  
46  
47

48  
49 379 members and data management staff for their excellent work.  
50  
51  
52  
53  
54  
55  
56  
57  
58  
59  
60

1  
2  
3 381  
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

For peer review only



## 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 ( $\Delta$  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 ( $\Delta$  HAZ = HAZ<sub>24</sub> - HAZ<sub>0</sub>) 0-24 months within the MINIMat cohort in rural Bangladesh.

1  
2 401  
3  
4  
5  
6 402  
7  
8  
9  
10  
11  
12 403  
13  
14  
15  
16 404  
17  
18 405  
19  
20  
21 406  
22  
23 407  
24  
25 408  
26  
27 409  
28  
29  
30 410  
31 411  
32  
33  
34 412  
35  
36 413  
37  
38 414  
39  
40 415  
41  
42  
43 416  
44  
45 417  
46  
47 418  
48  
49 419  
50  
51  
52 420  
53 421  
54  
55  
56 422  
57  
58 423  
59  
60

## References

- 1 Pradhan M, Sahn DE, Younger SD. Decomposing world health inequality. *Journal of Health Economics* 2003;**22**:271–93. doi:10.1016/S0167-6296(02)00123-6
- 2 Victora CG, Adair L, Fall C, *et al*. Maternal and child undernutrition: consequences for adult health and human capital. *Lancet* 2008;**371**:340–57. doi:10.1016/S0140-6736(07)61692-4
- 3 Black RE, Victora CG, Walker SP, *et al*. Maternal and child undernutrition and overweight in low-income and middle-income countries. *The Lancet* 2013.
- 4 General Assembly, UNICEF. Transforming our world: the 2030 Agenda for Sustainable Development. *United Nations* 2015;;1–35.
- 5 Victora CG, de Onis M, Hallal PC, *et al*. Worldwide timing of growth faltering: revisiting implications for interventions. *Pediatrics* 2010;**125**:e473–80. doi:10.1542/peds.2009-1519
- 6 Black RE, Allen LH, Bhutta ZA, *et al*. Maternal and child undernutrition: global and regional exposures and health consequences. *The Lancet* 2008;**371**:243–60. doi:10.1016/S0140-6736(07)61690-0
- 7 Martorell R, Ramakrishnan U, Schroeder DG, *et al*. Intrauterine growth retardation, body size, body composition and physical performance in adolescence. *Eur J Clin Nutr* 1998;**52 Suppl 1**:S43–52–discussionS52–3.
- 8 Frongillo EA Jr. *Symposium: Causes and Etiology of Stunting*. *J Nutr* 1999;**129**:529S–530S. doi:10.1016/S0140-6736(94)91337-4
- 9 Cole TJ, Parkin JM. Infection and its effect on the growth of young children: A comparison of The Gambia and Uganda. *Trans R Soc Trop Med Hyg* 1977;**71**:196–8. doi:10.1016/0035-9203(77)90005-0
- 10 Investigators M-EN. Childhood stunting in relation to the pre- and postnatal environment during the first 2 years of life: The MAL-ED longitudinal birth cohort study. *PLoS Medicine* 2017;;1–21. doi:10.1371/journal.pmed.1002408

- 1  
2 424 11 Espo M, Kulmala T, Maleta K, *et al*. Determinants of linear growth and predictors of severe stunting during  
3 425 infancy in rural Malawi. *Acta Paediatrica* 2002;**91**:1364–70. doi:10.1111/j.1651-2227.2002.tb02835.x  
4  
5  
6  
7 426 12 Rahman A, Chowdhury S. Determinants of Chronic Malnutrition Among Preschool Children In Bangladesh. *J*  
8 427 *Biosoc Sci* 2006;**39**:161. doi:10.1017/S0021932006001295  
9  
10  
11 428 13 MS MRAM, RN MWMPL, BS AFMC. Factors Associated with Stunting Among 0-23 Months-Old Children in Rural  
12 429 Bangladesh. *Journal of Nutrition Education and Behavior* 2017;**49**:S21. doi:10.1016/j.jneb.2017.05.061  
13  
14  
15  
16 430 14 Hong R, Banta JE, Betancourt JA. Relationship between household wealth inequality and chronic childhood under-  
17 431 nutrition in Bangladesh. *International Journal for Equity in Health* 2006;**5**:1–10. doi:10.1186/1475-9276-5-15  
18  
19  
20 432 15 Kim R, Mejía-Guevara I, Corsi DJ, *et al*. Relative importance of 13 correlates of child stunting in South Asia: Insights  
21 433 from nationally representative data from Afghanistan, Bangladesh, India, Nepal, and Pakistan. *Social Science &*  
22 434 *Medicine* 2017;**187**:144–54. doi:10.1016/j.socscimed.2017.06.017  
23  
24  
25  
26 435 16 Christian P, Lee SE, Donahue Angel M, *et al*. Risk of childhood undernutrition related to small-for-gestational age  
27 436 and preterm birth in low- and middle-income countries. *Int J Epidemiol* 2013;**42**:1340–55. doi:10.1093/ije/dyt109  
28  
29  
30  
31 437 17 World Health Organization. WHO Guideline: Daily iron and folic acid supplementation in pregnant women. *WHO*  
32 438 2012;;1–32.  
33  
34  
35 439 18 World Health Organization. Essential Nutrition Actions: improving maternal, newborn, infant and young child  
36 440 health and nutrition. *WHO* 2013;;1–144.  
37  
38  
39 441 19 Bhutta ZA, Das JK, Arjumand R, *et al*. Maternal and Child Nutrition 2 Evidence-based interventions for  
40 442 improvement of maternal and child nutrition: what can be done and at what cost? *The Lancet* 2013;**382**:452–77.  
41 443 doi:10.1016/S0140-6736(13)60996-4  
42  
43  
44  
45 444 20 Stewart CP, Iannotti L, Dewey KG, *et al*. Contextualising complementary feeding in a broader framework for  
46 445 stunting prevention. *Matern Child Nutr* 2013;**9**:27–45. doi:10.1111/mcn.12088  
47  
48  
49  
50 446 21 Breiman L, Friedman J, Stone CJ, *et al*. *Classification and Regression Trees*. Chapman and Hall/CRC 1984.  
51  
52  
53 447 22 Persson LÅ, Arifeen S, Ekström E-C, *et al*. Effects of Prenatal Micronutrient and Early Food Supplementation on  
54 448 Maternal Hemoglobin, Birth Weight, and Infant Mortality Among Children in Bangladesh: The MINIMat  
55 449 Randomized Trial. *JAMA* 2012;**307**:2050–9. doi:10.1001/jama.2012.4061  
56  
57  
58  
59 450 23 Khan AI, Hawkesworth S, Ekström E-C, *et al*. Effects of exclusive breastfeeding intervention on child growth and  
60 451 body composition: the MINIMat trial, Bangladesh. *Acta Paediatrica* 2013;**102**:815–23. doi:10.1111/apa.12282

- 1  
2 452 24 Gwatkin DR, Rustein S, Johnson K, *et al.* *Socio-economic differences in health nutrition and population in*  
3 *Bangladesh*. Washington, DC: HNP/Poverty Thematic Group of the World Bank 2000.  
4 453  
5  
6  
7 454 25 Frongillo EA, Chowdhury N, Ekström E-C, *et al.* Understanding the Experience of Household Food Insecurity in  
8 Rural Bangladesh Leads to a Measure Different from That Used in Other Countries  
9 455  
10  
11 456 . *J Nutr* 2003;:4158–62.  
12  
13  
14 457 26 Straus MA, Douglas EM. A Short Form of the Revised Conflict Tactics Scales, and Typologies for Severity and  
15 Mutuality. *Violence and Victims* 2004;19:1–15.  
16 458  
17  
18 459 27 WHO. WHO Multi-country Study on Women's Health and Domestic Violence against Women. *WHO* 2005;:1–38.  
19  
20  
21 460 28 Asling-Monemi K, Naved RT, Persson LA. Violence against women and the risk of fetal and early childhood growth  
22 impairment: a cohort study in rural Bangladesh. *Arch Dis Child* 2009;94:775–9. doi:10.1136/adc.2008.144444  
23 461  
24  
25 462 29 Rahman A, Vahter M, Smith AH, *et al.* Arsenic Exposure During Pregnancy and Size at Birth: A Prospective  
26 Cohort Study in Bangladesh. *American Journal of Epidemiology* 2008;169:304–12. doi:10.1093/aje/kwn332  
27 463  
28  
29  
30 464 30 Arifeen SE, Black RE, Caulfield LE, *et al.* Infant growth patterns in the slums of Dhaka in relation to birth weight,  
31 intrauterine growth retardation, and prematurity. *Am J Clin Nutr* 2000;72:1010–7.  
32 465  
33  
34 466 31 World Health Organization. Indicators for assessing breastfeeding practices. *WHO* 1991;:1–16.  
35  
36  
37 467 32 Moore SE, Prentice AM, Coward WA, *et al.* Use of stable-isotope techniques to validate infant feeding practices  
38 reported by Bangladeshi women receiving breastfeeding counseling. *Am J Clin Nutr* 2007;:1075–82.  
39 468  
40  
41 469 33 Kallioinen M, Ekström E-C, Khan AI, *et al.* Prenatal early food and multiple micronutrient supplementation trial  
42 reduced infant mortality in Bangladesh, but did not influence morbidity. *Acta Paediatrica* 2017;382:427–8.  
43 470  
44 doi:10.1111/apa.14009  
45 471  
46  
47 472 34 Arnold BF, Galiani S, Ram PK, *et al.* Optimal Recall Period for Caregiver-reported Illness in Risk Factor and  
48 Intervention Studies: A Multicountry Study. *American Journal of Epidemiology* 2013;177:361–70.  
49 473  
50 doi:10.1093/aje/kws281  
51 474  
52  
53 475 35 Onis M de, Onyango AW, Borghi E, *et al.* Development of a WHO growth reference for school-aged children and  
54 adolescents. *Bulletin of the World Health Organization* 2007;85:660–7. doi:10.1590/S0042-96862007000900010  
55 476  
56  
57  
58 477 36 Jönsson P, Wohlin C. An Evaluation of K-Nearest Neighbour Imputation Using Likert Data. *Proceedings of th*  
59 *International Symposium on Software Metrics* 2004;:108–18.  
60 478

- 1  
2 479 37 Stekhoven DJ, Buhlmann P. MissForest—non-parametric missing value imputation for mixed-type data.  
3 480 *Bioinformatics* 2012;;112–8.  
4  
5  
6  
7 481 38 Torsten Hothorn KHAZ. ctree: Conditional Inference Trees. www.cran.r-project.org. 2016;;1–21. [https://cran.r-](https://cran.r-project.org/web/packages/partykit/vignettes/ctree.pdf)  
8 482 [project.org/web/packages/partykit/vignettes/ctree.pdf](https://cran.r-project.org/web/packages/partykit/vignettes/ctree.pdf) (accessed 24 Mar2017).  
9  
10  
11 483 39 Hastie T, Tibshirani R, Friedman J. *The Elements of Statistical Learning*. New York, NY: : Springer Science &  
12 484 Business Media 2013.  
13  
14  
15  
16 485 40 R Core Team. R: A language and environment for stastical computing. 2017. <http://www.R-project.org/>  
17  
18 486 41 Achim Zeileis THKH. party: A Laboratory for Recursive Partytioning. 2017;;1–18.  
19  
20  
21 487 42 Wills AK, Chinchwadkar MC, Joglekar CV, *et al*. Maternal and paternal height and BMI and patterns of fetal  
22 488 growth: The Pune Maternal Nutrition Study. *Early Human Development* 2010;**86**:535–40.  
23 489 doi:10.1016/j.earlhumdev.2010.07.002  
24  
25  
26  
27 490 43 Christian P, Lee SE, Donahue Angel M, *et al*. Risk of childhood undernutrition related to small-for-gestational age  
28 491 and preterm birth in low- and middle-income countries. *Int J Epidemiol* 2013;**42**:1340–55. doi:10.1093/ije/dyt109  
29  
30  
31  
32 492 44 Schmidt MK, Muslimatun S, West CE, *et al*. Nutritional Status and Linear Growth of Indonesian Infants in West  
33 493 Java Are Determined More by Prenatal Environment than by Postnatal Factors. *J Nutr* 2002;;2202–7.  
34  
35  
36 494 45 Corsi DJ, Mejía-Guevara I, Subramanian SV. Risk factors for chronic undernutrition among children in India:  
37 495 Estimating relative importance, population attributable risk and fractions. *Social Science & Medicine* 2016;**157**:165–  
38 496 85. doi:10.1016/j.socscimed.2015.11.014  
39  
40  
41  
42 497 46 Kramer MS. The Epidemiology of Adverse Pregnancy Outcomes: An Overview. *J Nutr* 2003;;1592S–1596S.  
43  
44  
45 498 47 A WHO Collaborative Study of Maternal Anhhropomehtry and Pregnancy Outcomes. *International Journal of*  
46 499 *Gynecology and Obstetrics* 1996;**57**:1–15.  
47  
48  
49 500 48 Özaltın E. Association of Maternal Stature With Offspring Mortality, Underweight, and Stunting in Low- to  
50 501 Middle-Income Countries. *JAMA* 2010;**303**:1507–21. doi:10.1001/jama.2010.450  
51  
52  
53  
54 502 49 Svefors P, Rahman A, Ekström E-C, *et al*. Stunted at 10 Years. Linear Growth Trajectories and Stunting from Birth  
55 503 to Pre-Adolescence in a Rural Bangladeshi Cohort. *PLoS ONE* 2016;**11**:e0149700–18.  
56 504 doi:10.1371/journal.pone.0149700  
57  
58  
59  
60 505 50 Van de Poel E, Hosseinpoor AR, Speybroeck N, *et al*. Socioeconomic inequality in malnutrition in developing  
506 countries. *Bulletin of the World Health Organization* 2008;**86**:282–91. doi:10.1590/S0042-96862008000400013

- 1  
2 507 51 Bradley RH, Corwyn RF. Socioeconomic Status and Child Development. *Annual Review of Psychology* 2002;**53**:371–  
3 508 99.  
4  
5  
6  
7 509 52 Ramalingaswami V, UJ, J R. The Asian enigma. 1996.  
8  
9  
10 510 53 Shroff M, Griffiths P, Adair L, *et al*. Maternal autonomy is inversely related to child stunting in Andhra Pradesh,  
11 511 India. *Matern Child Nutr* 2009;**5**:64–74. doi:10.1111/j.1740-8709.2008.00161.x  
12  
13  
14 512 54 Chowdhury AMR, Bhuiya A, Chowdhury ME, *et al*. The Bangladesh paradox: exceptional health achievement  
15 513 despite economic poverty. *The Lancet* 2013;**382**:1734–45. doi:10.1016/S0140-6736(13)62148-0  
16  
17  
18 514 55 Asadullah MN, Chaudhury N. Girls school enrolment figures are good in Bangladesh but drop out rates are  
19 515 worrying. *Development and Cooperation* 2017;**59**:447–72. doi:10.1086/681929  
20  
21  
22  
23 516 56 UNESCO. EFA 2015 national review: Bangladesh; 2014. unesdoc.unesco.org.  
24 517 <http://unesdoc.unesco.org/images/0023/002305/230507e.pdf> (accessed 13 Feb 2018).  
25  
26  
27 518 57 Checkley W, Buckley G, Gilman RH, *et al*. Multi-country analysis of the effects of diarrhoea on childhood stunting.  
28 519 *Int J Epidemiol* 2008;**37**:816–30. doi:10.1093/ije/dyn099  
29  
30  
31  
32 520 58 Koster FT, Curlin GC, Aziz KMA, *et al*. Synergistic impact of measles and diarrhoea on nutrition and mortality in  
33 521 Bangladesh. *Bulletin of the World Health Organization* 1981;:901–8.  
34  
35  
36 522 59 Arifeen DrPH El S, MPH AC, ScD LR, *et al*. Community-based approaches and partnerships: innovations in  
37 523 health-service delivery in Bangladesh. *The Lancet* 2013;**382**:2012–26. doi:10.1016/S0140-6736(13)62149-2  
38  
39  
40 524 60 WHO, UNICEF. Bangladesh: WHO and UNICEF estimates of immunization coverage: 2016 revision. 2016;:1–30.  
41  
42  
43 525 61 Chowdhury AMR, Bhuiya A, Chowdhury ME, *et al*. The Bangladesh paradox: exceptional health achievement  
44 526 despite economic poverty. *The Lancet* 2013;**382**:1734–45. doi:10.1016/S0140-6736(13)62148-0  
45  
46  
47 527 62 Khan AI, Kabir I, Ekstrom EC, *et al*. Effects of prenatal food and micronutrient supplementation on child growth  
48 528 from birth to 54 months of age: a randomized trial in Bangladesh. *Nutr J* 2011.  
49  
50  
51  
52 529 63 Martorell R. Physical growth and development of the malnourished child: Contributions from 50 years of research  
53 530 at INCAP. *Food Nutr Bull* 2010;**31**:68–82.  
54  
55  
56 531 64 Imdad A, Yakoob MY, Bhutta ZA. Impact of maternal education about complementary feeding and provision of  
57 532 complementary foods on child growth in developing countries. *BMC Public Health* 2011;**11**:S25. doi:10.1186/1471-  
58 533 2458-11-S3-S25  
59  
60

- 1  
2 534 65 Weedon MN, Frayling TM. Reaching new heights: insights into the genetics of human stature. *Trends in Genetics*  
3 535 2008;**24**:595–603. doi:10.1016/j.tig.2008.09.006  
4  
5  
6  
7 536 66 Silventoinen K, Sammalisto S, Perola M, *et al*. Heritability of Adult Body Height: A Comparative Study of Twin  
8 537 Cohorts in Eight Countries. *Twin Research* 2003;**6**:399–408. doi:10.1375/twin.6.5.399  
9  
10  
11 538 67 Simeone P, Alberti S. Epigenetic heredity of human height. *Physiological Reports* 2014;**2**:e12047–7.  
12 539 doi:10.14814/phy2.12047  
13  
14  
15 540 68 Silventoinen K, Kaprio J, Lahelma E, *et al*. Relative Effect of Genetic and Environmental Factors on Body Height:  
16 541 Differences Across Birth Cohorts Among Finnish Men and Women. *American Journal of Public Health*  
17 542 2000;**90**:627–30.  
18  
19  
20  
21  
22 543 69 Imdad A, Bhutta ZA. Effect of balanced protein energy supplementation during pregnancy on birth outcomes.  
23 544 *BMC Public Health* 2011;**11**:S17. doi:10.1186/1471-2458-11-S3-S17  
24  
25  
26 545 70 Fall CHD, Fisher DJ, Osmond C, *et al*. Multiple Micronutrient Supplementation during Pregnancy in Low-Income  
27 546 Countries: A Meta-Analysis of Effects on Birth Size and Length of Gestation. *Food Nutr Bull* 2009;**30**:S533–46.  
28 547 doi:10.1177/15648265090304S408  
29  
30  
31  
32 548 71 Ramakrishnan U, Grant FK, Goldenberg T, *et al*. Effect of Multiple Micronutrient Supplementation on Pregnancy  
33 549 and Infant Outcomes: A Systematic Review. *Paediatric and Perinatal Epidemiology* 2012;**26**:153–67.  
34 550 doi:10.1111/j.1365-3016.2012.01276.x  
35  
36  
37  
38 551 72 West KP, Shamim AA, Mehra S, *et al*. Effect of Maternal Multiple Micronutrient vs Iron–Folic Acid  
39 552 Supplementation on Infant Mortality and Adverse Birth Outcomes in Rural Bangladesh: The JiVitA-3 Randomized  
40 553 Trial. *JAMA* 2014;**312**:2649–58. doi:10.1001/jama.2014.16819  
41  
42  
43  
44 554 73 Johnson W, Darboe MK, Sosseh F, *et al*. Association of prenatal lipid-based nutritional supplementation with fetal  
45 555 growth in rural Gambia. *Matern Child Nutr* 2017;**13**:e12367. doi:10.1111/mcn.12367  
46  
47  
48  
49 556 74 Ashorn P, Alho L, Ashorn U, *et al*. Supplementation of Maternal Diets during Pregnancy and for 6 Months  
50 557 Postpartum and Infant Diets Thereafter with Small-Quantity Lipid-Based Nutrient Supplements Does Not  
51 558 Promote Child Growth by 18 Months of Age in Rural Malawi: A Randomized Controlled Trial. *J Nutr*  
52 559 2015;**145**:1345–53. doi:10.3945/jn.114.207225  
53  
54  
55  
56 560 75 Adu-Afarwuah S, Lartey A, Okronipa H, *et al*. Lipid-based nutrient supplement increases the birth size of infants  
57 561 of primiparous women in Ghana. *Am J Clin Nutr* 2015;**101**:ajcn.091546–846. doi:10.3945/ajcn.114.091546  
58  
59  
60

- 1  
2 562 76 Huybregts L, Roberfroid D, Lanou H, *et al.* Prenatal food supplementation fortified with multiple micronutrients  
3 increases birth length: a randomized controlled trial in rural Burkina Faso. *American Journal of Clinical Nutrition*  
4 563 2009;**90**:1593–600. doi:10.3945/ajcn.2009.28253  
5  
6 564  
7  
8 565 77 Gluckman PD, Hanson MA, Cooper C, *et al.* Effect of In Utero and Early-Life Conditions on Adult Health and  
9 Disease. <http://dxdoiorg/101056/NEJMrao708473> 2009;**359**:61–73. doi:10.1056/NEJMrao708473  
10 566  
11  
12  
13 567 78 Potdar RD, Sahariah SA, Gandhi M, *et al.* Improving women's diet quality preconceptionally and during gestation:  
14 effects on birth weight and prevalence of low birth weight--a randomized controlled efficacy trial in India (Mumbai  
15 568 Maternal Nutrition Project). *American Journal of Clinical Nutrition* 2014;**100**:1257–68. doi:10.3945/ajcn.114.084921  
16 569  
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



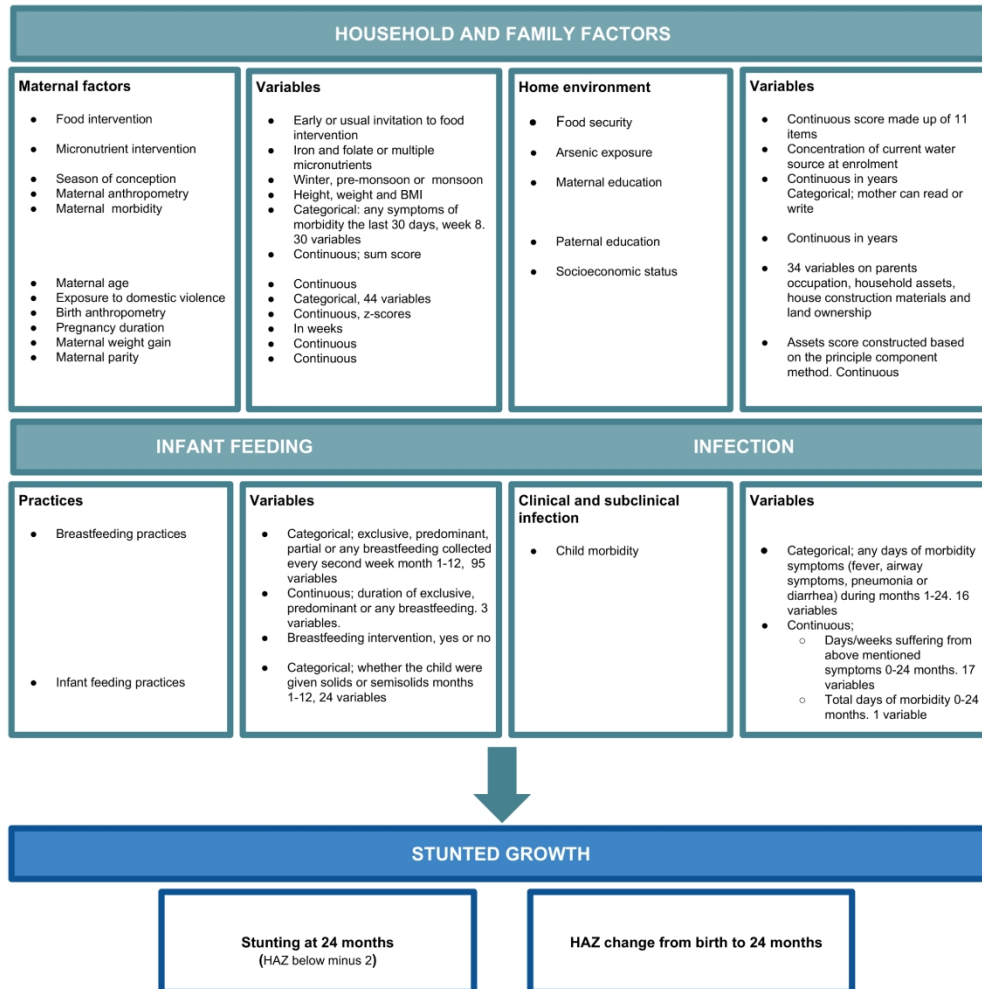
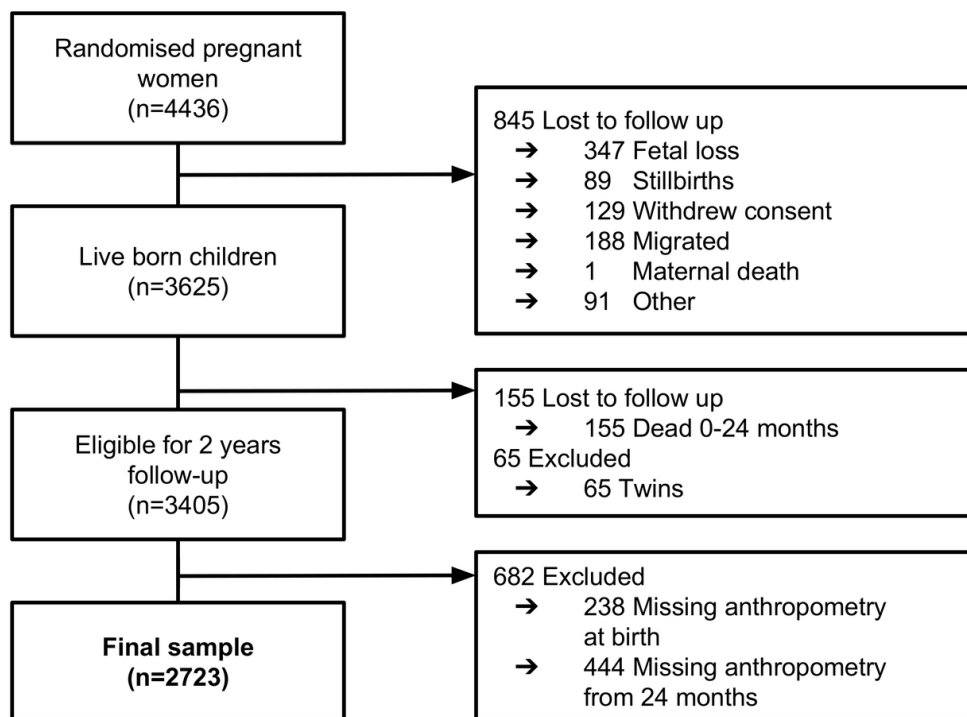


Figure 1. Factors, variables and outcomes included in the analysis of data from the MINIMat cohort, Bangladesh. Grouping according to the WHO conceptual framework on childhood stunting [20]

190x190mm (300 x 300 DPI)



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

33 106x80mm (300 x 300 DPI)

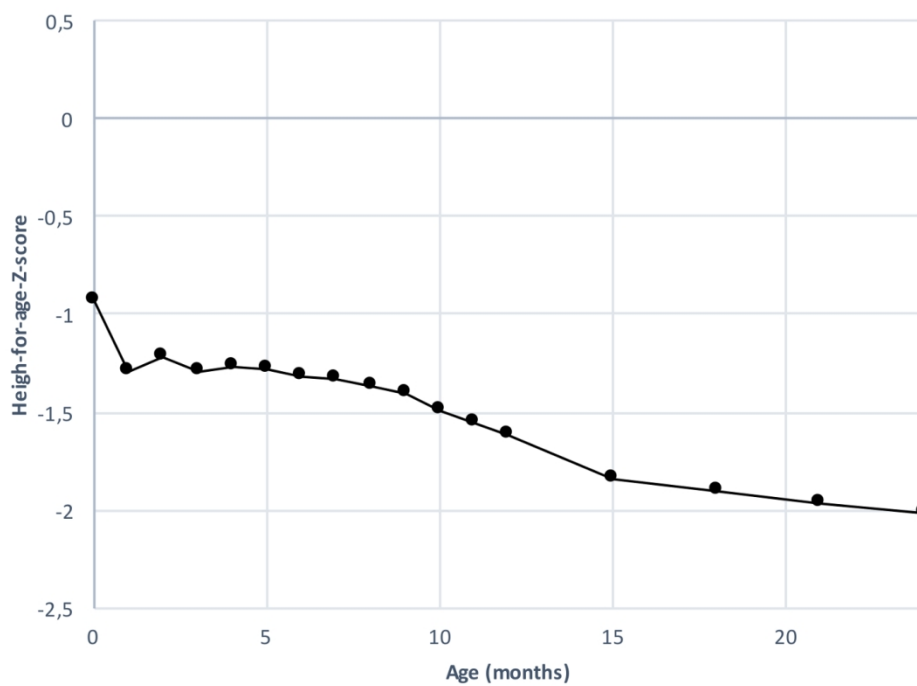


Figure 3. Height-for-age Z-scores from birth to 24 months in the MINIMat cohort in rural Bangladesh.

131x96mm (300 x 300 DPI)

1  
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

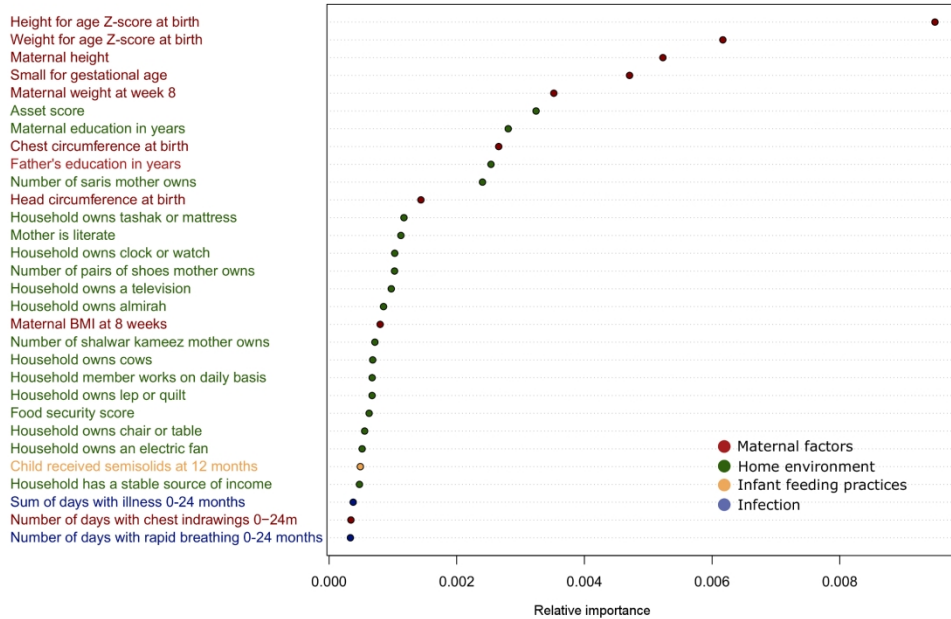


Figure 4. Conditional random forest plot ranking the relative importance of 30 predictors with regard to their ability to explain the presence of stunting at 24 months of age. The MINIMat cohort in rural Bangladesh. Colour coding according to the WHO conceptual framework on causes of stunting.

190x134mm (300 x 300 DPI)

1  
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

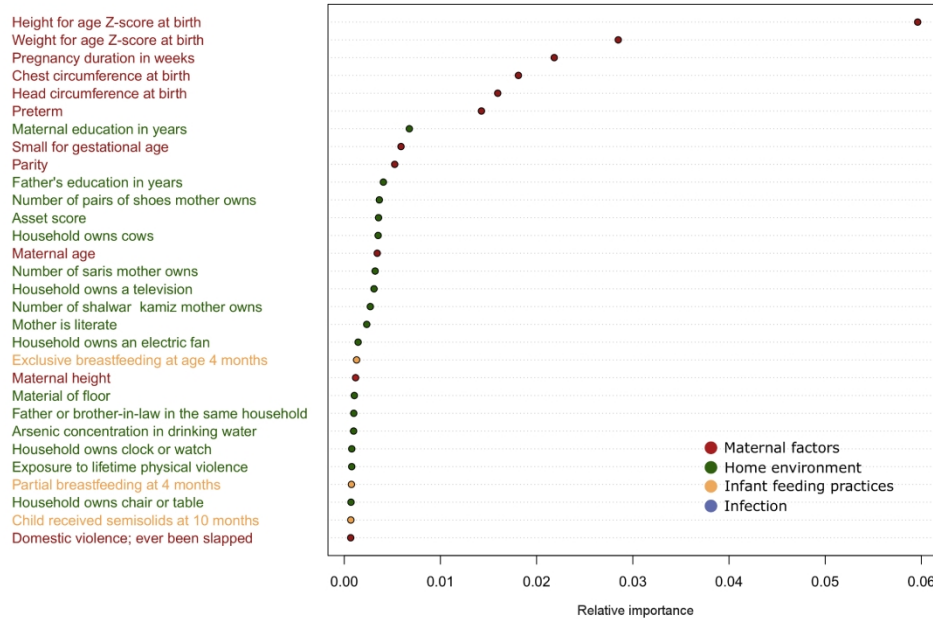


Figure 5. Conditional random forest plot ranking the relative importance of 30 predictors with regard to their ability to explain the variation in change in HAZ ( $\Delta$  HAZ) from birth to 24 months of age. The MINIMat cohort in rural Bangladesh. Colour coding according to the WHO conceptual framework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

190x134mm (300 x 300 DPI)

1  
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

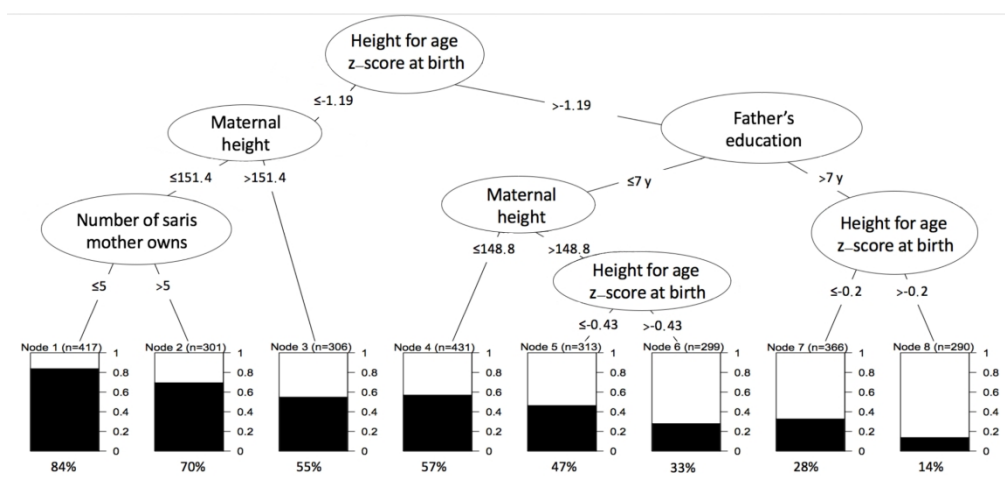


Figure 6. Conditional inference tree identifying sub-groups with different probabilities of stunting at 24 months. The MINIMat cohort in rural Bangladesh.

190x88mm (300 x 300 DPI)

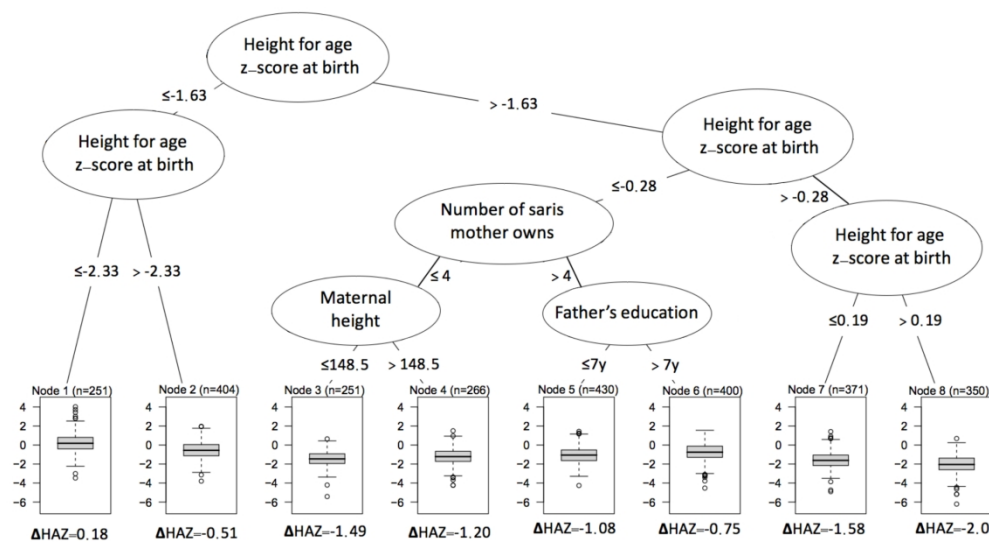


Figure 7. Conditional inference tree identifying sub-groups with different mean change in HAZ ( $\Delta HAZ=HAZ_{24}-HAZ_0$ ) 0–24 months within the MINIMat cohort in rural Bangladesh.

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].

**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> Misclassification rate

<sup>3</sup> Mean square error

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

$\beta$  = proportion of the original data sampled



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

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

<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  $MR^2$  and the  $MSE^3$  for different imputation methods, computed from  $m=100$  samples,  $\alpha = 0.2$ ,  $\beta = 0.05$

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

<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, Bühlmann P. MissForest—non-parametric missing value imputation for mixed-type data. *Bioinformatics*. 2012;: 112–118.
2. Jönsson P, Wohlin C. An Evaluation of K-Nearest Neighbour Imputation Using Likert Data. *Proceedings of the 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	
<b>Title and abstract</b>	1	(a) Indicate the study's design with a commonly used term in the title or the abstract	1
		(b) Provide in the abstract an informative and balanced summary of what was done and what was found	2
<b>Introduction</b>			
Background/rationale	2	Explain the scientific background and rationale for the investigation being reported	4-5
Objectives	3	State specific objectives, including any prespecified hypotheses	6
<b>Methods</b>			
Study design	4	Present key elements of study design early in the paper	5
Setting	5	Describe the setting, locations, and relevant dates, including periods of recruitment, exposure, follow-up, and data collection	6
Participants	6	(a) <i>Cohort study</i> —Give the eligibility criteria, and the sources and methods of selection of participants. Describe methods of 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	6
		(b) <i>Cohort study</i> —For matched studies, give matching criteria and number of exposed and unexposed <i>Case-control study</i> —For matched studies, give matching criteria and the number of controls per case	
Variables	7	Clearly define all outcomes, exposures, predictors, potential confounders, and effect modifiers. Give diagnostic criteria, if applicable	9 Figure 1
Data sources/ measurement	8*	For each variable of interest, give sources of data and details of methods of assessment (measurement). Describe comparability of assessment methods if there is more than one group	7-9
Bias	9	Describe any efforts to address potential sources of bias	7-9
Study size	10	Explain how the study size was arrived at	Not applicable
Quantitative variables	11	Explain how quantitative variables were handled in the analyses. If applicable, describe which groupings were chosen and why	7-9
Statistical methods	12	(a) Describe all statistical methods, including those used to control for confounding	10-11
		(b) Describe any methods used to examine subgroups and interactions	10-11
		(c) Explain how missing data were addressed	10-11
		(d) <i>Cohort study</i> —If applicable, explain how loss to follow-up was addressed <i>Case-control study</i> —If applicable, explain how matching of cases and controls was addressed	10

1  
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

*Cross-sectional study*—If applicable, describe analytical methods taking account of sampling strategy

(g) Describe any sensitivity analyses

10

Continued on next page10

For peer review only

**Results**

Participants	13*	(a) Report numbers of individuals at each stage of study—eg numbers potentially eligible, examined for eligibility, confirmed eligible, included in the study, completing follow-up, and analysed	Fig 2 12
		(b) Give reasons for non-participation at each stage	Fig 2
		(c) Consider use of a flow diagram	Figure 2
Descriptive data	14*	(a) Give characteristics of study participants (eg demographic, clinical, social) and information on exposures and potential confounders	Table 1, 13
		(b) Indicate number of participants with missing data for each variable of interest	13
		(c) <i>Cohort study</i> —Summarise follow-up time (eg, average and total amount)	13
Outcome data	15*	<i>Cohort study</i> —Report numbers of outcome events or summary measures over time	13
		<i>Case-control study</i> —Report numbers in each exposure category, or summary measures of exposure	
		<i>Cross-sectional study</i> —Report numbers of outcome events or summary measures	
Main results	16	(a) Give unadjusted estimates and, if applicable, confounder-adjusted estimates and their precision (eg, 95% confidence interval). Make clear which confounders were adjusted for and why they were included	14
		(b) Report category boundaries when continuous variables were categorized	14
		(c) If relevant, consider translating estimates of relative risk into absolute risk for a meaningful time period	
Other analyses	17	Report other analyses done—eg analyses of subgroups and interactions, and sensitivity analyses	14

**Discussion**

Key results	18	Summarise key results with reference to study objectives	15,18
Limitations	19	Discuss limitations of the study, taking into account sources of potential bias or imprecision. Discuss both direction and magnitude of any potential bias	15,16,17
Interpretation	20	Give a cautious overall interpretation of results considering objectives, limitations, multiplicity of analyses, results from similar studies, and other relevant evidence	18-22
Generalisability	21	Discuss the generalisability (external validity) of the study results	17

**Other information**

Funding	22	Give the source of funding and the role of the funders for the present study and, if applicable, for the original study on which the present article is based	23
---------	----	---	----

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

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

# BMJ Open

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

Journal:	<i>BMJ Open</i>
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

SCHOLARONE™  
Manuscripts

1  
2  
3 1 **Relative importance of pre- and postnatal determinants of stunting: data mining**  
4  
5  
6  
7

8 2 **approaches to the MINIMat cohort, Bangladesh**  
9  
10  
11  
12

13 3  
14  
15

16 4 Pernilla Svefors<sup>\*</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>,  
17  
18

19 5 Ashraful Islam Khan<sup>4</sup>, Katarina Ekholm Selling<sup>†</sup>  
20  
21  
22

23  
24 6 <sup>1</sup>International Maternal and Child Health, Department of Women's and Children's Health, Uppsala University, Uppsala, Sweden. <sup>2</sup>  
25  
26

27 7 Department of Computer and Information Sciences, Linköping University, Linköping, Sweden, <sup>3</sup>Department of Disease Control,  
28  
29

30 8 London School of Hygiene and Tropical Medicine, London, UK, <sup>4</sup>International Centre for Diarrhoeal Disease Research, Bangladesh  
31  
32

33 9 (iccd,r), Dhaka, Bangladesh.  
34  
35  
36  
37  
38

39 10 <sup>\*</sup>Corresponding author [pernilla.svefors@kbh.uu.se](mailto:pernilla.svefors@kbh.uu.se)  
40  
41  
42  
43  
44  
45  
46  
47  
48  
49  
50  
51  
52  
53  
54  
55  
56  
57  
58  
59  
60

## Abstract

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

**Methods** Conditional inference-tree-based methods were applied to the extensive Maternal and Infant Nutrition Interventions in Matlab (MINIMat) trial database with 309 variables of 2,723 children, their parents, and living conditions, including socioeconomic, nutritional and other biological characteristics of the parents; maternal exposure to violence; household food security; breast and complementary feeding; and measurements of morbidity of the mothers during pregnancy and repeatedly of their children up to 24 months of age. Child anthropometry was measured monthly from birth to 12 months, thereafter quarterly to 24 months.

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

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



1  
2 29 **Strengths and limitations of this study**  
3  
4  
5

- 6 30 • Includes high-quality longitudinal data with low rates of missing data on child growth and a wide range of pre and  
7  
8  
9  
10 31 postnatal household, family, and environmental factors, child characteristics at birth, infant feeding, and morbidity.  
11  
12  
13  
14 32 • Employs decision-tree-based methods that permit the inclusion of a high number of predictor variables, variables of  
15  
16  
17 33 different types and automatically discover complex interactions between predictor variables and include them in the  
18  
19  
20  
21 34 model.  
22  
23  
24  
25 35 • Some potentially important determinants of linear growth were not present in the database.  
26  
27  
28  
29 36 • The study does not include stratified analyses for girls and boys  
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

## 38 Introduction

39 Linear growth is considered to be the best overall indicator of children's present and future health<sup>[1,2]</sup> and the reduction of growth  
40 failure is one of the targets within the sustainable development agenda. Stunted growth is associated with short-term morbidity  
41 and mortality, impaired cognitive development, lower future productivity, and increased risk of adult chronic diseases [3]. In 2012,  
42 the WHO adopted a resolution on maternal and child undernutrition, targeting a reduction of stunting by 40% by 2025 [4]. Linear  
43 growth is most susceptible to environmentally modifiable factors from conception up to two years of age, i.e., the first 1000 days  
44 when most of the growth faltering takes place [5, 6]. To develop and deliver appropriate interventions, it is imperative to establish  
45 the relative importance of stunting risk factors. In addition, the sustainable development health goal has emphasized the  
46 personalized perspective under the universal coverage of health care. Precision public health interventions by identifying and  
47 targeting high-risk subgroups can be one of the strategies to reach this goal<sup>[7]</sup>.

48 Previous studies employing classical statistical methods have identified a wide range of pre- and post-natal factors  
49 associated with impaired growth [8-13]. Low birth weight, maternal height, maternal education, poverty and inadequate  
50 complementary feeding practices have been recognized as important risk factors [14-16]. Some analyses emphasize the importance  
51 of fetal growth restriction for later stunted growth, but rarely is the relative importance of pre- and post-natal factors assessed [17].  
52 Despite these findings, policy documents and recommendations emphasize interventions especially after birth, and pre-natal  
53 recommendations are usually limited to routine micronutrient supplementation for pregnant women [18-20].

1  
2 54           Despite a wealth of literature relating to the determinants of stunting, studies with a holistic approach, which  
3  
4  
5  
6 55           concurrently account for household, environmental, nutritional, biological, and socioeconomic influences are few. Moreover,  
7  
8  
9 56           individuals and groups may be stunted for various reasons and thus respond differently to interventions. Studies that identify risk  
10  
11  
12  
13 57           groups with different probabilities of stunting are, to the best of our knowledge, not yet available. The available studies with a  
14  
15  
16 58           multifactorial approach have frequently had a cross-sectional design and have applied traditional statistical methods. As visualized  
17  
18  
19 59           in the WHO's conceptual framework for childhood stunting [21], the causes of stunted linear growth are complex. The number of  
20  
21  
22  
23 60           risk factors and the complexity of the associations of these risk factors with linear growth restriction make traditional statistical  
24  
25  
26 61           models ineffective from a predictive perspective. Moreover, classical statistical methods do not have the capacity to identify groups  
27  
28  
29  
30 62           with different risks based on combinations of predictors. Decision trees are popular data mining (DM) methods, which allows for  
31  
32  
33 63           the inclusion of a high number of predictor variables, handling variables of different types, automatically discovering complex  
34  
35  
36 64           interactions between predictor variables and including them in the model [22]. Decision-tree-based algorithms can be used to rank  
37  
38  
39  
40 65           a high number of predictors according to their relative importance for the outcome and to identify subgroups with different risk  
41  
42  
43 66           patterns.

44  
45  
46  
47 67           The Maternal and Infant Nutrition Interventions in Matlab (MINIMat) was a randomized prenatal food and multiple  
48  
49  
50  
51 68           micronutrient trial carried out in rural Bangladesh. The frequent follow-up of mothers and children participating in this trial  
52  
53  
54 69           resulted in an extensive database, including frequent pre- and post-natal anthropometric assessments, socioeconomic and  
55  
56  
57  
58 70           biological characteristics of the mother and father, information on maternal exposure to violence, household food security, breast-  
59  
60  
71           and infant-feeding practices, and measurement of morbidity of the mothers during pregnancy and repeatedly of children up to 24

1  
2 72 months of age. The aim of this study is to, within this Bangladeshi cohort, assess the relative importance of determinants of linear  
3  
4  
5  
6 73 growth from 0–24 months and risk factors for stunting at two years, and to identify risk groups with negative growth trajectories  
7  
8  
9 74 and high prevalence of stunting at two years.  
10  
11  
12

## 14 75 **Methods**

### 17 76 **Study setting, participants and study design**

20 77 The MINIMat trial (Maternal and Infant Nutrition Interventions in Matlab, isrcrn.org identifier: ISRCTN16581394) was carried out  
21  
22  
23 78 in Matlab, Bangladesh, a rural delta region located 57 km southeast of the capital Dhaka. In this area, a health and demographic  
24  
25  
26 79 surveillance system enables early pregnancy identification and longitudinal follow-up. Pregnant women were enrolled in the  
27  
28  
29  
30 80 MINIMat trial and the follow-up included their offspring. MINIMat was a factorial randomized trial primarily evaluating the effect  
31  
32  
33 81 of an early invitation to prenatal food supplementation (versus usual timing) combined with multiple micronutrient  
34  
35  
36  
37 82 supplementation (versus usual program iron-folate) to pregnant women on maternal hemoglobin, birth weight, gestational age at  
38  
39  
40 83 birth, and infant mortality [23]. Further, the participating women were randomly assigned to either counselling for exclusive  
41  
42  
43  
44 84 breastfeeding or a different health education message of equivalent intensity [24]. The MINIMat trial recruited pregnant women  
45  
46  
47 85 from November 2001 to October 2003. When a woman reported to a community health worker that her menstruation was delayed  
48  
49  
50 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  
51  
52  
53  
54 87 was missing, the gestational age assessment was based on ultrasound examination. In total, 4436 pregnant women participated,  
55  
56  
57 88 giving birth to 3625 live born infants from April 2002 to June 2004. The pregnant women were enrolled at around gestational week  
58  
59  
60 89 8. In this analysis, the mothers and children were followed through pregnancy, birth, and up to two years of age.

1  
2 90 Written and oral informed consent was obtained from all participating women and from the parents of the participating  
3  
4  
5  
6 91 children. The Ethical Review Committee at the International Centre for Diarrhoeal Disease Research, Bangladesh, approved the  
7  
8  
9 92 study (approval registration numbers 2000-025; 2002-031; 2005-004)  
10  
11  
12

13 93

## 18 94 **Data collection**

19  
20  
21 95 Predictor and outcome variables are presented in Figure 1, grouped according to the WHO conceptual framework of stunting [21].  
22  
23

24 96 Data were collected using questionnaires, physical examinations, and laboratory analyses. At enrolment, well-trained field workers  
25  
26

27  
28 97 collected information on women's age, parity, marital status, educational level, occupation, maternal morbidity, socioeconomic  
29  
30

31 98 characteristics, and household food security. Socioeconomic status was assessed based on a range of household assets, and a  
32  
33

34 99 continuous household asset score, with a mean value of zero, was constructed based on a principal component analysis [25]. A  
35  
36

37  
38 100 validated household food security scale was created from eleven items with data on frequency of food purchased, cooked, borrowed  
39  
40

41 101 or lent (food and money), and whether there was ready access to adequate meals and snacks [26]. The participating women were  
42  
43

44  
45 102 also asked whether they had suffered any of thirty morbidity symptoms from twelve different categories, including airway, urinary  
46  
47

48 103 tract, fever, circulation, bowel, or pain symptoms during the last month. A sum score ranging from zero to twelve was created based  
49  
50

51 104 on absence of symptoms or those not recorded for each category.  
52  
53  
54

55  
56 105 Home visits were followed by clinic visits at local health sub-centers. Maternal height and weight were measured at  
57  
58

59 106 around eight weeks of gestation using a stadiometer to the nearest 0.1 cm and an electronic scale (Uniscale; SECA) with a precision  
60

1  
2 107 of 0.10 kg. In the third trimester, paramedics interviewed the participating women in privacy regarding their experiences of  
3  
4  
5  
6 108 domestic violence. A modified version of the WHO collaborative study questionnaire was used [27,28], based on the conflict tactic  
7  
8  
9 109 scale covering physical, sexual and emotional violence and controlling behavior [29]. Household drinking water was analyzed for  
10  
11  
12  
13 110 arsenic concentration [30].

17 111 A birth notification system allowed birth anthropometry to be measured within 72 hours. In the few cases where the  
18  
19  
20 112 newborns were reached after 72 hours, the measurements were adjusted to the time of birth using an SD score transformation,  
21  
22  
23  
24 113 assuming that the infants remained in the same relative position in the anthropometric distribution during this period [31]. At  
25  
26  
27 114 birth, data on sex, birth weight, length, and breastfeeding practices were collected. During the subsequent two-year study period,  
28  
29  
30  
31 115 the mother-and-child pairs were visited monthly in their homes during the first year, and every three months during the second  
32  
33  
34 116 year. On these occasions, data on infant feeding practices, child morbidity and anthropometry were collected. The mothers were  
35  
36  
37 117 interviewed about breastfeeding and complementary feeding practices. Breastfeeding practices were categorized into exclusive,  
38  
39  
40  
41 118 predominant, partial, or any breastfeeding for each month from one to twelve months. The total time for exclusive, predominant,  
42  
43  
44 119 and any breastfeeding was calculated. The WHO recommendations guided the breastfeeding assessment [32] and results were  
45  
46  
47 120 validated with a stable-isotope technique. The classification of exclusive breastfeeding was found to suffer from limited  
48  
49  
50  
51 121 misclassification in both directions and to be accurate at the group level [33]. The food given to the infant was categorized into  
52  
53  
54 122 semi-solids and solids each month from one to twelve months. The data collection did not include full dietary assessments or  
55  
56  
57  
58 123 classification of dietary diversity and meal frequency.

1  
2 124 The mothers were also asked whether the child had had any of the following symptoms during the last week; fever, cough,  
3  
4  
5  
6 125 difficult breathing, chest in-drawing, rapid breathing, diarrhea, bloody diarrhea and the duration of these symptoms [34].  
7  
8  
9 126 Categories were created based on whether the child had suffered from fever, respiratory symptoms, suspected pneumonia, or  
10  
11  
12  
13 127 diarrhea, and the sum of days with each symptom and total morbidity calculated from birth to 24 months. To reduce the risk of  
14  
15  
16 128 recall bias the mothers were visited monthly with an interview recall period of seven days for child morbidity. One week has been  
17  
18  
19 129 found to be optimal for this kind of morbidity recall assessment [35].  
20  
21  
22

23  
24 130 Children's weight was measured by SECA beam and electronic scales (UNICEF Uniscale; SECA GmbH & Co, Hamburg,  
25  
26  
27 131 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  
28  
29  
30 132 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  
31  
32  
33  
34 133 stadiometer. Head and chest circumference were measured with a measuring tape. Two measurements were recorded on each  
35  
36  
37 134 occasion and the mean was calculated. The equipment was calibrated daily and refresher training on data collection methods,  
38  
39  
40  
41 135 including the standardization of anthropometric measurements, was conducted periodically.  
42  
43  
44

## 45 136 **Outcomes**

46  
47  
48 137 Height-for-age z-scores (HAZ) were calculated from the measured length and height data using the program WHOAnthro, based  
49  
50  
51 138 on the WHO growth reference for children [36]. Children with a HAZ below minus two SD-scores were classified as stunted. Two  
52  
53  
54  
55 139 outcomes were analyzed: stunting at 24 months and the change in HAZ from birth to 24 months, referred to as  $\Delta$  HAZ and  
56  
57  
58 140 calculated by subtracting HAZ at birth from HAZ at 24 months i.e.  $\Delta$ HAZ = HAZ at 24 months - HAZ at birth.  
59  
60

## 141 **Statistical analysis**

142 A database was created with 309 variables characterizing mothers and children in the MINIMat cohort from enrolment in early  
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  
144 and 24 months was selected ( $n=2\ 723$ ). The average percent of missing values among all the predictors were 4 %. The highest  
145 percent missing were among maternal morbidity data during pregnancy (22%) and categorical monthly child morbidity data (ill  
146 or not), ranging from 0% to 35% with the highest number of missing observations in the first months. The continuous child  
147 morbidity data however (sum of days with different types illnesses), had no missing values. The most important variables identified  
148 by the random forest analyses and the variables included by the conditional inference trees had less than 1% missing values. The  
149 missing values of the predictor variables were imputed. To find the best method to impute the missing data we made a simulation  
150 study of the performance of the following imputation methods: imputation by variable mean, K-nearest neighbor imputation [37],  
151 and random forest imputation [38]. The design of the study followed a procedure similar to the strategy described in Jonsson et al.  
152 [37], see S appendix. Accordingly, we imputed the data by use of the random forest as the simulation study revealed that this  
153 method provided the most accurate imputations.

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  
156 the model. Traditional decision trees, such as Classification and Regression Trees (CART) have been shown to be biased [39]. This  
157 motivated us to select the Conditional Inference Trees (CIT) framework, a method that embeds a statistical hypothesis-testing  
158 framework into a recursive partitioning algorithm used for model building [39]. Conditional inference trees were used in order to



1  
2 159 identify sub-groups characterized by combinations of levels of certain predictors with distinct values of  $\Delta$  HAZ or prevalence of  
3  
4  
5  
6 160 stunting at 24 months. Cross-validation, a well-established model selection method that selects a tree with an optimal predictive  
7  
8  
9 161 performance for new unseen data, was applied. Cross-validation splits the data set into different train and test sets repeatedly,  
10  
11  
12  
13 162 estimates the model in one set and validates the prediction on another set, followed by an aggregation of the predictions[40]. To  
14  
15  
16 163 ensure public health relevance, the minimum number of observations in each terminal node (subgroup) was set to 250.

17  
18  
19  
20 164 Conditional random forest (CRF) analyses were performed to assess and rank the importance of predictors with regard to  
21  
22  
23  
24 165 their ability to explain the variation of the continuous outcome of the change in HAZ from birth to 24 months and the presence of  
25  
26  
27 166 stunting at 24 months of age. In conditional random forest analysis, an ensemble of conditional inference trees is created by means  
28  
29  
30  
31 167 of drawing subsamples from the original data and fitting a unique randomized conditional inference tree to each sample. Possible  
32  
33  
34 168 predictors at each split are selected randomly from the complete set of predictors, which leads to a better predictive performance of  
35  
36  
37 169 the tree ensemble [40]. The importance of a variable is computed by comparing the predictive mean squared error (MSE) from the  
38  
39  
40  
41 170 original data and a dataset where the corresponding variable values are specified incorrectly, which makes the variable irrelevant  
42  
43  
44 171 for the prediction. If the variable does not contribute to the prediction, the MSE is expected to be small when the values of the  
45  
46  
47  
48 172 variable are permuted. An aggregated difference between the MSE values over the given ensemble of trees makes up the relative  
49  
50  
51 173 importance of a variable. The random forests analyses were created based on 3000 trees, and the 30 variables with the highest  
52  
53  
54 174 importance measure are presented. The programming language R version 3.2.4 [41] and the 'party' package [42] were used for all  
55  
56  
57  
58 175 analyses.

## 176 Patient and public involvement

177 No participants were involved in developing the hypothesis, the specific aims or the research questions, nor were they involved in  
178 developing plans for design or implementation of the study. No participants were involved in the interpretation of study results or  
179 write up of the manuscript. There are no plans to disseminate the results of the research to study participants.

## 180 Results

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

187 The characteristics of the households, mothers, fathers at eight weeks of gestation, and children at birth are given in Table

188 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  
189 (SD 6.8) at recruitment. One-third of the women were underweight, with a BMI below 18.5 at pregnancy week eight. The average  
190 number of years of education was similar for mothers and fathers (5 years). The sample of children comprised an equal proportion  
191 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 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  
4  
5  
6 193 years of -2.0 (Figure 3) and 50% being stunted (girls 51.1%, boys 48.5%)  
7  
8  
9

10 194

11  
12  
13  
14 195 **Table 1.** Baseline characteristics, prevalence of stunting at 24 months, and mean  $\Delta$  HAZ (change in height-for-age Z-score) 0–24  
15  
16  
17  
18 196 months in the MINIMat cohort, Bangladesh.  
19  
20

21  
22  
23 197

Characteristics	n/n (%)	Stunted at 24 months n/n (%)	Mean $\Delta$ HAZ 0-24 months
<b>Mother's age (years)</b>			
<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
<b>Mother's education</b>			
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
<b>Father's education</b>			
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
<b>Parity</b>			
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
<b>Number of saris mother owns</b>			
<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
<b>Child at birth</b>			
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

54  
55  
56  
57  
58  
59  
60

1  
2  
3 210 **Relative importance of predictors for stunting at 24 months and change in height scores from**  
4  
5 211 **birth to 24 months**

6  
7  
8 212 The relative importance of predictors with respect to their ability to explain the probability of stunting at 24 months and the  
9  
10  
11 213 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  
12  
13  
14  
15 214 most important predictors of stunting at 24 months, followed by maternal height, Small for Gestational Age (SGA), maternal  
16  
17  
18 215 weight at eight weeks of gestation, household asset score, and parental education. The most important factors for  $\Delta$  HAZ were HAZ  
19  
20  
21 216 and WAZ at birth, pregnancy duration, head and chest circumference at birth, and maternal education.  
22  
23  
24  
25

26 217 **Subgroups with different levels of stunting at 24 months and levels of change in height scores from**  
27  
28 218 **birth to 24 months**

29  
30  
31 219 The conditional inference trees presented in Figure 6 and 7 display subgroups with different probability of stunting at 24 months  
32  
33  
34 220 and levels of  $\Delta$  HAZ 0-24 months due to distinctive combinations of levels of certain predictors. The conditional inference trees for  
35  
36  
37  
38 221 stunting and  $\Delta$ HAZ were composed of subgroups defined by the same predictors, specifically; HAZ at birth, maternal height,  
39  
40  
41 222 father's educational level, and the number of saris owned by the mother. The probability of stunting ranged from 14% to 84%.  
42  
43  
44  
45 223 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  
46  
47  
48 224 highest probability of stunting at 24 months, at 84%. Children of a father with more than seven years of education, who had HAZ  
49  
50  
51 225 at birth above -0.2, had the lowest probability of stunting at 24 months, at 14% (Figure 6). The difference in  $\Delta$  HAZ between the  
52  
53  
54  
55 226 identified subgroups of children with the most negative change and the subgroup with the most positive change was 2.22 HAZ.  
56  
57  
58  
59  
60

1  
2 227 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  
3  
4  
5  
6 228 HAZ), while children who were born with a HAZ above 0.19 had the most negative  $\Delta$  HAZ (-2.04 HAZ) (Figure 7).  
7  
8  
9

10 229

14 230

## 19 231 Discussion

22 232 In our analysis of 309 predictors characterizing household, environmental, biological, and socioeconomic factors, we found birth  
23  
24  
25 233 size, maternal anthropometry and parental education to be the most influential for linear growth up to and stunting at 24 months.  
26  
27  
28  
29 234 Conditions after birth, such as feeding practices and morbidity, were less important for linear growth trajectories and stunting at  
30  
31  
32 235 two years. The difference between the identified subgroups of children with the highest and lowest probabilities of stunting was as  
33  
34  
35 236 high.

40 237 The most important predictors of stunting at 24 months were different indicators of size at birth, maternal height, asset  
41  
42  
43 238 score and maternal education. These findings are in line with a multi-country longitudinal study that found birth or enrollment  
44  
45  
46 239 weight of the infant and maternal height to have the highest cumulative odds ratios for linear growth deficit up to two years of age  
47  
48  
49  
50 240 [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  
51  
52  
53 241 a pooled analysis of 19 birth cohorts with longitudinal follow-up, 20% of stunting was attributable to small-for-gestational-age  
54  
55  
56 242 weight at birth [17]. That study did not include any post-natal factors in the analysis. In a study in Indonesia, neonatal length and  
57  
58  
59  
60 243 weight were the strongest predictors of nutritional status and increases in weight and length during infancy [44]. Our study

1

2 244 included both pre- and post-natal factors and, in contrast to most other studies, assessed not only the relative importance of  
3  
4  
5  
6 245 different potential predictors, but also the public health importance of each element.  
7  
8  
9

10 246 In a study with pooled data from five Demographic and Health Surveys in South Asia, maternal height and underweight,  
11  
12  
13 247 household wealth, maternal education, and minimum dietary diversity were found to be the most important factors among  
14  
15  
16  
17 248 children aged 6–23 months [16]. Similar results were reported from a study in India [45]. These studies were, however, cross-  
18  
19  
20 249 sectional, without access to birth characteristics.  
21  
22  
23

24 250 Maternal height is a strong determinant of fetal growth [46] that indirectly reflect the epigenetic heredity. Maternal  
25  
26  
27  
28 251 height is directly associated with the uterine volume [47], cephalo-pelvic disproportion and subsequent infant and childhood  
29  
30  
31 252 stunting, and child mortality [48,49]. In a previous analysis of the MINIMat cohort, a short maternal height was strongly associated  
32  
33  
34  
35 253 with stunting all the way up to 10 years of age [49]. Thus, factors that well precede pregnancy generate a vicious intergenerational  
36  
37  
38 254 cycle, where small mothers give birth to small children of whom a high proportion become and remain stunted. In the conditional  
39  
40  
41 255 inference trees for stunting at 24 months, children who were born with a higher HAZ but who had shorter mothers were as likely to  
42  
43  
44  
45 256 be stunted as children with lower HAZ at birth but with a taller mother. This finding suggests that intergenerational improvements  
46  
47  
48 257 in height are achievable and that interventions with a particular focus on adolescents and women of reproductive health are needed  
49  
50  
51  
52 258 to break the vicious intergenerational cycle.  
53  
54  
55

56 259 A strong relationship between stunting and poverty has been reported from many low-middle income settings [50]. Asset  
57  
58  
59 260 score and other socioeconomic markers, such as the number of shoes and saris the mother owned, were highly ranked in the  
60

1  
2  
3 261 random forest analysis and categorized subgroups with a higher probability of stunting and undesirable linear growth trajectories.  
4  
5  
6 262 Poverty is associated with unfavorable food and sanitation practices that can lead to poor nutrition and an increased occurrence of  
7  
8  
9 263 infections during pregnancy, infancy, and childhood. Poverty increases the risk of maternal stress, depression [51] and weak  
10  
11  
12  
13 264 mother-to-child interaction and stimulation.  
14  
15  
16  
17 265 The number of shoes and saris the mother owns might also be markers of the woman's status in the household. During  
18  
19  
20 266 the last few decades, the importance of women's position in household and society for child nutrition has been emphasized [52].  
21  
22  
23  
24 267 Maternal status is associated with food allocation to mother and child, and a higher level of maternal autonomy has been  
25  
26  
27 268 associated with better child weight and lower levels of stunting [53]. The subordinate position of women in South Asia has been  
28  
29  
30  
31 269 suggested to be a contributor to the high prevalence of child undernutrition in the region, compared to other areas with equivalent  
32  
33  
34 270 levels of economic growth and food security [52].  
35  
36  
37  
38 271 An acknowledged way of increasing women's position is through improved education. The remarkable health  
39  
40  
41  
42 272 achievements in Bangladesh over the past two decades can partly be attributed to the progress in access to education, especially at  
43  
44  
45 273 primary level and for girls [54]. However, there is a considerable risk of not completing primary school for both girls and boys [55].  
46  
47  
48  
49 274 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  
50  
51  
52 275 conditional decision trees models for stunting and change in HAZ, the cut-off values for paternal and maternal education in the  
53  
54  
55 276 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,  
56  
57  
58  
59 277 furthering the importance of girls and boys not only enrolling in but also continuing at school.  
60

1  
2 278 It may seem contradictory that children who were born with a very short length had the smallest change in HAZ. This

3  
4  
5  
6 279 finding most likely reflects a situation where linear growth had already been severely restricted in fetal life.

7  
8  
9  
10 280 A multi-country pooled analysis of cohort studies showed that a higher cumulative burden of diarrhea increased the risk

11  
12  
13 281 of stunting [57]. In situations, where measles still occurred, its impact on growth and mortality risks were repeatedly documented

14  
15  
16  
17 282 [58]. One explanation to the discrepancy between our results and previous findings could be Bangladesh's remarkable success in

18  
19  
20 283 achieving the globally highest coverage of oral rehydration therapy in diarrhea [59], which may have reduced the impact on linear

21  
22  
23 284 growth. Another factor is the almost universal immunization coverage [60,61] that has reduced or partly eliminated immunization-

24  
25  
26  
27 285 preventable morbidity and the subsequent effect on growth. Our previous publications on the MINIMat prenatal nutrition

28  
29  
30 286 interventions' effects on child growth and mortality were not mediated through morbidity [23,62], further supporting the modest

31  
32  
33 287 impact of child morbidity on linear growth in our sample [34]. In other settings with lower coverage of diarrhea treatment and

34  
35  
36  
37 288 immunization, the relative importance of these factors may be greater.

38  
39  
40  
41 289 Suboptimal infant and early childhood feeding practices have, in earlier studies, been reported as significant risk factors

42  
43  
44  
45 290 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–

46  
47  
48 291 24 months had been randomized to appropriate complementary foods [64]. The infant feeding variables included in our analysis

49  
50  
51  
52 292 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

53  
54  
55 293 documented effects of complementary feeding programs on stunting, these interventions are often the priority in efforts to combat

56  
57  
58 294 stunting.



1  
2 295 The nutrition interventions from pre-conception to two years of age currently recommended by the WHO include efforts  
3  
4  
5  
6 296 to ensure exclusive breastfeeding, adequate complementary feeding, appropriate nutritional care of sick and malnourished  
7  
8  
9 297 children and proper intake of vitamin A, iron and iodine for women and children [19]. All of these, except micronutrient  
10  
11  
12  
13 298 supplementation to pregnant women, are focused on the postnatal period from birth up to two years. Our results strengthen the  
14  
15  
16 299 evidence that the process of becoming stunted already begins in utero, as well as the importance of intergenerational effects.  
17  
18  
19 300 Although worthwhile, the present focus on postnatal interventions results in missed opportunities to intervene before or during the  
20  
21  
22  
23 301 first nine months when the process of stunting is established.

24  
25  
26  
27 302 So, what possibilities do we have to improve the postnatal linear growth trajectories prenatally? Attained height is mainly  
28  
29  
30  
31 303 dependent on one's genetic potential for linear growth, in turn determined by DNA sequence polymorphism [65,66] and epigenetic  
32  
33  
34 304 heredity [67], and to some extent the environment. The modulation of non-DNA sequence epigenetic heredity has been proposed to  
35  
36  
37 305 be one of the leading factors explaining variations in height and height changes over generations[67], especially in more deprived  
38  
39  
40  
41 306 populations [68]. Postnatal interventions can influence factors in the environment that constrain the ability to increase linear  
42  
43  
44 307 growth, while prenatal interventions also have the potential to modulate the actual growth potential through an epigenetic  
45  
46  
47 308 modification that results from changes to gene expression in response to the fetal environment.

48  
49  
50  
51 309 Established prenatal nutritional interventions include balanced energy-protein supplementation, multiple micronutrient  
52  
53  
54  
55 310 supplements, and nutritional counseling and education. Unfortunately, most studies evaluating these interventions report only  
56  
57  
58 311 birth weight, not length, which is why evidence to directly assess the effect on fetal linear growth is limited. Meta-analyses and  
59  
60

1  
2  
3 312 randomized trials evaluating these interventions report their positive impact on birth weight and a reduced risk of LBW [69-76].  
4  
5  
6 313 Effect sizes vary from increases in birth weight of 20–200g, with the smallest effects seen in studies of multiple micronutrients and  
7  
8  
9 314 bigger effects seen by balanced energy-protein and lipid-based nutrient supplements. Considerable heterogeneity in growth  
10  
11  
12  
13 315 response is common, and is related to the mother's nutritional status when entering pregnancy and possibly also to the genetic  
14  
15  
16 316 potential to benefit. In the MINIMat food and micronutrient interventions, all women received food supplementation, but they were  
17  
18  
19 317 randomized to an early invitation to supplementation (week 9) or the usual program start of supplementation (week 20). Children  
20  
21  
22  
23 318 of mothers who participated in food supplementation from early pregnancy (versus the usual start) had a 13% reduction in  
24  
25  
26 319 stunting up to five years [62].  
27  
28  
29

30  
31 320 There is increasing evidence that preconception interventions may be even more appropriate[77]. A few trials examining  
32  
33  
34 321 the effect of interventions initiated before pregnancy are underway, but few results have so far been published [78]. Preconception  
35  
36  
37 322 interventions have the potential to bring about epigenetic modulation and improved growth in present and future generations.  
38  
39  
40  
41 323 Thus, the launch and evaluation of interventions targeting adolescent and women of reproductive age that focus on adequate  
42  
43  
44 324 health, education, and nutrition before and during pregnancy is needed, especially in South Asia with its high burden of maternal  
45  
46  
47 325 undernutrition and young age at first pregnancy [79]. Targeting high-risk subgroups, in this setting characterized by short, poor,  
48  
49  
50  
51 326 women with low education, can be another strategy to address the intractable problem of stunting.  
52  
53  
54

## 55 327 **Strengths and limitations**

56  
57  
58  
59  
60

1  
2 328 The extensive database that was available for our analysis covered a wide range of household, family, and environmental factors,  
3  
4  
5  
6 329 child characteristics at birth, feeding, and morbidity. Infant and young child growth was carefully assessed from birth up to two  
7  
8  
9 330 years. The MINIMat cohort was implemented in an excellent research infrastructure that fulfills the prerequisites for obtaining  
10  
11  
12 331 high-quality longitudinal data. Experienced field workers and study nurses collected data on the 309 variables during pregnancy  
13  
14  
15  
16 332 and the following two years. They received repeated training, including standardization exercises, and were supervised by senior  
17  
18  
19 333 medical doctors.

20  
21  
22  
23 334 Some potential determinants were not present in the database. Household water, sanitation, and hygiene (WASH)  
24  
25  
26  
27 335 characteristics were limited to information on arsenic contamination of the drinking water, but diarrhea and other morbidity  
28  
29  
30 336 information were included in our analyses. Further, the cohort did not include the collection of stools for the study of  
31  
32  
33  
34 337 enteropathogens in the child, which may be associated with the risk of stunting [11]. Paternal height, which may be related to fetal  
35  
36  
37 338 growth, was not available [80]. The mothers' smoking habits were not represented in the data, as smoking was extremely rare  
38  
39  
40  
41 339 among women in the study area.

42  
43  
44  
45 340 There were slight differences in basic characteristics of the analyzed and non-analyzed groups. These differences had  
46  
47  
48 341 most likely no influence on the primary outcomes of this study. There were no or few missing values of the critical variables that  
49  
50  
51  
52 342 ranked high in the random forest and defined the sub-groups in the conditional inference trees. A sub-study was carried out to  
53  
54  
55 343 ensure the most accurate method to impute missing data. Thus, it is also highly unlikely that missing data influenced the main  
56  
57  
58 344 findings.  
59  
60

1  
2  
3 345 Some of the included variables like “household asset score” are composite variables, which depend on individual variables  
4  
5  
6 346 like TV ownership, number of cows, etc. Presence of both composite and individual variables creates computational problems for  
7  
8  
9 347 traditional models like linear regression and for some machine learning models due to a possible high correlation between the  
10  
11  
12  
13 348 individual and the composite variables. However, CIT methods perform automatic variable selection by choosing the most relevant  
14  
15  
16 349 variable (with the strongest association to the response) at each decision tree split step [39]. Accordingly, these methods  
17  
18  
19 350 automatically choose either a composite variable or an individual variable at each split step based on the relevance of this variable  
20  
21  
22  
23 351 to the response.  
24  
25  
26

27 352 Traditional methods like linear regression often have lower predictive power than data mining methods. In some cases,  
28  
29  
30 353 the traditional methods are not even possible to compute due to a high number of predictor variables and complex interactions.  
31  
32  
33  
34 354 The method used in this work, Conditional Inference Trees, belongs to the class of Interpretable Machine Learning models and  
35  
36  
37 355 display precise information on the priority, size, and direction of the association of the predictors with the outcome. In addition, the  
38  
39  
40  
41 356 risk group identification, including the prioritization and relevant cut-offs of risk factors, can be of high public health relevance for  
42  
43  
44 357 the design and targeting of appropriate interventions with the most significant benefit. Thus, we believe that the CIT framework  
45  
46  
47 358 has a large potential in public health and medical applications.  
48  
49  
50

51  
52 359 It can be noted that the CRF and the CIT models are not fully comparable. This can be explained by two factors. Firstly,  
53  
54  
55 360 many predictors that were important in the CRF model are relatively highly correlated and thus have a similar relationship to the  
56  
57  
58  
59 361 response. Once one of these variables is selected by the decision tree in a split, there is a high chance that the remaining correlated  
60

1  
2 362 variables (although also important according to the CRF) will not be picked up as the next splitting variable. Secondly, the CRF  
3  
4  
5  
6 363 models and the CIT models cannot be matched directly. The CRF is a combination of many trees and is thus a more flexible model  
7  
8  
9 364 than a CIT. However, CRFs are nearly black-box models: the only interpretable information that these models deliver is the variable  
10  
11  
12  
13 365 importance measure. On the contrary, CITs are “transparent” and interpretable models but have a smaller predictive power. This is  
14  
15  
16 366 another reason why these models are not generally capable of efficiently embedding all the variables that are important in the  
17  
18  
19 367 CRFs.

20  
21  
22  
23  
24 368 Another potential limitation is that decision trees do not deliver *p*-values or confidence intervals. The cross-validation  
25  
26  
27 369 method, however, ensures that the selected tree is optimal. This validation method was chosen superior to other model validation  
28  
29  
30 370 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  
31  
32  
33  
34 371 computational burden.

35  
36  
37  
38 372 The study setting was a low socioeconomic area in rural Bangladesh, where maternal and child undernutrition in early  
39  
40  
41 373 life still is widespread. The growth trajectories of our cohort were consistent with established growth trajectories in South Asia,  
42  
43  
44  
45 374 where children are born below the WHO growth reference and falter dramatically up to 24 months of age [5]. The sub-continent of  
46  
47  
48 375 South Asia and Sub-Saharan Africa share similar proportions of stunted children and faltering patterns. The sub-Saharan African  
49  
50  
51  
52 376 children are however, on average born slightly bigger than children in South Asia [5], which makes our results mainly relevant for  
53  
54  
55 377 the South Asian context.

## 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 and postnatal data, household and family information, environmental factors, child characteristics at birth, infant feeding, and  
7  
8  
9 381 morbidity. Prenatal factors including birth size, the mother's anthropometry, and parental education were the most critical factors  
10  
11  
12  
13 382 for stunting at 24 months. These results should be seen in contrast to present practice and recommendations that mainly are  
14  
15  
16 383 limited to child interventions. The findings emphasize the benefit of interventions before conception and during pregnancy to  
17  
18  
19 384 reach a substantial reduction in stunting.  
20  
21  
22  
23  
24 385  
25  
26  
27  
28 386  
29  
30  
31  
32 387  
33  
34  
35  
36 388  
37  
38  
39  
40

## 41 389 **Contributors**

42  
43  
44 390 PS contributed to study design, data analysis and interpretation of the results and had the main responsibility of writing the paper.  
45  
46

47 391 LÁP and SEA were principal investigators of the MINIMat project. ECE, LÁP and KS contributed to the study design. ECE, RN, AR  
48  
49

50  
51 392 and AIK took part in and supervised data collection. PS, OS, and KS analysed the data. All authors contributed to the preparation  
52  
53

54 393 of the database, interpretation of the results and reviewed and approved the final version of the manuscript.  
55  
56  
57  
58

## 59 394 **Competing interests**

60 395 The authors declare that they have no competing interests

1  
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

## 396 **Data sharing statement**

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

## 400 **Funding**

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

## 406 **Acknowledgements**

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

1  
2  
3 410  
4  
5 411  
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

For peer review only



1  
2 413 **Legend to Figures**

3  
4  
5 414 **Figure 1.** Factors, variables and outcomes included in the analysis of data from the MINIMat cohort, Bangladesh. Grouping  
6  
7  
8 415 according to the WHO conceptual framework on childhood stunting [21]  
9

10  
11  
12 416 **Figure 2.** Flow chart of pregnant women and their children included in the data mining analyses of the MINIMat cohort from  
13  
14  
15  
16 417 conception to two years of age.  
17

18  
19  
20 418 **Figure 3.** Height-for-age Z-scores from birth to 24 months in the MINIMat cohort in rural Bangladesh.  
21

22  
23  
24 419 **Figure 4.** Conditional random forest plot ranking the relative importance of 30 predictors with regard to their ability to explain the  
25  
26  
27  
28 420 presence of stunting at 24 months of age. The MINIMat cohort in rural Bangladesh. Colour coding according to the WHO  
29  
30  
31 421 conceptual framework on causes of stunting.  
32

33  
34  
35 422 **Figure 5.** Conditional random forest plot ranking the relative importance of 30 predictors with regard to their ability to explain the  
36  
37  
38  
39 423 variation in change in HAZ ( $\Delta$  HAZ) from birth to 24 months of age. The MINIMat cohort in rural Bangladesh. Colour coding  
40  
41  
42 424 according to the WHO conceptual framework on causes of stunting.  
43

44  
45  
46 425 **Figure 6.** Conditional inference tree identifying sub-groups with different probabilities of stunting at 24 months. The MINIMat  
47  
48  
49  
50 426 cohort in rural Bangladesh.  
51

52  
53  
54 427 **Figure 7.** Conditional inference tree identifying sub-groups with different mean change in HAZ ( $\Delta$  HAZ = HAZ<sub>24</sub> - HAZ<sub>0</sub>) 0-24  
55  
56  
57 428 months within the MINIMat cohort in rural Bangladesh.  
58

59  
60  
429

1  
2 430  
3  
4  
5  
6  
7 431  
8  
9  
10  
11  
12 432 **References**

- 13 433 1 Pradhan M, Sahn DE, Younger SD. Decomposing world health inequality. *Journal of Health Economics*  
14 2003;**22**:271–93. doi:10.1016/S0167-6296(02)00123-6  
15 434  
16  
17  
18 435 2 Victora CG, Adair L, Fall C, *et al.* Maternal and child undernutrition: consequences for adult health and human  
19 capital. *Lancet* 2008;**371**:340–57. doi:10.1016/S0140-6736(07)61692-4  
20 436  
21  
22 437 3 Black RE, Victora CG, Walker SP, *et al.* Maternal and child undernutrition and overweight in low-income and  
23 middle-income countries. *The Lancet* 2013.  
24 438  
25  
26 439 4 General Assembly, UNICEF. Transforming our world: the 2030 Agenda for Sustainable Development. *United*  
27 Nations 2015;**1**–35.  
28 440  
29  
30  
31 441 5 Victora CG, de Onis M, Hallal PC, *et al.* Worldwide timing of growth faltering: revisiting implications for  
32 interventions. *Pediatrics* 2010;**125**:e473–80. doi:10.1542/peds.2009-1519  
33 442  
34  
35 443 6 Black RE, Allen LH, Bhutta ZA, *et al.* Maternal and child undernutrition: global and regional exposures and health  
36 consequences. *The Lancet* 2008;**371**:243–60. doi:10.1016/S0140-6736(07)61690-0  
37 444  
38  
39  
40 445 7 Bayer R, Galea S. Public Health in the Precision-Medicine Era. - PubMed - NCBI. *N Engl J Med* 2015;**373**:499–501.  
41 doi:10.1056/NEJMp1506241  
42 446  
43  
44 447 8 Martorell R, Ramakrishnan U, Schroeder DG, *et al.* Intrauterine growth retardation, body size, body composition  
45 and physical performance in adolescence. *Eur J Clin Nutr* 1998;**52 Suppl 1**:S43–52–discussionS52–3.  
46 448  
47  
48  
49 449 9 Frongillo EA Jr. *Symposium: Causes and Etiology of Stunting.* *J Nutr* 1999;**129**:529S–530S. doi:10.1016/S0140-  
50 450 6736(94)91337-4  
51  
52  
53 451 10 Cole TJ, Parkin JM. Infection and its effect on the growth of young children: A comparison of The Gambia and  
54 Uganda. *Trans R Soc Trop Med Hyg* 1977;**71**:196–8. doi:10.1016/0035-9203(77)90005-0  
55 452  
56  
57 453 11 Investigators M-EN. Childhood stunting in relation to the pre- and postnatal environment during the first 2 years  
58 of life: The MAL-ED longitudinal birth cohort study. *PLoS Medicine* 2017;**1**–21. doi:10.1371/journal.pmed.1002408  
59 454  
60

- 1  
2  
3 455 12 Espo M, Kulmala T, Maleta K, *et al.* Determinants of linear growth and predictors of severe stunting during  
4 456 infancy in rural Malawi. *Acta Paediatrica* 2002;**91**:1364–70. doi:10.1111/j.1651-2227.2002.tb02835.x  
5  
6  
7 457 13 Rahman A, Chowdhury S. Determinants of Chronic Malnutrition Among Preschool Children In Bangladesh. *J*  
8 458 *Biosoc Sci* 2006;**39**:161. doi:10.1017/S0021932006001295  
9  
10  
11 459 14 Amin R, Murumi MW, Carabaza AFM. Factors Associated with Stunting Among 0-23 Months-Old Children in  
12 Rural Bangladesh. *Journal of Nutrition Education and Behavior* 2017;**49**:S21. doi:10.1016/j.jneb.2017.05.061  
13 460  
14  
15  
16 461 15 Hong R, Banta JE, Betancourt JA. Relationship between household wealth inequality and chronic childhood under-  
17 462 nutrition in Bangladesh. *International Journal for Equity in Health* 2006;**5**:1–10. doi:10.1186/1475-9276-5-15  
18  
19  
20 463 16 Kim R, Mejía-Guevara I, Corsi DJ, *et al.* Relative importance of 13 correlates of child stunting in South Asia: Insights  
21 464 from nationally representative data from Afghanistan, Bangladesh, India, Nepal, and Pakistan. *Social Science &*  
22 465 *Medicine* 2017;**187**:144–54. doi:10.1016/j.socscimed.2017.06.017  
23  
24  
25  
26 466 17 Christian P, Lee SE, Donahue Angel M, *et al.* Risk of childhood undernutrition related to small-for-gestational age  
27 467 and preterm birth in low- and middle-income countries. *Int J Epidemiol* 2013;**42**:1340–55. doi:10.1093/ije/dyt109  
28  
29  
30  
31 468 18 World Health Organization. WHO Guideline: Daily iron and folic acid supplementation in pregnant women. *WHO*  
32 469 2012;;1–32.  
33  
34  
35 470 19 World Health Organization. Essential Nutrition Actions: improving maternal, newborn, infant and young child  
36 471 health and nutrition. *WHO* 2013;;1–144.  
37  
38  
39 472 20 Bhutta ZA, Das JK, Arjumand R, *et al.* Maternal and Child Nutrition 2 Evidence-based interventions for  
40 473 improvement of maternal and child nutrition: what can be done and at what cost? *The Lancet* 2013;**382**:452–77.  
41 474 doi:10.1016/S0140-6736(13)60996-4  
42  
43  
44  
45 475 21 Stewart CP, Iannotti L, Dewey KG, *et al.* Contextualising complementary feeding in a broader framework for  
46 476 stunting prevention. *Matern Child Nutr* 2013;**9**:27–45. doi:10.1111/mcn.12088  
47  
48  
49  
50 477 22 Breiman L, Friedman J, Stone CJ, *et al.* *Classification and Regression Trees*. Chapman and Hall/CRC 1984.  
51  
52  
53 478 23 Persson LÅ, Arifeen S, Ekström E-C, *et al.* Effects of Prenatal Micronutrient and Early Food Supplementation on  
54 479 Maternal Hemoglobin, Birth Weight, and Infant Mortality Among Children in Bangladesh: The MINIMat  
55 Randomized Trial. *JAMA* 2012;**307**:2050–9. doi:10.1001/jama.2012.4061  
56 480  
57  
58  
59 481 24 Khan AI, Hawkesworth S, Ekström E-C, *et al.* Effects of exclusive breastfeeding intervention on child growth and  
60 482 body composition: the MINIMat trial, Bangladesh. *Acta Paediatrica* 2013;**102**:815–23. doi:10.1111/apa.12282

- 1  
2 483 25 Gwatkin DR, Rustein S, Johnson K, *et al.* *Socio-economic differences in health nutrition and population in*  
3 484 *Bangladesh*. Washington, DC: HNP/Poverty Thematic Group of the World Bank 2000.  
4  
5  
6  
7 485 26 Frongillo EA, Chowdhury N, Ekström E-C, *et al.* Understanding the Experience of Household Food Insecurity in  
8 486 Rural Bangladesh Leads to a Measure Different from That Used in Other Countries  
9  
10  
11 487 . *J Nutr* 2003;:4158–62.  
12  
13  
14 488 27 Straus MA, Douglas EM. A Short Form of the Revised Conflict Tactics Scales, and Typologies for Severity and  
15 489 Mutuality. *Violence and Victims* 2004;19:1–15.  
16  
17  
18 490 28 WHO. WHO Multi-country Study on Women's Health and Domestic Violence against Women. *WHO* 2005;:1–38.  
19  
20  
21 491 29 Asling-Monemi K, Naved RT, Persson LA. Violence against women and the risk of fetal and early childhood growth  
22 492 impairment: a cohort study in rural Bangladesh. *Arch Dis Child* 2009;94:775–9. doi:10.1136/adc.2008.144444  
23  
24  
25 493 30 Rahman A, Vahter M, Smith AH, *et al.* Arsenic Exposure During Pregnancy and Size at Birth: A Prospective  
26 494 Cohort Study in Bangladesh. *American Journal of Epidemiology* 2008;169:304–12. doi:10.1093/aje/kwn332  
27  
28  
29  
30 495 31 Arifeen SE, Black RE, Caulfield LE, *et al.* Infant growth patterns in the slums of Dhaka in relation to birth weight,  
31 496 intrauterine growth retardation, and prematurity. *Am J Clin Nutr* 2000;72:1010–7.  
32  
33  
34 497 32 World Health Organization. Indicators for assessing breastfeeding practices. *WHO* 1991;:1–16.  
35  
36  
37 498 33 Moore SE, Prentice AM, Coward WA, *et al.* Use of stable-isotope techniques to validate infant feeding practices  
38 499 reported by Bangladeshi women receiving breastfeeding counseling. *Am J Clin Nutr* 2007;:1075–82.  
39  
40  
41 500 34 Kallioinen M, Ekström E-C, Khan AI, *et al.* Prenatal early food and multiple micronutrient supplementation trial  
42 501 reduced infant mortality in Bangladesh, but did not influence morbidity. *Acta Paediatrica* 2017;382:427–8.  
43 502 doi:10.1111/apa.14009  
44  
45  
46  
47 503 35 Arnold BF, Galiani S, Ram PK, *et al.* Optimal Recall Period for Caregiver-reported Illness in Risk Factor and  
48 504 Intervention Studies: A Multicountry Study. *American Journal of Epidemiology* 2013;177:361–70.  
49 505 doi:10.1093/aje/kws281  
50  
51  
52  
53 506 36 Onis M de, Onyango AW, Borghi E, *et al.* Development of a WHO growth reference for school-aged children and  
54 507 adolescents. *Bulletin of the World Health Organization* 2007;85:660–7. doi:10.1590/S0042-96862007000900010  
55  
56  
57  
58 508 37 Jönsson P, Wohlin C. An Evaluation of K-Nearest Neighbour Imputation Using Likert Data. *Proceedings of th*  
59 509 *International Symposium on Software Metrics* 2004;:108–18.  
60

- 1  
2 510 38 Stekhoven DJ, Buhlmann P. MissForest—non-parametric missing value imputation for mixed-type data.  
3 511 *Bioinformatics* 2012;;112–8.  
4  
5  
6  
7 512 39 Torsten Hothorn KHAZ. ctree: Conditional Inference Trees. www.cran.r-project.org. 2016;;1–21.https://cran.r-  
8 513 project.org/web/packages/partykit/vignettes/ctree.pdf (accessed 24 Mar2017) page 655.  
9  
10  
11 514 40 Hastie T, Tibshirani R, Friedman J. *The Elements of Statistical Learning*. New York, NY: : Springer Science &  
12 Business Media 2013.  
13 515  
14  
15 516 41 R Core Team. R: A language and environment for stastical computing. 2017. httpa://www.R-project.org/  
16  
17  
18 517 42 Achim Zeileis THKH. party: A Laboratory for Recursive Partytioning. 2017;;1–18.  
19  
20  
21 518 43 Christian P, Lee SE, Donahue Angel M, *et al.* Risk of childhood undernutrition related to small-for-gestational age  
22 and preterm birth in low- and middle-income countries. *Int J Epidemiol* 2013;42:1340–55. doi:10.1093/ije/dyt109  
23 519  
24  
25 520 44 Schmidt MK, Muslimatun S, West CE, *et al.* Nutritional Status and Linear Growth of Indonesian Infants in West  
26 Java Are Determined More by Prenatal Environment than by Postnatal Factors. *J Nutr* 2002;;2202–7.  
27 521  
28  
29  
30 522 45 Corsi DJ, Mejía-Guevara I, Subramanian SV. Risk factors for chronic undernutrition among children in India:  
31 Estimating relative importance, population attributable risk and fractions. *Social Science & Medicine* 2016;157:165–  
32 523 85. doi:10.1016/j.socscimed.2015.11.014  
33 524  
34  
35  
36 525 46 Kramer MS. The Epidemiology of Adverse Pregnancy Outcomes: An Overview. *J Nutr* 2003;;1592S–1596S.  
37  
38  
39 526 47 A WHO Collaborative Study of Maternal Anthropometry and Pregnancy Outcomes. *International Journal of*  
40 527 *Gynecology and Obstetrics* 1996;57:1–15.  
41  
42  
43 528 48 Özaltın E. Association of Maternal Stature With Offspring Mortality, Underweight, and Stunting in Low- to  
44 Middle-Income Countries. *JAMA* 2010;303:1507–21. doi:10.1001/jama.2010.450  
45 529  
46  
47 530 49 Svefors P, Rahman A, Ekström E-C, *et al.* Stunted at 10 Years. Linear Growth Trajectories and Stunting from Birth  
48 to Pre-Adolescence in a Rural Bangladeshi Cohort. *PLoS ONE* 2016;11:e0149700–18.  
49 531 doi:10.1371/journal.pone.0149700  
50 532  
51  
52  
53 533 50 Van de Poel E, Hosseinpoor AR, Speybroeck N, *et al.* Socioeconomic inequality in malnutrition in developing  
54 534 countries. *Bulletin of the World Health Organization* 2008;86:282–91. doi:10.1590/S0042-96862008000400013  
55 535  
56  
57  
58 535 51 Bradley RH, Corwyn RF. Socioeconomic Status and Child Development. *Annual Review of Psychology* 2002;53:371–  
59 536 99.  
60

- 1  
2 537 52 Ramalingaswami V, UJ, J R. *The Asian enigma*. 1996.
- 3  
4  
5 538 53 Shroff M, Griffiths P, Adair L, *et al*. Maternal autonomy is inversely related to child stunting in Andhra Pradesh,  
6 India. *Matern Child Nutr* 2009;**5**:64–74. doi:10.1111/j.1740-8709.2008.00161.x
- 7 539  
8  
9 540 54 Chowdhury AMR, Bhuiya A, Chowdhury ME, *et al*. The Bangladesh paradox: exceptional health achievement  
10 541 despite economic poverty. *The Lancet* 2013;**382**:1734–45. doi:10.1016/S0140-6736(13)62148-0
- 11  
12  
13 542 55 Asadullah MN, Chaudhury N. Girls school enrolment figures are good in Bangladesh but drop out rates are  
14 543 worrying. *Development and Cooperation* 2017;**59**:447–72. doi:10.1086/681929
- 15  
16  
17  
18 544 56 UNESCO. EFA 2015 national review: Bangladesh; 2014. unesdoc.unesco.org.  
19 545 <http://unesdoc.unesco.org/images/0023/002305/230507e.pdf> (accessed 13 Feb 2018).
- 20  
21  
22  
23 546 57 Checkley W, Buckley G, Gilman RH, *et al*. Multi-country analysis of the effects of diarrhoea on childhood stunting.  
24 547 *Int J Epidemiol* 2008;**37**:816–30. doi:10.1093/ije/dyn099
- 25  
26  
27 548 58 Koster FT, Curlin GC, Aziz KMA, *et al*. Synergistic impact of measles and diarrhoea on nutrition and mortality in  
28 549 Bangladesh. *Bulletin of the World Health Organization* 1981;:901–8.
- 29  
30  
31  
32 550 59 Arifeen DrPH El S, MPH AC, ScD LR, *et al*. Community-based approaches and partnerships: innovations in  
33 551 health-service delivery in Bangladesh. *The Lancet* 2013;**382**:2012–26. doi:10.1016/S0140-6736(13)62149-2
- 34  
35  
36 552 60 WHO, UNICEF. Bangladesh: WHO and UNICEF estimates of immunization coverage: 2016 revision. 2016;:1–30.
- 37  
38  
39 553 61 Chowdhury AMR, Bhuiya A, Chowdhury ME, *et al*. The Bangladesh paradox: exceptional health achievement  
40 554 despite economic poverty. *The Lancet* 2013;**382**:1734–45. doi:10.1016/S0140-6736(13)62148-0
- 41  
42  
43 555 62 Khan AI, Kabir I, Ekstrom EC, *et al*. Effects of prenatal food and micronutrient supplementation on child growth  
44 556 from birth to 54 months of age: a randomized trial in Bangladesh. *Nutr J* 2011.
- 45  
46  
47 557 63 Martorell R. Physical growth and development of the malnourished child: Contributions from 50 years of research  
48 558 at INCAP. *Food Nutr Bull* 2010;**31**:68–82.
- 49  
50  
51  
52 559 64 Imdad A, Yakoob MY, Bhutta ZA. Impact of maternal education about complementary feeding and provision of  
53 560 complementary foods on child growth in developing countries. *BMC Public Health* 2011;**11**:S25. doi:10.1186/1471-  
54 561 2458-11-S3-S25
- 55  
56  
57  
58 562 65 Weedon MN, Frayling TM. Reaching new heights: insights into the genetics of human stature. *Trends in Genetics*  
59 563 2008;**24**:595–603. doi:10.1016/j.tig.2008.09.006

- 1  
2 564 66 Silventoinen K, Sammalisto S, Perola M, *et al.* Heritability of Adult Body Height: A Comparative Study of Twin  
3 Cohorts in Eight Countries. *Twin Research* 2003;6:399–408. doi:10.1375/twin.6.5.399  
4 565  
5  
6  
7 566 67 Simeone P, Alberti S. Epigenetic heredity of human height. *Physiological Reports* 2014;2:e12047–7.  
8 doi:10.14814/phy2.12047  
9 567  
10  
11 568 68 Silventoinen K, Kaprio J, Lahelma E, *et al.* Relative Effect of Genetic and Environmental Factors on Body Height:  
12 Differences Across Birth Cohorts Among Finnish Men and Women. *American Journal of Public Health*  
13 569 2000;90:627–30.  
14  
15 570  
16  
17 571 69 Imdad A, Bhutta ZA. Effect of balanced protein energy supplementation during pregnancy on birth outcomes.  
18 *BMC Public Health* 2011;11:S17. doi:10.1186/1471-2458-11-S3-S17  
19 572  
20  
21  
22 573 70 Fall CHD, Fisher DJ, Osmond C, *et al.* Multiple Micronutrient Supplementation during Pregnancy in Low-Income  
23 Countries: A Meta-Analysis of Effects on Birth Size and Length of Gestation. *Food Nutr Bull* 2009;30:S533–46.  
24 574 doi:10.1177/15648265090304S408  
25 575  
26  
27  
28 576 71 Ramakrishnan U, Grant FK, Goldenberg T, *et al.* Effect of Multiple Micronutrient Supplementation on Pregnancy  
29 and Infant Outcomes: A Systematic Review. *Paediatric and Perinatal Epidemiology* 2012;26:153–67.  
30 577 doi:10.1111/j.1365-3016.2012.01276.x  
31 578  
32  
33  
34 579 72 West KP, Shamim AA, Mehra S, *et al.* Effect of Maternal Multiple Micronutrient vs Iron–Folic Acid  
35 Supplementation on Infant Mortality and Adverse Birth Outcomes in Rural Bangladesh: The JiVitA-3 Randomized  
36 580 Trial. *JAMA* 2014;312:2649–58. doi:10.1001/jama.2014.16819  
37 581  
38  
39  
40 582 73 Johnson W, Darboe MK, Sosseh F, *et al.* Association of prenatal lipid-based nutritional supplementation with fetal  
41 growth in rural Gambia. *Matern Child Nutr* 2017;13:e12367. doi:10.1111/mcn.12367  
42 583  
43  
44  
45 584 74 Ashorn P, Alho L, Ashorn U, *et al.* Supplementation of Maternal Diets during Pregnancy and for 6 Months  
46 585 Postpartum and Infant Diets Thereafter with Small-Quantity Lipid-Based Nutrient Supplements Does Not  
47 Promote Child Growth by 18 Months of Age in Rural Malawi: A Randomized Controlled Trial. *J Nutr*  
48 586 2015;145:1345–53. doi:10.3945/jn.114.207225  
49 587  
50  
51  
52 588 75 Adu-Afarwuah S, Lartey A, Okronipa H, *et al.* Lipid-based nutrient supplement increases the birth size of infants  
53 of primiparous women in Ghana. *Am J Clin Nutr* 2015;101:ajcn.091546–846. doi:10.3945/ajcn.114.091546  
54 589  
55  
56  
57 590 76 Huybregts L, Roberfroid D, Lanou H, *et al.* Prenatal food supplementation fortified with multiple micronutrients  
58 591 increases birth length: a randomized controlled trial in rural Burkina Faso. *American Journal of Clinical Nutrition*  
59 2009;90:1593–600. doi:10.3945/ajcn.2009.28253  
60 592

- 1  
2 593 77 Gluckman PD, Hanson MA, Cooper C, *et al.* Effect of In Utero and Early-Life Conditions on Adult Health and  
3 594 Disease. <http://dxdoiorg/101056/NEJMrao708473> 2009;**359**:61–73. doi:10.1056/NEJMrao708473  
4  
5  
6  
7 595 78 Potdar RD, Sahariah SA, Gandhi M, *et al.* Improving women's diet quality preconceptionally and during gestation:  
8 596 effects on birth weight and prevalence of low birth weight—a randomized controlled efficacy trial in India (Mumbai  
9 597 Maternal Nutrition Project). *American Journal of Clinical Nutrition* 2014;**100**:1257–68. doi:10.3945/ajcn.114.084921  
10  
11  
12  
13 598 79 The World Bank. World Development Indicators: Reproductive health. 2016;:1–6.  
14  
15  
16 599 80 Wills AK, Chinchwadkar MC, Joglekar CV, *et al.* Maternal and paternal height and BMI and patterns of fetal  
17 600 growth: The Pune Maternal Nutrition Study. *Early Human Development* 2010;**86**:535–40.  
18 601 doi:10.1016/j.earlhumdev.2010.07.002  
19  
20  
21  
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



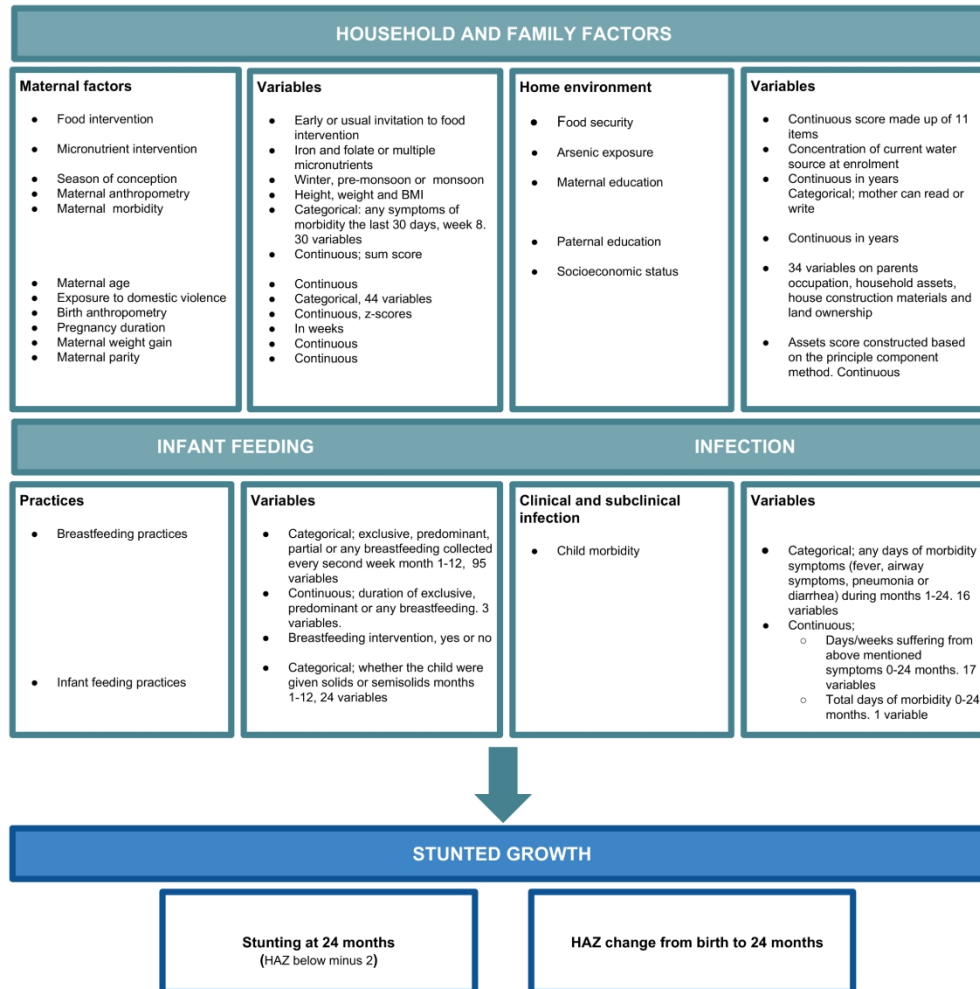
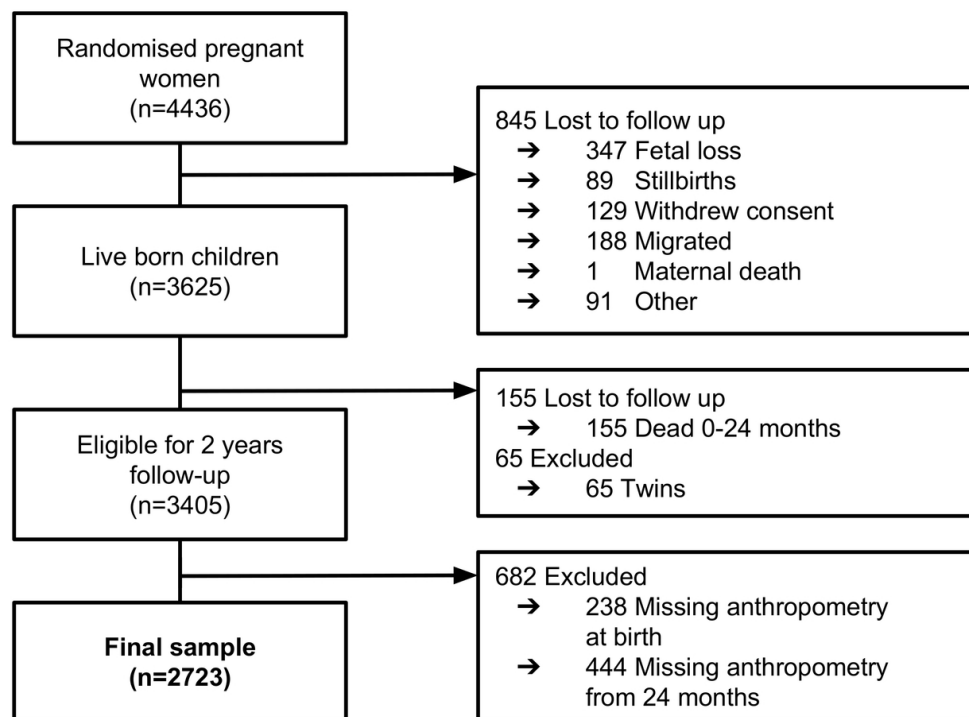


Figure 1. Factors, variables and outcomes included in the analysis of data from the MINIMat cohort, Bangladesh. Grouping according to the WHO conceptual framework on childhood stunting [20]

190x190mm (300 x 300 DPI)



31 Figure 2. Flow chart of pregnant women and their children included in the data mining analyses of the  
32 MINIMat cohort from conception to two years of age.  
33

34 106x80mm (300 x 300 DPI)  
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  
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

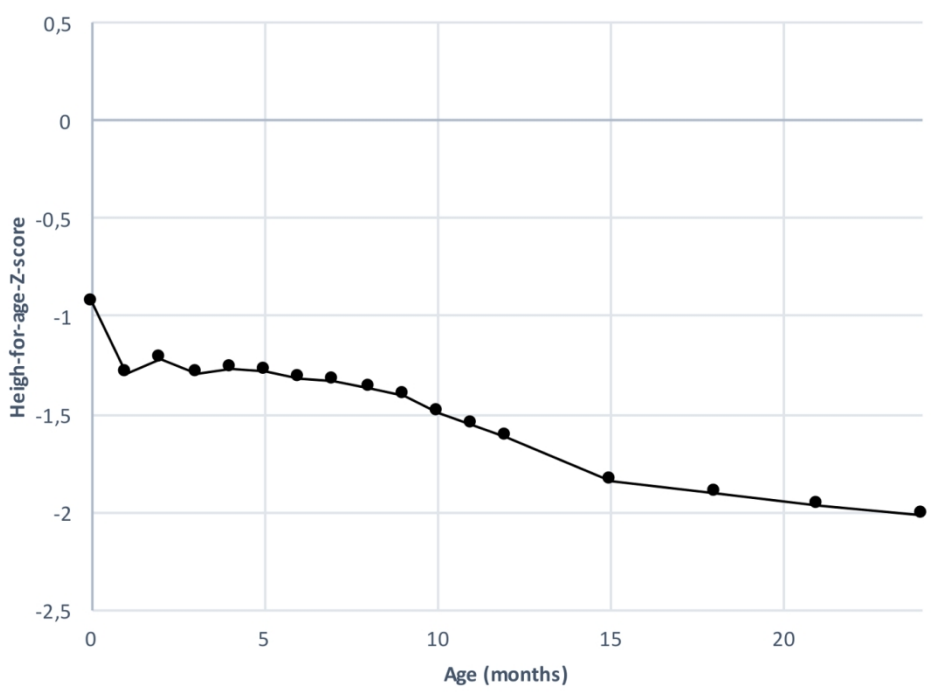


Figure 3. Height-for-age Z-scores from birth to 24 months in the MINIMat cohort in rural Bangladesh.

131x96mm (300 x 300 DPI)

1  
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

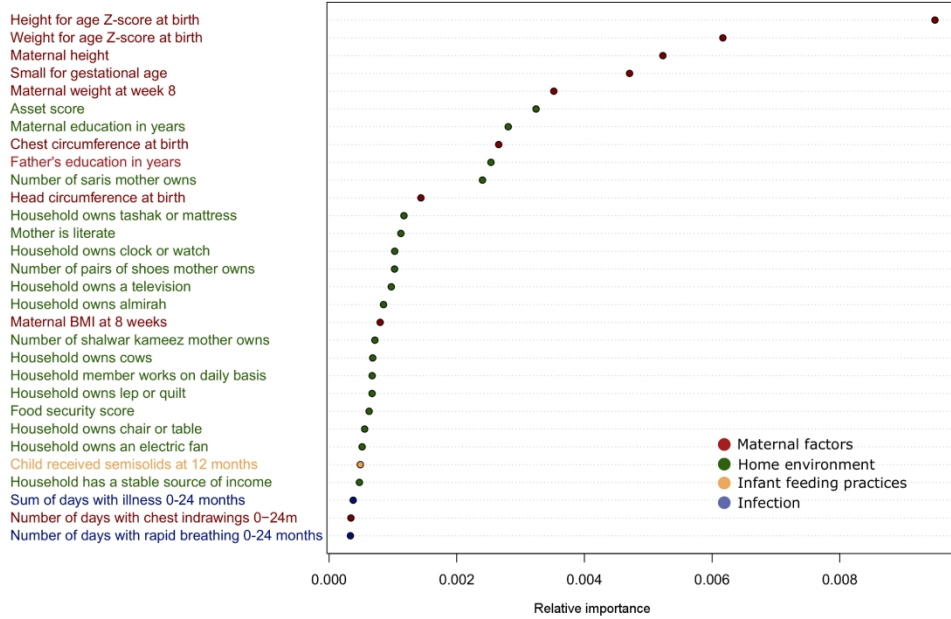


Figure 4. Conditional random forest plot ranking the relative importance of 30 predictors with regard to their ability to explain the presence of stunting at 24 months of age. The MINIMat cohort in rural Bangladesh. Colour coding according to the WHO conceptual framework on causes of stunting.

190x134mm (300 x 300 DPI)

1  
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

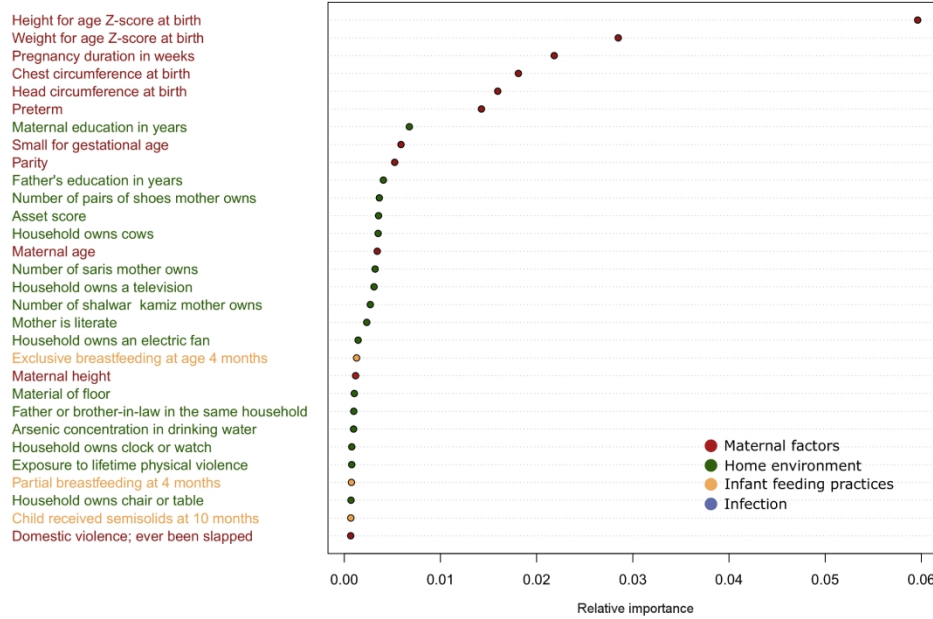


Figure 5. Conditional random forest plot ranking the relative importance of 30 predictors with regard to their ability to explain the variation in change in HAZ ( $\Delta$  HAZ) from birth to 24 months of age. The MINIMat cohort in rural Bangladesh. Colour coding according to the WHO conceptual framework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

190x134mm (300 x 300 DPI)

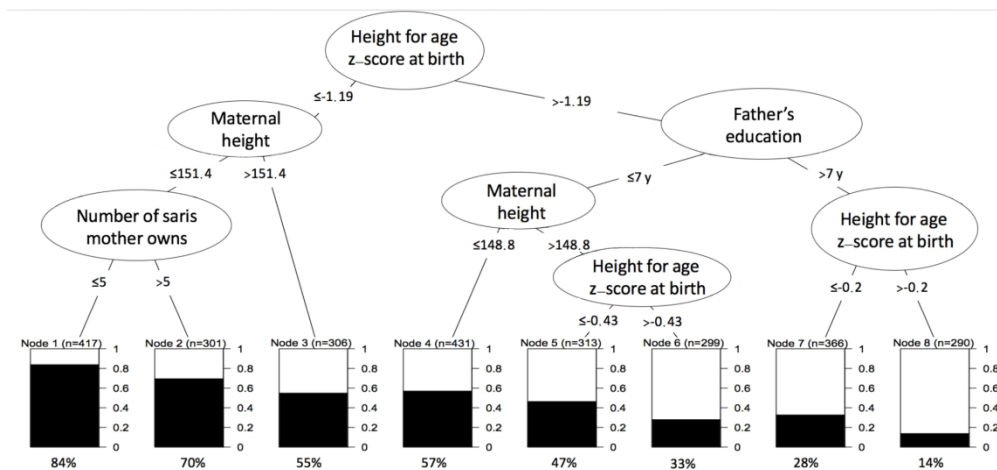


Figure 6. Conditional inference tree identifying sub-groups with different probabilities of stunting at 24 months. The MINIMat cohort in rural Bangladesh.

190x88mm (300 x 300 DPI)

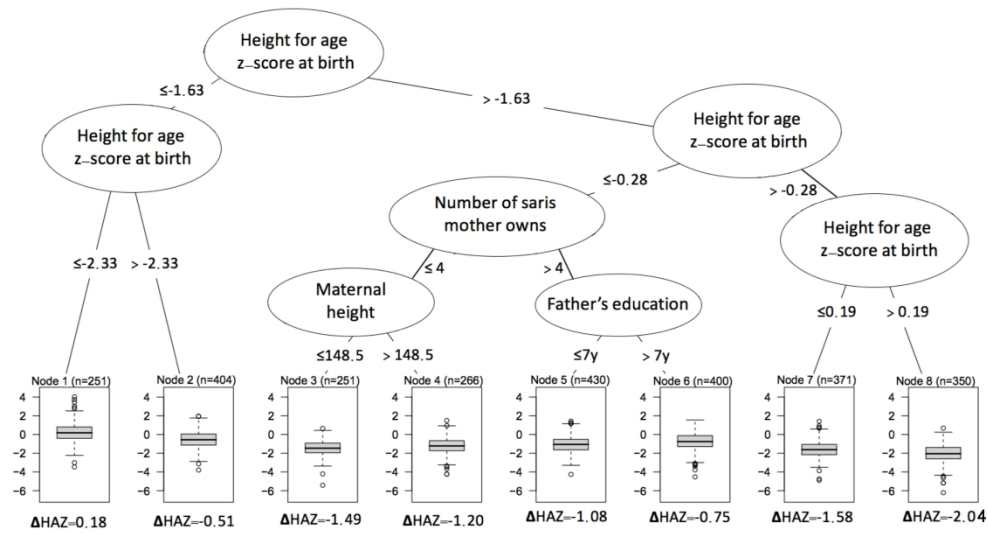


Figure 7. Conditional inference tree identifying sub-groups with different mean change in HAZ ( $\Delta\text{HAZ} = \text{HAZ}_{24} - \text{HAZ}_0$ ) 0–24 months within the MINIMat cohort in rural Bangladesh.

193x101mm (300 x 300 DPI)

## Supplementation appendix

### Simulation study of the predictive performance of three different imputation methods

The following strategy was used to study the imputation accuracy of various methods for the input variables in our analyses. First, we standardized numerical variables in the data and took a sample of the entire data ( $\alpha$ ) 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> Misclassification rate

<sup>3</sup> Mean square error

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

$\beta$  = proportion of the original data sampled



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

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

<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  $MR^2$  and the  $MSE^3$  for different imputation methods, computed from  $m=100$  samples,  $\alpha = 0.2$ ,  $\beta = 0.05$

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

<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, Bühlmann P. MissForest—non-parametric missing value imputation for mixed-type data. *Bioinformatics*. 2012;: 112–118.
2. Jönsson P, Wohlin C. An Evaluation of K-Nearest Neighbour Imputation Using Likert Data. *Proceedings of the 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	
<b>Title and abstract</b>	1	(a) Indicate the study's design with a commonly used term in the title or the abstract	1
		(b) Provide in the abstract an informative and balanced summary of what was done and what was found	2
<b>Introduction</b>			
Background/rationale	2	Explain the scientific background and rationale for the investigation being reported	4-5
Objectives	3	State specific objectives, including any prespecified hypotheses	6
<b>Methods</b>			
Study design	4	Present key elements of study design early in the paper	5
Setting	5	Describe the setting, locations, and relevant dates, including periods of recruitment, exposure, follow-up, and data collection	6
Participants	6	(a) <i>Cohort study</i> —Give the eligibility criteria, and the sources and methods of selection of participants. Describe methods of 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	6
		(b) <i>Cohort study</i> —For matched studies, give matching criteria and number of exposed and unexposed <i>Case-control study</i> —For matched studies, give matching criteria and the number of controls per case	
Variables	7	Clearly define all outcomes, exposures, predictors, potential confounders, and effect modifiers. Give diagnostic criteria, if applicable	9 Figure 1
Data sources/ measurement	8*	For each variable of interest, give sources of data and details of methods of assessment (measurement). Describe comparability of assessment methods if there is more than one group	7-9
Bias	9	Describe any efforts to address potential sources of bias	7-9
Study size	10	Explain how the study size was arrived at	Not applicable
Quantitative variables	11	Explain how quantitative variables were handled in the analyses. If applicable, describe which groupings were chosen and why	7-9
Statistical methods	12	(a) Describe all statistical methods, including those used to control for confounding	10-11
		(b) Describe any methods used to examine subgroups and interactions	10-11
		(c) Explain how missing data were addressed	10-11
		(d) <i>Cohort study</i> —If applicable, explain how loss to follow-up was addressed <i>Case-control study</i> —If applicable, explain how matching of cases and controls was addressed	10

1  
2 *Cross-sectional study*—If applicable, describe analytical  
3 methods taking account of sampling strategy

---

4 (g) Describe any sensitivity analyses

10

5 Continued on next page10  
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  
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**Results**

Participants	13*	(a) Report numbers of individuals at each stage of study—eg numbers potentially eligible, examined for eligibility, confirmed eligible, included in the study, completing follow-up, and analysed	Fig 2 12
		(b) Give reasons for non-participation at each stage	Fig 2
		(c) Consider use of a flow diagram	Figure 2
Descriptive data	14*	(a) Give characteristics of study participants (eg demographic, clinical, social) and information on exposures and potential confounders	Table 1, 13
		(b) Indicate number of participants with missing data for each variable of interest	13
		(c) <i>Cohort study</i> —Summarise follow-up time (eg, average and total amount)	13
Outcome data	15*	<i>Cohort study</i> —Report numbers of outcome events or summary measures over time	13
		<i>Case-control study</i> —Report numbers in each exposure category, or summary measures of exposure	
		<i>Cross-sectional study</i> —Report numbers of outcome events or summary measures	
Main results	16	(a) Give unadjusted estimates and, if applicable, confounder-adjusted estimates and their precision (eg, 95% confidence interval). Make clear which confounders were adjusted for and why they were included	14
		(b) Report category boundaries when continuous variables were categorized	14
		(c) If relevant, consider translating estimates of relative risk into absolute risk for a meaningful time period	
Other analyses	17	Report other analyses done—eg analyses of subgroups and interactions, and sensitivity analyses	14

**Discussion**

Key results	18	Summarise key results with reference to study objectives	15,18
Limitations	19	Discuss limitations of the study, taking into account sources of potential bias or imprecision. Discuss both direction and magnitude of any potential bias	15,16,17
Interpretation	20	Give a cautious overall interpretation of results considering objectives, limitations, multiplicity of analyses, results from similar studies, and other relevant evidence	18-22
Generalisability	21	Discuss the generalisability (external validity) of the study results	17

**Other information**

Funding	22	Give the source of funding and the role of the funders for the present study and, if applicable, for the original study on which the present article is based	23
---------	----	---	----

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

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