Modelling Under-Five Mortality through Multilevel Structured Additive Regression with Varying Coefficients for Asia and Sub-Saharan Africa

KENNETH HARTTGEN*, STEFAN LANG**, JUDITH SANTER** & JOHANNES SEILER***
*Department of Humanities, Social and Political Sciences, ETH Zurich, Zurich, Switzerland, **Department of Statistics, University of Innsbruck, Innsbruck, Austria

(Original version submitted November 2017; final version accepted December 2018)

ABSTRACT Despite improvements in global child health during the last three decades, under-five mortality rates remain significantly high in sub-Saharan Africa and Asia. Both regions did not achieve the MDG target of reducing under-five mortality rates by two thirds by 2015. The underlying causes of under-five mortality differ significantly between countries and between regions, which highlights the need to expand our understanding of the determinants of child health in developing countries. By comparing the two geographic regions of the world with the highest under-five mortality rates, we aim to identify differences between the determinants of under-five mortality in these regions. We analyse a large sample of DHS data sets consisting of 35 sub-Saharan African countries and 13 Asian countries from 1992 to 2016. Using a discrete-time survival model that takes advantage of a recently developed multilevel framework in a Bayesian setting, allowing for important non-linear effects and cluster specific heterogeneity. We find strong non-linear effects for the baseline hazard, the household size, the year of birth, and the mother’s BMI. We find considerable differences in determinants between Asian and sub-Saharan African countries. This highlights the necessity to expand our current knowledge of the underlying mechanisms, and helps to formulate policy advices.

1. Introduction

Child health has been improved throughout the last three decades in low- and middle-income countries around the globe. Most regions experienced a significant decline in under-five mortality rates since 1990 (You, New, & Wardlaw, 2012, 2014, 2015). Since 2000, most regions experienced an accelerated decline in under-five mortality compared to the period 1990–2000. From the period 1990–2000 to the period 2000–2015, the annual reduction rate increased in sub-Saharan Africa from 1.6 per cent to 4.1 per cent, in Central Asia from 1.4 per cent to 4.6 per cent, and in Southern Asia from 3.2 per cent to 3.6 per cent. Despite this significant progress towards achieving Millennium Development Goal (MDG) four, high under-five mortality rates are still concentrated in these regions, where rates remain higher than targeted by the MDG for 2015 (You, New, et al., 2015). Therefore, analysing and understanding the link between health outcomes of children and
determinants of under-five mortality is important for future policies and aid. A better understanding of this relationship, and applying this knowledge to future policies, can effectively help to achieve the newly formulated Sustainable Development Goals\(^1\) (SDG).

The empirical and theoretical literature on studying determinants of child health in general, and under-five mortality in particular, has identified the following factors of child wellbeing: individual and household characteristics, environmental conditions, and community or regional conditions. These characteristics directly and indirectly determine proximate or intermediate input variables for child health outcomes (Mosley & Chen, 1984; Schultz, 1984; Wolpin, 1997). Based on these considerations, a broad stream of empirical literature exists, which uses regression methods to analyse survival outcomes of infants, and children under a certain age (for example 36 months, or 60 months) in low- and middle-income countries, as it can be seen as a reverse indicator of child health (see for instance Boco, 2010; Hobcraft, McDonald, & Rutstein, 1985; Houweling, Kunst, Looman, & Mackenbach, 2005; Kandala & Ghilagaber, 2006; Liu et al., 2016; Harttgen & Misselhorn, 2006; Rutstein, 2000, 2005; Wang, 2003; You, Jones, Hill, Wardlaw, & Chopra, 2010). The findings of these studies highlight the importance of socio-economic and household specific factors related to under-five mortality and child mortality in general, which are confirmed in our analysis based on 48 countries.

One shortcoming of the existing literature is that most studies that analyse individual level data focus on single countries, or a small group of countries within a specific geopolitical region (for example Adebayo & Fahrmeir, 2005; Ayele, Zewotir, & Mwambi, 2015; Chamurbagwala, 2010). Limited empirical evidence exists on the association between child health and household, environmental, and socio-economic characteristics using pooled household survey data from several countries and years, while simultaneously considering the clustered structure of the survey data. Studies that account for the clustered data structure are for example Kuate-Defo and Diallo (2002) or Boco (2010).

A comparison across countries and geographic regions is even regarded as not helpful by some authors (for example Black, Morris, & Bryce, 2003), due to the different causes and different magnitude of the effects on under-five mortality. However, the increased availability of standardised household survey data across countries in recent years, as well as recent theoretical advances in hierarchical modelling in a Bayesian framework (Belitz et al., 2015; Lang, Umlauf, Wechselberger, Harttgen, & Kneib, 2014) allow to jointly analyse under-five mortality in Asia and sub-Saharan Africa at the individual level, while accounting for the clustered data structure and regional or country specific heterogeneity.

In addition, most existing studies present linear estimates of determinants of child health, which implies that the effect is independent of the level of the determinant. Assuming a non-linear relationship is more precise, and the range of the determinant in which the biggest effect on child health can be excepted is depicted. Underlying mechanisms of health transmission suggest that it is important to allow for non-linearities. For example, Harttgen, Lang, and Santer (2015) find strong non-linearities in the relationship between under-five mortality and its determinants. Several studies for individual countries identified non-linearities in the relationship between under-five mortality and its determinants. For instance, Ayele et al. (2015) find for an Ethiopian sample besides spatial differences, non-linear covariate effects. Black et al. (2013) shows that the effect of the nutritional status of the mother is associated with the mortality risk of the child in a non-linear way. Similarly, for the age of the mother a non-linear pattern is observed, too. This pattern is described and measured for instance by Harttgen et al. (2015), Kandala and Ghilagaber (2006), and Kandala et al. (2014), reiterating the importance of considering non-linear covariate effects.

The main objective of this study is to systematically analyse under-five mortality, by comparing determinants across geopolitical regions, and by allowing for non-linear covariate effects using a multilevel survival model. We aim at better understanding why under-five mortality decreased in such great variation between low- and middle-income countries in sub-Saharan Africa and Asia, and why high rates of under-five mortality remain in many countries. In doing so, we contribute to this strand of literature by aiming to close four shortcomings in the existing literature. First, we analyse a large dataset of standardised cross-country household surveys to study determinants of under-five mortality, while taking into account the hierarchical structure of the micro-data. Second, we explicitly...
account for differences in the determinants of under-five mortality between geopolitical regions, as well as differences occurring between and within countries. Third, we take into account possible non-linearities in the determinants of under-five mortality. Fourth, we employ a discrete survival model to consider the most recent period of five years preceding the survey.

We find strong non-linear covariate effects for child and maternal characteristics, as well as for socio-economic characteristics on under-five mortality. Besides considerable differences in the determinants between the two analysed geopolitical regions, considerable differences between countries are revealed when allowing for cluster specific heterogeneity.

The remainder of the paper is organised as follows: In Section 2, we provide an overview of the used dataset and a description of the covariates. In Section 3, we describe the methodology. In Section 4, we present the results. Finally, in Section 5 we conclude.

2. Data

2.1. Data source and data limitations

2.1.1. Demographic and Health Surveys. Data for this analysis is taken from Demographic and Health Surveys (DHS), which is collected by Macro International Inc., Calverton, Maryland, and local administration and are funded by USAID. The data is collected in waves, with currently the seventh wave ongoing, with about a period of normally five years between each wave. DHS are nationally representative and standardised across countries and provide information on child health, infant and child mortality, education, and maternal health, household’s durables, and quality of the household’s sanitation and dwelling.

The study contains information on 836,286 children from 48 countries which were sampled in 133 individual surveys over the period from 1992 to 2016. This data is complemented by per capita GDP taken from the Penn World Table (PWT) (Version 9.0) (Feenstra, Inklaar, & Timmer, 2015). The sample consists of 13 Asian and 35 sub-Saharan African countries, which can further be divided into 358 distinct regions. The left panel of Figure 1 depicts information about the included countries, and countries that could not be included.

2.1.2. Limitations. Conducting this analysis is associated with limitations that should be recognised and are primarily caused by data restrictions.

First, the DHS data do not provide information on important determinants for deceased children, implying that these variables could not be included into our estimation at all, or at the individual level. For instance, this is the case for the nutritional status, which is one of the main underlying causes of under-five mortality found by several authors (for example Black et al., 2003, 2013; Pelletier & Frongillo, 2003). DHS provides information about the nutritional status only for those children who were still living on the day the interview was conducted. This allows us to only include the regional mean of the standardised anthropometric indicator. This aggregation gives rise to a potential bias due to the close relationship of undernutrition and under-five mortality. It can be assumed that the effect of undernutrition on the risk of dying is underestimated by using the regional aggregate.

Second, questionnaires are prone to problems in the data collecting process. Missing values can be present in the data, as the respondent fails to answer a particular question. For instance, respondents could not remember the birth weight of their children, causing many missing values, or fatal incidences, especially when laying further back in the past are under reported within sampled households.

Third, information on HIV prevalence and malaria incidence rates is too scarce to be included in the analysis. Due to this, under-five mortality for countries with high HIV or malaria incidence rates will be underestimated. Further research needs to control for HIV and malaria prevalence, however this requires a significant improvement of data availability.
2.2. Dependent variable

Analysing under-five mortality is equivalent to modelling the probability of dying between birth and the fifth birthday (Rutstein & Rojas, 2006). Information on under-five mortality comes from the birth history section of the DHS, which offer basic information for deceased children (for example sex, date of birth, and date of death) and more detailed information for living children.

Two approaches can be considered for this analysis: the first approach is to use binary regression models. Using this type of model, however, restricts the sample to children fully exposed to the risk of dying before the fifth birthday. This corresponds to the under-five mortality rates for the interval five to nine years preceding the survey. Birth history data for this period, for instance, due to misreporting of the date of birth and the age at death, of dead children, are thought to be less precise, and can lead to an underestimation of the corresponding mortality rates (Korenromp, Arnold, Williams, Nahlen, & Snow, 2004). The second approach is to use survival models. This type of model allows to consider the more recent period of five years preceding the survey, and due to this is thought to be more precise and will be used in our analysis.

Noticeable is that the observed mortality rate for sub-Saharan Africa (mortality rate, per 1,000 children: 114) is almost twice as high compared to the observed mortality rate of Asia (mortality rate, per 1,000 children: 72), an observation already pointed by Klasen (2000) and Harttgen and Misselhorn (2006).

2.3. Independent variables

The explanatory variables used in the analysis can be grouped into child specific factors, maternal characteristics, household specific and demographic factors, and environmental and socio-economic characteristics. For the choice of covariates entering the model, we follow the theoretical framework of analysing the underlying causes of child health in developing countries proposed by Mosley and Chen (1984). This framework builds the basis for many empirical studies that have analysed child health and child mortality, particularly in developing countries (see for instance Adebayo & Fahrmeir, 2005; Black et al., 2003; Boco, 2010; Rutstein, 2000; Rutstein & Rojas, 2006).
2.3.1. Child specific factors. Child specific determinants include, amongst other covariates, a categorical covariate for the sex of the child, to control for a potential gender gap. The risk of dying is expected to be not evenly distributed over the whole age range from 0 to 59 months. Thus, the effect of the age is captured in the baseline hazard. The birth order is included, as a well-established relationship between the birth order and the risk of dying exists (see for instance Harttgen et al., 2015; Hobcraft et al., 1985; Rutstein, 2005). To complete the child specific factors at the individual level, a categorical indicator whether an older sibling of child \(i\) died is included into our analysis.

The children’s nutritional status is also found to be a major cause of child mortality (see for instance, Black et al., 2003; Pelletier & Frongillo, 2003), the nutritional status is most commonly assessed using standardised anthropometric measures. Information on the child’s anthropometric status is only available for non-deceased children, and it can only be included into the model at the regional level as regional aggregate. Another factor associated with child health is whether the child received vaccinations against diseases such as measles, polio, and tetanus within their first year. Unfortunately, information on vaccinations is only observed for non-deceased children, hence, it is not possible to control for the immunisation status at the child level.

2.3.2. Maternal characteristics. Maternal factors include the mother’s nutritional status, measured as the body mass index (BMI), the age of the mother, which is included as time-varying covariate, the spacing between two births, as well as the mother’s years of education.

Several studies found these covariates to be an important determinant of the risk of dying (see for example, Adebayo & Fahrmeir, 2005; Bhalotra & Van Soest, 2008; Black et al., 2013). For instance, a causal relationship was found between maternal education and the risk of dying (see for instance, Breierova & Duflo, 2004; Grépin & Bharadwaj, 2015; Makate & Makate, 2016) and the spacing between two births was found to influence the survival chances in the first five years (Mosley & Chen, 1984; Rutstein, 2000).

In our model, we include all maternal factors except for the spacing between two births, as this cannot be observed for first born children. However, as a robustness check, the final model is re-estimated including this factor.

2.3.3. Household specific characteristics. A household’s material wealth is positively related with the health status of the children living in this household (Fink et al., 2017; Vollmer, Harttgen, Kupka, & Subramanian, 2017). To measure the wealth of a household, an asset index is included into the analysis. This index is derived through a principal components analysis based on assets (for example television, fridge, or motorbike) possessed by household \(i\) and dwelling and sanitation characteristics of the household. To capture the effect of the size of the household, the total number of household members is included in the model.

2.3.4. Environmental and socio-economic factors. To control for differences between rural and urban areas, a categorical covariate is included to capture these potential differences. To control for the countries’ economic wealth, the per capita GDP of a given year is included into the analysis.

Declining mortality rates are observed in the last three decades (Lozano et al., 2011), to control for this time trend, the year of birth is included in the analysis. This trend is also illustrated in the right panel of Figure 1 that shows the trend in under-five mortality by birth year and geopolitical region.

3. Modelling under-five mortality

The statistical analysis is conducted using the open source statistical software BayesX (Belitz et al., 2015). Results are processed using the statistical software R (R Core Team, 2016) and the corresponding packages BayesX (Kneib, Heinzl, Brezger, Sabanés Bové, & Klein, 2014) and R2BayesX (Umlauf, Adler, Kneib, Lang, & Zeileis, 2015).
3.1. Survival model and expected effects

3.1.1. Discrete-time survival model. Discrete survival models can be specified as regression models with binary response, for instance as logit or probit model. To make an estimation feasible, the data needs to be augmented.\textsuperscript{6}

DHS provides monthly information on the age of the child at the date of the interview, information on the duration is accordingly measured on a discrete time scale, with duration time interval $T \in \{1, 2, ..., q; \ q = 22\}$.\textsuperscript{7} Besides the duration time $T$, the vector of covariates $x_t$ is observed, then $x_t = (x_{t1}, x_{t2}, ..., x_{tn})$ denotes the information at each interval up to interval $t$. Following Adebayo and Fahrmeir (2005), the discrete hazard function can be written as $\lambda(t; x_t) = P(T = t | T \geq t, x_t)$, with $t = 1, ..., 22$. $\lambda(t; x_t)$ is the discrete hazard function rewritten as the conditional probability of death in interval $t$, given child $i$ survived until interval $t$.

Information about the survival status is recorded using the binary event indicator $U_{it}$, with $U_{it} = 0$, if child $i$ is alive in interval $t$, and $U_{it} = 1$, if child $i$ died in interval $t$. In the former case, $t_i$ corresponds to the current age interval of child $i$ at the interview, and in the latter case $t_i$ corresponds to the observed age interval the child died. The underlying survival process of child $i$ can be interpreted as a series of dichotomous choices at each interval $t$. Assuming a probit link for the discrete-time survival model, the response function is given by $\pi_{it} = \phi(\eta_{it})$, where $\phi(\eta_{it})$ is the cumulative distribution function of the standard normal distribution and the link function is given by $\phi^{-1}(\pi_{it}) = \eta_{it}$ (for details see Fahrmeir, Kneib, Lang, & Marx, 2013, Chapter 5). Assuming for the continuous covariates a non-linear relationship the additive predictor $\eta_{it}$, is given by:

$$\eta_{it} = f_0(\text{Baseline hazard}) + f_1(\text{Asset index}_i) + f_2(\text{Year of birth}_i) + f_3(\text{Education mother}_i) + f_4(\text{Household size}_i) + f_5(\text{BMI mother}_i) + f_6(\text{Age mother}_it, \text{Birth order}_i) + f_7(\text{Stunting regional mean}_i) + f_8(\text{Regional immunization rate}_i) + f_9(\text{GDP}_it) + \gamma_1(\text{Older dead sibling}_i) + \gamma_2(\text{Gender}_i) + \gamma_3(\text{Place of living}_i) + \gamma_4(\text{Continent}_i) = f_0(\text{Baseline hazard}) + f_1(\text{Asset index}_i) + ... + f_9(\text{GDP}_it) + x'_{it}, \tag{1}$$

where $f_0$ (Baseline hazard) captures the potential non-linear effect of the children’s age, $f_1$ to $f_9$ are the smooth effects of the continuous covariates that enter the model with an assumed non-linear relationship, and $x'_{it}$ subsumes the effect coded categorical covariates included in the analysis.

3.1.2. Expected effects. Table 1 summarises the included covariates and the expected effects based on previous findings in the literature. The following covariates enter the model as categorical covariates: whether an older sibling of child $i$ is deceased, the sex of the child, and the place of residence. To detect potential differences between Asian and sub-Saharan African countries a binary indicator is included as the interaction variable in the varying coefficient terms.

All other variables enter the model as continuous covariates: the deviation from the regional mean of the asset index, the birth order of child $i$ in the household, the years of education of the mother, the numbers of people living in the household, the mother’s age, the mother’s BMI, and to capture a trend over time the birth year is included at the individual level. At the regional and country level the analysis includes the share of fully vaccinated children, and the mean height-for-age z-score within a particular region, and the country’s GDP. Descriptive statistics for the sample, and the geopolitical regions are presented in Table A2. Table A1 provides information about mortality rates per country and survey year.
Table 1. Factors of child mortality and their relationship found in the literature

| Variable | Description | Author | Data source | Mortality variable | Level | Relationship |
|----------|-------------|--------|-------------|--------------------|-------|--------------|
| **Child specific factors** | | | | | | |
| Dead sibling | Neonatal mortality of previous sibling | Bhalotra and Van Soest (2008) | NFHS India 1998/99 | Neonatal | Child | Positive |
| Index child | Harttgen et al. (2015) | DHS 25 countries SSA | Under-5 (U-5) | Child | Positive |
| Sex of child | Binary indicator | Chamarbagwala (2010) | NFHS-III India | not further specified | Child | Depending on gender composition |
| Baseline hazard | Age in months | Adebayo and Fahrmeir (2005) | DHS Nigeria 1999 | U-5 | Child | Non-linear decreasing |
| Birth order | Birth order in household | Ayele et al. (2015) | DHS Ethiopia 2011 | U-5 | Child | U-shaped |
| Nutritional status | Z-score underweight | Pelletier and Frongillo (2003) | DHS 62 countries | U-5 | Child | Linear decreasing |
| | Regional mean stunting | Harttgen et al. (2015) | DHS 25 countries SSA | U-5 | Child | Non-linear decreasing |
| Vaccination coverage | Share of fully immunised children | Boco (2010) | DHS for 28 countries in SSA | U-5 | Child | Decreasing |
| **Maternal specific characteristics** | | | | | | |
| Age mother | Age mother at birth | Adebayo and Fahrmeir (2005) | DHS Nigeria 1999 | 0–35 months | Child | Linear increasing |
| | Age mother at birth | Kandala and Ghilagaber (2006) | DHS Malawi 2000 | U-5 | Child | U-shaped |
| | Age mother at birth | Kandala et al. (2014) | DHS 2007 Congo, Dem. Rep. | U-5 | Child | U-shaped |
| Nutrition mother | BMI | Black et al. (2013) | DHS LMICs | Neonatal, infant | Child | U-shaped |
| Education | Education parents | Breierova and Duflo (2004) | SU-PAS Indonesia | Child | Household | Linear decreasing |
| | Education mother | Grépin and Bharadwaj (2015) | DHS Zimbabwe 1988, 1994, 1999, 2006, 2011 | Infant, child | Child | Linear decreasing |
| Birth interval (BI) | Education mother preceding BI | Makate and Makate (2016) | DHS Malawi 2000, 2005, 2010 | U-5 | Child | Linear decreasing |
| | log preceding BI | Rutstein (2005) | DHS 17 countries | neonatal, infant, U-5 | Child | Linear decreasing |
| **Household specific characteristics** | | | | | | |
| Household’s wealth | Wealth index | Adebayo and Fahrmeir (2005) | DHS Nigeria 1999 | 0–35 months | Child | Linear decreasing |
| | Wealth index | Ayele et al. (2015) | DHS Ethiopia 2011 | U-5 | Child | Linear decreasing |

(continued)
Table 1. (Continued)

| Variable         | Description            | Author                          | Data source                  | Mortality variable | Level   | Relationship                  |
|------------------|------------------------|---------------------------------|------------------------------|--------------------|---------|------------------------------|
| Household size   | Number of people       | Ayele et al. (2015)             | DHS Ethiopia 2001            | U-5                | Child   | Non-linear increasing        |
| Residency        | City, town, slum, rural| Fink, Günther, and Hill (2014)  | DHS 73 countries             | Child              |         | Lower risk urban residency   |
| Environmental and socio-economic factors | GDP | $\log$ per capita GDP | Bhalotra (2006) | NFHS India 1998/99 | U-5 | State | Negative relationship |
|                  | $\log$ per capita GDP | Baird, Friedman, and Schady (2011) | DHS 59 countries | Infant | Country | Negative relationship |
|                  | $\log$ per capita GDP | Houweling et al. (2005) | 43 countries (World Bank, DHS) | U-5 | Country | Linear decreasing |
| Time trend       | Survey year            | Korenromp et al. (2004)        | DHS 41 African countries     | U-5                | Country | Linear decreasing            |
|                  | Survey year            | Kuate-Defo and Diallo (2002)   | DHS 28 African countries     | U-5                | Mother  | Linear decreasing            |

Notes: Incomplete list of key findings from studies found in the literature that analyses factors associated with neonatal, infant, child, and under-five mortality.
3.2. Statistical modelling approach

3.2.1. Smooth non-linear function $f$. The potential non-linear terms will be modelled through a non-linear function $f$ that can be estimated using penalised-splines (P-splines) (Fahrmeir et al., 2013, Chapter 8). The P-splines will be approximated by using an adequate linear combination of basis functions that take, for instance for the asset index, the form:

$$f_1(\text{Asset index}) = \sum_{j=1}^{d} \gamma_j B_j(\text{Asset index})$$ (2)

The basic idea of P-splines is to divide the range of the data into a relatively large number of intervals. The boundaries between two intervals are called knots, and we use 10 knots as the default choice for the number of knots. Within each interval usually a third degree polynomial is assumed, the function $f$ is forced to be twice differentiable at the knots to ensure smoothness. In a frequentist set-up this model could be estimated considering standard maximum likelihood procedure. In a Bayesian set-up the model will be estimated using a simulation procedure through Markov chain Monte Carlo (MCMC).

3.2.2. Correlated covariates. The birth order is highly correlated with the age of the mother ($\rho = 0.76$), and an interaction between the two covariates is most likely. Incorporating this smooth bivariate effect can be achieved by including a two-dimensional effect of the form, $f_6(\text{Age mother}_{it}, \text{Birth order}_i)$ into Equation (1). The association between under-five mortality and the interaction of the two correlated covariates is modelled and interpreted as surface with irregular boundaries, where $f_6$ is a smooth function depending on the two correlated covariates (Fahrmeir et al., 2013, Chapter 8 and Chapter 9). The effect of the combinations of the two metric covariates can be interpreted as a spatial effect arranged on an irregular grid, where the combinations are interpreted as discrete location variables. Estimation is based on Bayesian Markov random fields, as proposed and described by Fahrmeir and Lang (2001).

3.2.3. Capturing differences between Asian and sub-Saharan African countries. For determining potential differences between Asian and sub-Saharan African countries a varying coefficient term for the geographic location is included in the estimation. Including varying coefficients was first proposed by Hastie and Tibshirani (1993) and applications of varying coefficients can be found in Lang and Sunder (2003) and Lang et al. (2014). The effect of a covariate (for example the wealth index of the household child $i$ lives in), can vary with respect to the binary indicator of the geopolitical region. This will be modelled as follows,

$$\eta_{it}^{\text{add}} = f_0(\text{Baseline hazard}) + \ldots + f_2(\text{Asset index}_i) + f_{\text{Asset index|Continet}}(\text{Asset index}_i)(\text{Continet}) + \ldots + x_i'\gamma,$$ (3)

and includes a varying coefficient term, where the effect of the geopolitical indicator Continent is varying with respect to the observed value of the continuous covariate (here the wealth index). Including this type of interaction into the estimation allows us to detect potential differences in the effect of a specific covariate for Asia and sub-Saharan Africa.

3.2.4. Multilevel structured additive regression (STAR) framework. To account for the hierarchical structure of the data, and to allow for heterogeneity between countries, a multilevel STAR model is used. The 836,286 observed children are nested into 358 regions, which are again nested into 48 countries in Asia and sub-Saharan Africa. Nesting of the observations is also called clustering (Jain, Murty, & Flynn, 1999; Korenromp et al., 2004). Following Lang et al. (2014), we use a multilevel
approach that allows us to account for heterogeneity at the regional and country level, as well as to include non-linear covariate effects. The multilevel model with three levels is written as follows:

\[
\begin{align*}
\text{Level} & - 1 : \eta_{it} = \eta_{i}^{\text{individual}} + f_{14}(\text{Region}_i) \\
\text{Level} & - 2 : f_{14}(\text{Region}) = f_{14,1}(\text{Stunting regional mean}_\text{Region}) + \\
& f_{14,2}(\text{Stunting regional mean}_\text{Region})\text{Continent} + \\
& f_{14,3}(\text{Regional vaccination rate}_\text{Region}) + \\
& f_{14,4}(\text{Regional vaccination rate}_\text{Region})\text{Continent} + \\
& f_{14,5}(\text{Country}_\text{Region}) + f_{14,6}(\text{Region}) \\
\text{Level} & - 3 : f_{14,5,1}(\text{Country}) = f_{14,5,1}(\text{GDP}_\text{Country,}t) + \\
& f_{14,5,2}(\text{GDP}_\text{Country,}t)\text{Continent} + f_{14,5,3}(\text{Country}) + \\
& f_{14,5,1}\text{Continet}_\text{Country}
\end{align*}
\]

The model accounts for the hierarchical structure of the data, which allows us to model the effects of regional and country specific covariates on under-five mortality. To determine systematic differences between Asian and sub-Saharan African countries a varying coefficient is included in the STAR model. In Equation (4), \( \eta_{i}^{\text{individual}} \) subsumes the included categorical covariates, and the continuous covariates including the corresponding varying coefficient. Regional specific spatial heterogeneity is modelled through the level-2 equation of \( f_{14}(\text{Region}) \), which includes the smooth effects of the regional mean of stunting, the regional vaccination rate, and the corresponding varying coefficients. Country specific spatial heterogeneity is modelled through the level-3 equation of \( f_{14,5}(\text{Country}) \), and includes a smooth term for the countries per capita GDP, and the corresponding varying coefficient term.

3.2.5. Accounting for country specific covariate effects. Up to this point the model captures cluster specific heterogeneity by including i.i.d. Gaussian random effects of the region \( (f_{14,6}(\text{Region}) \sim N(0; \sigma_{f_{14,6}}^2)) \), and the country \( (f_{14,5,3}(\text{Country}) \sim N(0; \sigma_{f_{14,5,3}}^2)) \). To capture potential heterogeneity of the individual specific non-linear effects \( f_q \), and the individual specific linear effects \( \gamma_q \), for each individual specific covariate, a country specific random effect is included.

Models that include this type of interaction for an unordered grouping factor, such as countries, are also called models with random slopes (Belitz et al., 2015). This can be of particular use, when accounting for country or region-specific heterogeneity. The basic idea is to add random slopes to the individual specific covariates of Equation (4). For example, for the effect of the wealth of the household, this is achieved by including a random slope term of the form \( \tilde{f}_2(\text{Country})\text{Assetindex}_t \) into Equation (4). \( \tilde{f}_2(\text{Country}) \sim N(0; \sigma_{f_{14,2}}^2) \) is i.i.d. normal distributed with inverse gamma distributed variance \( \sigma_{f_{14,2}}^2 \). A more recent, alternative method to model the cluster specific variation of the non-linear terms is to use multiplicative random effects (also called random scaling) (see for instance, Brunauer, Lang, & Umlauf, 2013; Lang, Steiner, Weber, & Wechselberger, 2015; Razen, Lang, & Santer, 2016; Wechselberger, Lang, & Steiner, 2008). In this variant the main effect of Equation (4) is substituted by \( (1 + \tilde{f}_2(\text{Country})) \cdot f_2(\text{Assetindex}_t) \). A random effect smaller than zero \( (\tilde{f}_q(\text{Country}) < 0) \) yields to a scaling down of the main effect, whereas a random effect greater than zero \( (\tilde{f}_q(\text{Country}) > 0) \) yields the main effect to be scaled up. In countries with a positive random effect the effect is more pronounced, whereas in countries with a negative random effect the effect is less pronounced.

3.2.6. Model selection. The selection process is based on two steps: first, we roughly distinguish between different model types (linear model, additive model, multilevel model, multilevel model including country random slopes, or multiplicative random effects) using the deviance information criterion (DIC) (Spiegelhalter, Best, Carlin, & Van der Linde, 2002). The DIC is
considered to be the generalisation of the Akaike information criterion (AIC). Differences between models greater than 20 points are considered to be substantial, and the model with the lowest DIC is selected. Second, significant covariates are identified using Bayesian credible bands (Krivobokova, Kneib, & Claeskens, 2010), and insignificant covariates are dropped to obtain the ‘best’ model. This can be summarised as follows (see Table A3 for details): comparing the models of Equations (1), (3), and (4), strongly favours the hierarchical model of Equation (4), emphasising the necessity to consider non-linearities, differences between Asia and sub-Saharan Africa, and the hierarchical data structure. Accounting for country specific covariate effects, and adding either random slopes, or random scaling factors to Equation (4) further improves the model, and random scaling is strictly favourable compared to random slopes. Omitting insignificant effects yields the final model of Equation (5). The following insignificant effects have been removed: the varying effects of the interaction between the age of the mother and the birth order, the BMI of the mother, and the GDP.

\[
\begin{align*}
\text{Level } 1: \quad \eta_{it} &= (1 + \tilde{f}_0(\text{Country})) \cdot f_0(\text{Baseline hazard}) + \\
&+ f_1(\text{Baseline hazard})(\text{Continent}) + \ldots + \\
&+ f_{12}(\text{Region}_i) + \gamma_1(\text{Older dead sibling}_i) + \\
&+ \tilde{f}_1(\text{Country})(\text{Older dead sibling}_i) + \ldots + \\
&+ \tilde{f}_3(\text{Country})(\text{Place of living})
\end{align*}
\]

\[
\begin{align*}
\text{Level } 2 : \quad f_{12}(\text{Region}) &= f_{12,1}(\text{Stunting regional mean}_{\text{Region}}) + \\
&+ f_{12,2}(\text{Stunting regional mean}_{\text{Region}})(\text{Continent}) + \\
&+ f_{12,3}(\text{Regional vaccination rate}_{\text{Region}}) + \\
&+ f_{12,4}(\text{Regional vaccination rate}_{\text{Region}})(\text{Continent}) + \\
&+ f_{12,5}(\text{Country}_{\text{Region}}) + f_{12,6}(\text{Region})
\end{align*}
\]

\[
\begin{align*}
\text{Level } 3 : \quad f_{12,5}(\text{Country}) &= f_{12,5,1}(\text{GDP}_{\text{Country}, t}) + \\
&+ \gamma_{12,5,1}(\text{Continent}_{\text{Country}})
\end{align*}
\]

\section*{4. Results}

Results, unless explicitly stated, are presented for the model of Equation (5) that emerged from the selection process of Section 3.2.6. Robustness checks can be found in the online supplement, these include an analysis of the effect of the preceding birth interval, controlling for migration, whether the age was imputed or not, including time-varying effects, and weighting the final model by the population and the survey weights.

\subsection*{4.1. Linear effects}

Table 2 summarises the results of the linear effects. The table shows the posterior mean together with the 95 per cent credible interval. The only effect found to be insignificant in the model building phase was whether the household is led by a female or male. This covariate was not included in the final analysis. All other linear covariates are found to be significant.

\subsection*{4.1.1. Child specific factors.} The indicator of whether the child has an older sibling who died, is the covariate found to have the largest effect. Growing up in a household where an older sibling died is associated with a higher mortality risk in the first 60 months, compared to a household where no such fatality occurred. The estimated posterior mean is 0.33, which is in line with our expectations, as well as the literature (see for instance, Bhalotra & Van Soest, 2008; Harttgen et al., 2015; Hobcraft et al., 1985). It can be seen as a reflection of household specific risk factors that accumulate to a higher mortality risk, like the household’s access to sanitary facilities, hygiene standards within the household, and environmental risk factors the household is exposed to.
The estimated posterior mean of 0.03 of the gender of the child is also significant and consistent with our expectations, and the literature such as Arnold (1997). Compared to girls, boys tend to have a slightly higher risk of dying in their first five years, which is also reflected in these estimates. One follow-up remark needs to be made. In countries, such as India, with a strong male preference, girls are faced with a higher mortality risk (Khera, Jain, Lodha, & Ramakrishnan, 2014). This is also reflected in the bottom left panel of Figure 2 and emphasises the differences across countries, as well as the capability of the applied method to detect these cross-country differences.

4.1.2. Environmental and socio-economic factors. Children living in urban areas have higher survival probabilities compared to their rural counterparts. The estimated posterior mean of −0.02 is significant and negative, however, of small magnitude. These estimates reveal that children below five years, who live in rural areas, are confronted with a higher mortality risk. This gap between urban and rural areas can be explained through generally poorer health status and basic health infrastructure in rural areas (Günther & Harttgen, 2012).

The estimated results for the gender of the child, and the area of living should be interpreted with great care. When analysing large sample sizes already relatively small effects tend to become significant. Both effects are significant but scarcely important, considering the usual predictor range of probit models between −2 and 2.

4.2. Non-linear effects

Figure 3 to the top panel of Figure 6 show the results of the non-linear effects of the continuous covariates. The figures show from left to right, the mean effect together with the 80 per cent and 95 per cent credible intervals, and for covariates with significant differences between Asia and sub-Saharan Africa, the continent specific effects. The next panel depicts the main effect, respectively, the continent specific effects scaled by the country random effect of the continuous covariates.

4.2.1. Child specific factors. In more detail for the baseline hazard that shows a strong effect considering the usual predictor range of probit models between −2 and 2: the top left panel in Figure 3 depicts the posterior mean of the main effect including credible intervals. The mortality risk is found to be highest for neonates, and then more or less steadily declining. Two things are important to note: first, the mortality risk is highest in the first months, and the baseline hazard has more or less a concave shape. Second, the two local maxima around month 18 and month 48, respectively, the local minimum around month 36, and the sharp decline after the 50th month seems to be artefacts, a similar pattern was already pointed out by Adebayo and Fahrmeir (2005). As Rutstein and Rojas (2006) point out, DHS report for older children the age at death in full years, instead of a more accurate monthly reporting. Besides the baseline effect, the varying effect (see top left panel of Figure A1) is found to be significant. These two effects can be combined to the top right panel of Figure 3, which shows mayor differences between Asian and sub-Saharan African countries. In particular, a faster decrease of the probability of dying for children living in Asia. Sub-Saharan African

| Definition Variable | Mean | SD  | 95% CI (lower bound) | 95% CI (upper bound) |
|---------------------|------|-----|----------------------|----------------------|
| Overall intercept   | −2.65| 0.03| −2.71                | −2.59                |
| Continent (SSA = 1) | 0.12 | 0.03| 0.06                 | 0.17                 |
| Older sibling of child i died (yes = 1) | 0.33 | 0.01| 0.31                 | 0.35                 |
| Gender (male = 1)   | 0.03 | 0.00| 0.02                 | 0.03                 |
| Place living (urban = 1) | −0.02| 0.00| −0.03                | −0.01                |

Note: Estimation results of categorical covariates. All covariates are effect coded.
Source: DHS; calculation by authors.
countries have a flatter baseline hazard compared to Asian countries, indicating a mortality risk that is less dependent on the age of the child. Modelling country specific heterogeneity with multiplicative random effects yield substantial differences between countries, highlighting the need to account for heterogeneity beyond a random intercept. A random effect $<0$, such as for example in the case of Cambodia, Niger, Burkina Faso, or Uganda yields to scale down the baseline hazard, which results the country specific effect of the age being less dominant. A random effect $>0$, amplifies the main effect and scales the baseline effect up, for example in the case of Bangladesh, Pakistan, the Comoros, or Rwanda. Combining the country specific random effect (see top right panel of Figure 3) and the continent specific effect (top right panel of Figure A1) give the country specific effects illustrated in row two of Figure 3. This shows that between sub-Saharan African countries the variation compared to Asian countries is higher, probably caused by a higher dispersion between countries in sub-Saharan Africa. Analysing the effects of the other covariates is analogous.

The analysis includes the regional mean of the height-for-age z-score, and the composite regional immunisation rate for selected vaccines as recommended by Rutstein and Rojas (2006). The effects are depicted in the third and fourth row of Figure 3. A better nutritional status observed in a particular region reduces the individual mortality risk for observations in this region. The effect is more or less monotonically decreasing over the complete range of the data. The difference between Asian and sub-Saharan African countries, even though significant, are found to be of minor relevance due to the small size of the varying effect. This effect is consistent with the general findings of Pelletier and Frongillo (2003), Black et al. (2003), and

---

**Figure 2.** Categorical covariates, country random slopes and random intercepts.

*Notes:* Figure depicts the posterior means of the random intercept (top left), categorical variables dead sibling (top right), gender (bottom left), and place of living (bottom right) together with 80 per cent and 95 per cent simultaneous credible intervals. Asian countries are shown to the left of the vertical dashed line, while sub-Saharan African countries are on the right. International country codes (ISO-3) are used as abbreviations. *Source:* DHS; calculations by the authors.
Figure 3. Non-linear effects of the baseline hazard, the nutritional status, and regional immunisation rate. 

Notes: Shown are the mean effects, together with 95 per cent and 80 per cent simultaneous credible intervals. Additionally, the continent specific effects, and for the baseline hazard the with the country random effects scaled effects for Asia and sub-Saharan Africa are shown.

Source: DHS; calculations by the authors.
Black et al. (2013). It puts an emphasis on reducing malnutrition (measured as the height-for-age z-score) to values even above the threshold of −2 of the WHO for moderate stunting. Considering the regional immunisation rate, the effect starts to decrease monotonically and becomes significant, when the immunisation rate is above 50 per cent, showing that in regions with lower immunisation rates children are exposed to a higher risk of dying. Like for the nutritional status the varying effect is found to be negligible, due to the small size of the effect.

Both effects emphasise the importance of improving the nutritional status and increasing vaccination coverage in low and lower-middle income countries. These can be directly transferred to the advice for policy-makers and institutions that reducing the number of severely stunted children and increasing the share of children fully vaccinated against BCG, measles, DPT, and polio would also be reflected in decreasing mortality rates. This result emphasises the need for better vaccination campaigns to increase immunisation coverage as already recommended lately by Boco (2010) and Wang (2003). In the case of vaccination coverage, institutions should aim for a higher vaccination rate across countries, which follows closely the guidelines established by the WHO.

The age of the mother is highly correlated with the birth order, which biases the estimated effects. To account for this, a two-dimensional effect is estimated using Markov random fields. The effect is best visualised by plotting a grid with all possible combinations of the two covariates, as shown in Figure 4. In the top panel the main effect (left) together with the 95 per cent posterior probability (right) are illustrated. An interaction of the two covariates exists, and the mortality risk is highest for children of young mothers, and then decreases with the age of the mother, on the other hand, leaving out observations of young mothers the risk is constantly increasing with the birth order. This shows that not accounting for this interaction would have yielded invalid results. No significant differences between the two geopolitical regions were found, however, differences between countries exists as illustrated by the country random effects and the effect by country which are shown in the bottom left panel of Figure A1 and the Supplementary Materials.

4.2.2. Maternal characteristics. Contrary to Harttgen et al. (2015), who found an inverted U-shaped relationship between the mother’s BMI and child mortality, we find the effect to be U-shaped (row 2 of Figure 4).\textsuperscript{12} It decreases up to a BMI of around 20, and then increases again. This is particularly interesting since malnutrition of the mother and obesity of the mother increase child mortality. The effect is found to be rather small in magnitude, and the differences between the two geopolitical regions were found to be insignificant. Additionally, the effect is rather homogeneous across countries.

Another strong effect is the effect of the mother’s years of education (Figure 4). The effect is found to be more or less decreasing linearly and corresponds to the findings of Cleland and Van Ginneken (1988) and Vollmer, Bommer, Krishna, Harttgen, and Subramanian (2017).\textsuperscript{13} Investing in the education for expecting mothers, and women in general, seems to lower child mortality. Several explanations for this exist. Women with higher education can better assess relevant information, which, in the end, will benefit the children’s health and overall status, or have a better knowledge about health-related issues. Higher education has a monetary pay-off, since the chance to find better jobs increases. This result is strengthened by Breierova and Dufllo (2004), Grépin and Bharadwaj (2015), and Makate and Makate (2016) who find a causal relationship of the mother’s education on the risk of dying and emphasises the need to foster the educational opportunities of women. It opposes the results of Desai and Alva (1998), who argue that, even though a strong correlation between education and infant mortality exists, the educational level of the mother can be seen as an approximation of the family’s socio-economic status.

4.2.3. Household specific characteristics. The effect of the household’s wealth is decreasing more or less monotonically (top panels of Figure 5), showing that children born in wealthier households (relative to the regional mean) have a higher survival probability. This finding is very important considering that the wealth index as an indicator of long-term economics status of the household (compared to income, which can be considered as a more short-term indicator of standard of living)
Figure 4. Two-dimensional effect of the age of the mother and the birth order (top), non-linear effects of the BMI of the mother (middle), and the mother’s years of education (bottom).

Notes: Shown are the mean effects together with simultaneous 95 per cent posterior probability for the two-dimensional effect, the mean effects, together with 95 per cent and 80 per cent simultaneous credible intervals for the remaining effects. Additionally, the continent specific effects, and the with country random effects scaled effects for Asia and sub-Saharan Africa are shown for the BMI of the mother and the mother's years of education.

Source: DHS; calculations by the authors.
and which emphasises the necessity of pro-poor growth. This finding is of particular interest since reducing poverty rates in developing countries would also imply that child mortality will decline in these countries. The interaction with geographic indicator is significant and reveals differences between households located in Asian and sub-Saharan African countries. Between countries a large dispersion exists, in particular between Asian countries.

The size of the household is also found to influence under-five mortality significantly and shows a strong effect; children living in small households \((n \leq 3)\) are found to be faced with the highest mortality risk (bottom panels of Figure 5). Most likely this corresponds to the effect of growing up in a single parent household, as only one parent can provide income and food. The effect decreases monotonically until a household size of approximately seven or eight people. For larger households the effect remains more or less constant. The varying effect of the household size is found to be significant, but due to the small effect size negligible (left panel row 3 of Figure A2).

4.2.4. Environmental and socio-economic factors. The results reveal some interesting findings about declining under-five mortality rates over time and between sub-Saharan Africa and Asia. Declining under-five mortality rates are observed since the late 1990s and can be illustrated by plotting under-five mortality rates by birth year (see right-hand panel of Figure 1) and are found by various authors (see for instance Jamison, Murphy, & Sandbu, 2016; Lozano et al., 2011). We find that until the year 1995 under-five mortality decreased throughout Asia on a higher rate compared to sub-Saharan Africa, while the effect slowed down in Asia between 1996 and 2005 and continued to decrease from the year 2005 on (Figure 6). The effect of sub-Saharan Africa is monotonically declining, however, it decreases with a lower rate compared to Asia. This is consistent with a study by Black et al. (2003), where sub-Saharan African countries also experienced a drop in under-five mortality, however, the decline was of much smaller magnitude in comparison to Asian countries. While both regions did not fully achieve MDG four, Asian countries came closer to achieving this goal. This also puts an emphasis on the critique by Easterly (2009) and Lange and Klasen (2017) who claim that the target set by the MDG is ‘unfair’ against sub-Saharan Africa. This is due to the fact that MDG four is formulated in relative terms, and sub-Saharan African countries are faced with mortality rates around twice as high compared to Asia; in relative terms formulating the goal implies a reduction in absolute terms approximately twice as high for sub-Saharan Africa.

The effect of the countries’ per capita GDP is found to be significant and non-linear. No differences between Asian and sub-Saharan African countries are found. For low values of per capita GDP, the effect is strongly decreasing, while for higher values, it can be assumed to be constant even though the estimated effect increases. This is due to increased uncertainty, indicated by widening credible intervals, because of fewer observations. This shows that increasing per capita GDP in low income countries is also reflected in a reduction of under-five mortality rates. As mentioned by Wang (2003), an emphasis should be to allocate more resources to countries’ health systems.

4.3. Interaction of binary covariates and geopolitical region

Controlling for a possible interaction between the child’s gender and the geographic location of the household confirms the pattern illustrated by the Nelson-Aalen cumulative hazard (see Supplementary Materials). Boys in sub-Saharan Africa have the highest mortality risk (posterior mean: 0.11; 95 per cent credible interval: [0.06, 0.14]). Girls growing up in sub-Saharan Africa have the second highest mortality risk (posterior mean: 0.05; 95 per cent credible interval: [0.02, 0.09]). The posterior mean of Asian boys is −0.07; [−0.11, −0.02], the calculation of the effect for the reference category, girls in Asia, is feasible as effect coding is used and reveals that girls born in Asia have the lowest risk of dying. The mean effect is −0.09. This illustrates the existing differences
Figure 5. Non-linear effects of the household’s wealth (top), and the size of the household (bottom).

Notes: Shown are the mean effects, together with 95 per cent and 80 per cent simultaneous credible intervals. Additionally, the continent specific effects, and the with country random effects scaled effects for Asia and sub-Saharan Africa are shown.

Source: DHS; calculations by the authors.
between the two geopolitical regions and boys and girls, and that growing up in Asia is associated with a lower mortality risk.

Allowing for a similar interaction term for the household’s place of living and the geographic location, reveals that with respect to the geopolitical region, children born in urban areas – possibly due to better access to health care, sanitation, and drinking water – are exposed to a lower mortality risk compared to children growing up in rural areas. Table 3 shows these results in more detail.

![Figure 6](image)

**Figure 6.** Non linear time trend of the year of birth, and non linear effect of the GDP.
*Notes:* Shown are the mean effects, together with 95 per cent and 80 per cent simultaneous credible intervals, and the continent specific effect for the birth year.
*Source:* DHS, and PEN World Table 9.0; calculations by authors.

### Table 3. Results model with interaction place of living | continent

| Variable      | Mean | SD  | 95% CI (lower bound) | 95% CI (upper bound) |
|---------------|------|-----|----------------------|----------------------|
| Urban SSA     | 0.06 | 0.02| 0.02                 | 0.11                 |
| Rural SSA     | 0.10 | 0.02| 0.06                 | 0.15                 |
| Urban Asia    | −0.10| 0.02| −0.14                | −0.05                |
| Rural Asia    | −0.06|     |                      |                      |

*Notes:* Results categorical covariates interaction place of living | continent. All covariates are effect coded.
*Source:* DHS; calculation by authors.
5. Conclusion

On a global scale, under-five mortality rates fell approximately by one half since 1990. A decline is observed, both in relative terms (per 1,000 live-births), as well as in absolute numbers (You, Hug, et al., 2015). Despite this positive development, two regions were identified where the reduction was lacking behind the projections needed to achieve MDG four of reducing under-five mortality by two thirds between 1990 and 2015. These regions are Asia and sub-Saharan Africa (You, New, et al., 2015). In this study possible drivers of under-five mortality in sub-Saharan Africa and Asia are investigated, allowing for non-linear effects, in order to understand why child health increased in some low- and middle-income countries while it stagnated in others. This is thought to offer a starting point to improve policy towards achieving the target of SDG three: reducing under-five mortality to at least as low as 25 per 1,000 live births by 2030 (Liu et al., 2016).

The contribution of our analysis to the existing studies of determinants of child health in the empirical literature is fourfold and features several novelties. First, we analyse the largest pooled cross-sectional dataset currently available. Second, we explicitly take into account potential differences in determinants of under-five mortality between sub-Saharan Africa and Asia and allow for heterogeneity between countries. Third, we consider potential non-linearities in determinants of under-five mortality. Fourth, we employ a fully Bayesian discrete-time survival model.

We find strong non-linear effects that can be summarised as follows. Differences between Asia and sub-Saharan Africa exist with respect to the baseline hazard, the BMI of the mother, the household’s wealth and size, the birth year, and the countries per capita GDP. We also find a relatively strong interaction of the age of the mother and the birth order, due to the high correlation, for which we accounted for. Heterogeneity of associated determinants is explicitly accounted for and can be seen as one factor, why under-five mortality decreased in such great variation between low- and middle-income countries. Undernutrition and vaccination coverage, are both found to be important drivers of under-five mortality. While the regional mean of the height-for-age z-score decreases more or less linearly, a non-linear pattern emerges for the later: only when at least 50 per cent of the children are fully vaccinated in the first year does it influence under-five mortality. With respect to the differences between the analysed geopolitical regions, substantial differences have been found for the age of the child, the wealth of the household, and the birth year.

No single factor, however, could be isolated on which policy should focus, instead our analysis shows that a wide range of influential determinants on under-five mortality exist on which policy should focus:

- Improving maternal education, since several mechanisms exist through which improving maternal education reduces under-five mortality.
- Fostering policies that enhance pro-poor growth and are targeted to improve the living standards of children in poor households.
- Increase vaccination coverage through immunisation programmes. Immunisation programmes are found to be efficacious measures to prevent diseases which influence under-five mortality.
- Improving the situation of malnourished children, and provide better access to health and nutritional programmes for undernourished children. These programmes should be targeted to reduce the share of malnourished children.

These examples should be seen as an incomplete list of policy measures, on which policy-makers should focus to help to reach SDG three. Further research needs to control for HIV and malaria prevalence, however, this requires a significant improvement of data availability. Estimating spatial effects and identifying high risk areas could further help to provide effective policy implications aiming to reduce under-five mortality.

Acknowledgements

The computational results presented have been achieved (in part) using the HPC infrastructure LEO of the University of Innsbruck. We thank Macro International Inc., for collecting the DHS, and
providing access to the data and Catherine Laflamme and Antoinette Van der Merwe for proofreading the manuscript. We thank conference participants at the 16th Nordic Conference in Development Economics 2017, the 31st Annual Conference of the European Society for Population Economics (ESPE 2017), and participants at various seminars for their feedback on earlier versions of this paper. Finally, we also like to thank the managing editor, and the anonymous referees for giving valuable comments and constructive suggestions that helped to improve the paper. All remaining errors are our own. Raw data is public available, the BayesX and R codes used for processing, and analysing the data for this study, are available from the authors upon request.

**Funding**

Johannes Seiler gratefully acknowledges financial support from a doctoral scholarship by the University of Innsbruck.

**Disclosure statement**

No potential conflict of interest was reported by the authors.

**Notes**

1. The SDGs were adopted by the United Nations in 2015 and can be seen as a successor of the MDGs. In contrast to the MDGs they cover not only development related topics, instead the focus of the SDGs is on sustainable development.
2. The regions are harmonised across DHS surveys, which provides us with a panel data structure at the regional level allowing us to include region fixed effects.
3. For a more detailed explanation on the calculation of the z-score see, for example, Harttgen et al. (2015). The regional mean of the height-for-age z-score in region \( j \) and survey year \( t \) is calculated as follows:

\[
\text{Stunting}_{jt} = \frac{1}{n} \sum_{i=1}^{n} z - \text{stunting}_{ijt}.
\]

The height-for-age z-score is used for two reasons: first, according to de Onis, Blössner, and Borghi (2012), the prevalence of undernutrition is higher for stunted children compared to the other two standard anthropometric measures. Second, more observations at the regional level are available for the height-for-age z-score in contrast to the anthropometric measures for weight-for-age and weight-for-height. Including instead anthropometric measures for weight-for-age and weight-for-height, respectively, would have reduced the sample to less than 358 regions, as in some regions only height-for-age z-score was observed.
4. As recommended by Rutstein and Rojas (2006), we calculate the immunisation rate for children age 12–23 months at a regional level. Included vaccines are: BCG (Bacillus Calmette-Guérin against tuberculosis), measles, and three doses of DPT (vaccine against diphtheria, pertussis, and tetanus), and polio vaccine each. To calculate the immunisation rate, children aged 12–23 months are included, since immunisation should be completed in the first year.
5. Even though the DHS data already provides a wealth index of households (hv270), the asset index is calculated using the information provided in the data on the households’ assets, following the approach used besides others by Filmer and Pritchett (2001) and Sahn and Stifel (2003). This is done, as the DHS wealth index is not calculated for all surveys. Recalculating the asset index allows us to use more surveys and more observations.
6. For a more elaborate discussion of the data augmentation consult, for instance, the methodology manual of BayesX.
7. To make an estimation tractable in BayesX, we further divide the monthly duration time scale into intervals, as follows, \([0, 1), [1, 2), [2, 3), [3, 4, 5, 6), [6, 7, 8, 9), ... , [53, 54, 55, 56), [56, 57, 58, 59)\).
8. For a more thorough description of the estimation see Fahrmeir et al. (2013, Chapter 8 and Chapter 9) or Harttgen et al. (2015).
9. Including time-varying effects improves the model further. However, with the exemption of the effect for the mother’s educational attainment the size of the effects is found to be to small (ranging from −0.05 to 0.05) to be relevant.
10. Weighting the model by population and the DHS survey weights did not change the results.
11. A linear effect is assumed to be significant if zero is not included in the 95 per cent credible interval.
12. The WHO groups the BMI of adults as follows: BMI <18.5, Underweight; 18.5 ≤ BMI <25.0, Normal-weight; 25.0 ≤ BMI <30.0, Pre-obesity; 30.0 ≤ BMI <35.0, Obesity class-I; 35.0 ≤ BMI <40.0, Obesity class-II; BMI >40.0, Obesity class-III.
13. In addition, an interaction with the baseline hazard is found to be significant and relevant (see Supplementary Materials). The baseline hazard shows a stronger decrease for children of better educated mothers. For older children above four years the effect reverses, which should be considered as artefact due to few observations in this range of the age.

ORCID

Johannes Seiler http://orcid.org/0000-0001-5714-9234

References

Adebayo, S. B., & Fahrmeir, L. (2005). Analysing child mortality in Nigeria with geoadditive discrete-time survival models. *Statistics in Medicine*, 24(5), 709–728.

Arnold, F. (1997). *Gender preferences for children. DHS comparative studies no. 23*. Calverton, MD: Macro International Inc.

Ayele, D. G., Zewotir, T. T., & Mwambi, H. G. (2015). Structured additive regression models with spatial correlation to estimate under-five mortality risk factors in Ethiopia. *BMC Public Health*, 15(1), 268.

Baird, S., Friedman, J., & Schady, N. (2011). Aggregate income shocks and infant mortality in the developing world. *The Review of Economics and Statistics*, 93(3), 847–856.

Belitz, C., Brezger, A., Klein, N., Kneib, T., Lang, S., & Umlauf, N. (2015, July). BayesX: Software for Bayesian inference in structured additive regression models. Version 3.0.2 [Computer software manual]. Retrieved from http://www.BayesX.org/

Bhalotra, S. (2006). *Childhood mortality and economic growth* (WIDER Research Paper, No. 2006/79). Helsinki: The United Nations University World Institute for Development Economic Research (UNU-WIDER).

Bhalotra, S., & Van Soest, A. (2008). Birth-spacing, fertility and neonatal mortality in India: Dynamics, frailty, and fecundity. *Journal of Econometrics*, 143(2), 274–290.

Black, R. E., Morris, S. S., & Bryce, J. (2003). Where and why are 10 million children dying every year? *The Lancet*, 361(9376), 2226–2234.

Black, R. E., Victora, C. G., Walker, S. P., Bhutta, Z. A., Christian, P., de Onis, M., ... Uauy, R. (2013). Maternal and child undernutrition and overweight in low-income and middle-income countries. *The Lancet*, 382(9890), 427–451.

Boco, A. G. (2010). *Individual and community level effects on child mortality: An analysis of 28 Demographic and Health Surveys in sub-Saharan Africa* (Tech. Rep.). Calverton, MD: Macro International.

Breierova, L., & Duflo, E. (2004). *The impact of education on fertility and child mortality: Do fathers really matter less than mothers?* (Working Paper No. 10513). Cambridge, MA: National Bureau of Economic Research.

Brunauer, W., Lang, S., & Umlauf, N. (2013). Modelling house prices using multilevel structured additive regression. *Statistical Modelling*, 13(2), 95–123.

Chamarbagwala, R. (2010). Sibling composition and selective gender-based survival bias. *Journal of Population Economics*, 24(3), 935–955.

Cleland, J. G., & Van Ginneken, J. K. (1988). Maternal education and child survival indeveloping countries: The search for pathways of influence. *Social Science & Medicine*, 27(12), 1357–1368.

de Onis, M., Blossner, M., & Borghi, E. (2012). Prevalence and trends of stunting among pre-school children, 1990–2020. *Public Health Nutrition*, 15(1), 142–148.

Desai, S., & Alva, S. (1998). Maternal education and child health: Is there a strong causal relationship? *Demography*, 35(1), 71–81.

Easterly, W. (2009). How the millennium development goals are unfair to Africa. *World Development*, 37(1), 26–35.

Fahrmeir, L., Kneib, T., Lang, S., & Marx, B. (2013). *Regression: Models, methods and applications* (1st ed.). Berlin: Springer-Verlag.

Fahrmeir, L., & Lang, S. (2001). Bayesian inference for generalized additive mixed models based on Markov random field priors. *Journal of the Royal Statistical Society: Series C (Applied Statistics)*, 50(2), 201–220.

Feenstra, R. C., Inklaar, R., & Timmer, M. P. (2015). The next generation of the Penn World Table. *American Economic Review*, 105(10), 3150–3182.

Filmer, D., & Pritchett, L. H. (2001). Estimating wealth effects without expenditure data — Or tears: An application to educational enrollments in states of India. *Demography*, 38(1), 115–132.

Fink, G., Günther, I., & Hill, K. (2014). Slum residence and child health in developing countries. *Demography*, 51(4), 1175–1197.

Fink, G., Victora, C. G., Harttgen, K., Vollmer, S., Vidaletti, L. P., & Barros, A. J. (2017). Measuring socioeconomic inequalities with predicted absolute incomes rather than wealth quintiles: A comparative assessment using child stunting data from national surveys. *American Journal of Public Health*, 107(4), 550–555. PMID: 28207339

Grépin, K. A., & Bharadwaj, P. (2015). Maternal education and child mortality in Zimbabwe. *Journal of Health Economics*, 44, 97–117.

Günther, I., & Harttgen, K. (2012). Deadly cities? A note on spatial inequalities in mortality in sub-Saharan Africa. *Population and Development Review*, 38(3), 469–486.
A multilevel approach to explain child mortality and undernutrition in South Asia and sub-Saharan Africa (Tech. Rep. No. 20). Ibero America Institute for Economic Research.

Harttgen, K., Lang, S., & Santer, J. (2015). Multilevel modelling of child mortality in Africa (Working Paper No. 03). Innsbruck: Economics and Statistics University of Innsbruck.

Hastie, T., & Tibshirani, R. (1993). Varying-coefficient models. Journal of the Royal Statistical Society. Series B (Methodological), 55(4), 757–796.

Hobcraft, J., McDonald, J., & Rutstein, S. (1985). demographic determinants of infant and early child mortality: A comparative analysis. Population Studies, 39(3), 363–385.

Houweling, T. A. J., Kunst, A. E., Looman, C. W. N., & Mackenbach, J. P. (2005). Determinants of under-5 mortality among the poor and the rich: A cross-national analysis of 43 developing countries. International Journal of Epidemiology, 34(6), 1257–1265.

Jain, A. K., Murty, M. N., & Flynn, P. J. (1999). Data clustering: A review. ACM Computing Surveys, 31(3), 264–323.

Jamison, D. T., Murphy, S. M., & Sandhu, M. E. (2016). Why has under-5 mortality decreased at such different rates in different countries? Journal of Health Economics, 48, 16–25.

Kandala, N.-B., & Ghilagaber, G. (2006). A geo-additive Bayesian discrete-time survival model and its application to spatial analysis of childhood mortality in Malawi. Quality and Quantity, 40(6), 935–957.

Kandala, N.-B., Mandungu, T. P., Mbela, K., Nzita, K. P. D., Kalambahy, B. B., Kayembe, K. P., & Emina, J. B. O. (2014). Child mortality in the democratic republic of Congo: Cross-sectional evidence of the effect of geographic location and prolonged conflict from a national household survey. BMC Public Health, 14(1), 266.

Khera, R., Jain, S., Lodha, R., & Ramakrishnan, S. (2014). Gender bias in child care and child health: Global patterns. Archives of Disease in Childhood, 99(4), 369–374.

Klasen, S. (2000). Malnourished and surviving in South Asia, better nourished and dying young in Africa: What can explain this puzzle? (Tech. Rep.). University of Munich.

Kneib, T., Heinzl, F., Brezger, A., Sabanés Bové, D., & Klein, N. (2014). BayesX: R utilities accompanying the software package BayesX [Computer software manual]. (R package version 0.2–9).

Korenromp, E. L., Arnold, F., Williams, B. G., Nahlen, B. L., & Snow, R. W. (2004). Monitor- ing trends in under-5 mortality rates through national birth history surveys. International Journal of Epidemiology, 33(6), 1293–1301.

Krivobokova, T., Kneib, T., & Claeskens, G. (2010). Simultaneous confidence bands for penalized spline estimators. Journal of the American Statistical Association, 105(490), 852–863.

Kuate-Defo, B., & Diallo, K. (2002). Geography of child mortality clustering within African families. Health & Place, 8(2), 93–117.

Lang, S., Steiner, W. J., Weber, A., & Wechselberger, P. (2015). Accommodating heterogeneity and nonlinearity in price effects for predicting brand sales and profits. European Journal of Operational Research, 246(1), 232–241.

Lang, S., & Sunder, M. (2003). Non-parametric regression with BayesX: A flexible estimation of trends in human physical stature in 19th century America. Economics & Human Biology, 1(1), 77–89.

Lang, S., Umlauf, N., Wechselberger, P., Harttgen, K., & Kneib, T. (2014). Multilevel structured additive regression. Statistics and Computing, 24(2), 223–238.

Lange, S., & Klasen, S. (2017). How the new international goal for child mortality is unfair to sub-Saharan Africa (again). World Development, 90, 128–146.

Liu, L., Oza, S., Hogan, D., Chue, Y., Perin, J., Zhu, J., … Black, R. E. (2016). Global, regional, and national causes of under-5 mortality in 2000–15: An updated systematic analysis with implications for the sustainable development goals. The Lancet, 388(10063), 3027–3035.

Lozano, R., Wang, H., Foreman, K. J., Knoll Rajaratnam, J., Naghavi, M., Marcus, J. R., … Murray, C. J. L. (2011). Progress towards millennium development goals 4 and 5 on maternal and child mortality: An updated systematic analysis. The Lancet, 378(9797), 1139–1165.

Makate, M., & Makate, C. (2016). The causal effect of increased primary schooling on child mortality in Malawi: Universal primary education as a natural experiment. Social Science & Medicine, 168, 72–83.

Mosley, W. H., & Chen, L. C. (1984). An analytical framework for the study of child survival in developing countries. Population and Development Review, 10, 25–45.

Pelletier, D. L., & Frongillo, E. A. (2003). Changes in child survival are strongly associated with changes in malnutrition in developing countries. The Journal of Nutrition, 133(1), 107–119.

R Core Team. (2016). R: A language and environment for statistical computing [Computer software manual]. Vienna, Austria. Retrieved from https://www.R-project.org/

Razen, A., Lang, S., & Santer, J. (2016). Estimation of spatially correlated random scaling factors based on Markov random field priors (Working Paper No. 33). Innsbruck: Economics and Statistics University of Innsbruck.

Rutstein, S. O. (2000). Factors associated with trends in infant and child mortality in developing countries during the 1990s. Bulletin of the World Health Organization, 78, 1256–1270.

Rutstein, S. O. (2005). Effects of preceding birth intervals on neonatal, infant and under-five years mortality and nutritional status in developing countries: Evidence from the demo- graphic and health surveys. International Journal of Gynecology & Obstetrics, 89, 7–24. Optimal Birth Spacing: New Evidence on Newborn, Infant, Child and Maternal Health.

Rutstein, S. O., & Rujias, G. (2006). Guide to DHS statistics: Demographic and Health Surveys methodology (Tech. Rep.). Calverton, MD: ORC Macro.
Sahn, D. E., & Stifel, D. (2003). Exploring alternative measures of welfare in the absence of expenditure data. Review of Income and Wealth, 49(4), 463–489.

Schultz, T. P. (1984). Studying the impact of household economic and community variables on child mortality. Population and Development Review, 10, 215–235.

Spiegelhalter, D. J., Best, N. G., Carlin, B. P., & Van der Linde, A. (2002). Bayesian measures of model complexity and fit. Journal of the Royal Statistical Society. Series B (Statistical Methodology), 64(4), 583–639.

Umlauf, N., Adler, D., Kneib, T., Lang, S., & Zeileis, A. (2015). Structured additive regression models: An R interface to BayesX. Journal of Statistical Software, 63(1), 1–46.

Vollmer, S., Bommer, C., Krishna, A., Harttgen, K., & Subramanian, S. V. (2017). The association of parental education with childhood undernutrition in low- and middle-income countries: Comparing the role of paternal and maternal education. International Journal of Epidemiology, 46(1), 312–323.

Vollmer, S., Harttgen, K., Kupka, R., & Subramanian, S. V. (2017). Levels and trends of childhood undernutrition by wealth and education according to a composite index of anthropometric failure: Evidence from 146 Demographic and Health Surveys from 39 countries. BMJ Global Health, 2(2), e000206.

Wang, L. (2003). Determinants of child mortality in LDCs: Empirical findings from demo- graphic and health surveys. Health Policy (Amsterdam, Netherlands), 65(3), 277–299.

Wechselberger, P., Lang, S., & Steiner, W. J. (2008). Additive models with random scaling factors: Applications to modeling price response functions. Austrian Journal of Statistics, 37(3, 4), 255–270.

Wolpin, K. I. (1997). Determinants and consequences of the mortality and health of infants and children. (Vol. 1 A; M. Rosenzweig & O. Stark, Eds.) (No. chapter 10). Amsterdam: Elsevier.

You, D., Hug, L., Ejdemyr, S., Idele, P., Hogan, D., Mathers, C., … Alkema, L. (2015). Global, regional, and national levels and trends in under-5 mortality between 1990 and 2015, with scenario-based projections to 2030: A systematic analysis by the UN inter-agency group for child mortality estimation. The Lancet, 386(10010), 2275–2286.

You, D., Jones, G., Hill, K., Wardlaw, T., & Chopra, M. (2010). Levels and trends in child mortality, 1990–2009. The Lancet, 376(9745), 931–933.

You, D., New, J. R., & Wardlaw, T. (2012). Levels and trends in child mortality. Report 2012. Estimates developed by the UN inter-agency group for child mortality estimation (Tech. Rep.). United Nations.

You, D., New, J. R., & Wardlaw, T. (2014). Levels and trends in child mortality. Report 2014. Estimates developed by the UN inter-agency group for child mortality estimation (Tech. Rep.). United Nations.

You, D., New, J. R., & Wardlaw, T. (2015). Levels and trends in child mortality. Report 2015. Estimates developed by the UN inter-agency group for child mortality estimation (Tech. Rep.). United Nations.
Appendix

A.1. Under-five mortality rates by country and survey year

Table A1. Number of observations ($n$) and under-five mortality rates per 1,000 live births (U5M) for the five-year period preceding the survey by country, and survey year.

| Country     | ISO-3 | Year | $n$ | U5M | Country       | ISO-3 | Year | $n$ | U5M |
|-------------|-------|------|-----|-----|---------------|-------|------|-----|-----|
| Armenia     | ARM   | 2000 | 1,532 | 40 | Guinea        | GIN   | 2005 | 3,235 | 154 |
| Armenia     | ARM   | 2005 | 1,264 | 23 | Guinea        | GIN   | 2012 | 3,644 | 124 |
| Armenia     | ARM   | 2016 | 1,530 | 12 | India         | IND   | 1999 | 48,930 | 96  |
| Azerbaijan  | AZE   | 2006 | 2,074 | 49 | India         | IND   | 2006 | 46,043 | 69  |
| Bangladesh  | BGD   | 1997 | 5,720 | 115| Ivory Coast   | CIV   | 1994 | 5,273 | 158 |
| Bangladesh  | BGD   | 2000 | 6,612 | 85 | Ivory Coast   | CIV   | 1999 | 501   | 131 |
| Bangladesh  | BGD   | 2004 | 6,768 | 80 | Ivory Coast   | CIV   | 2012 | 3,748 | 117 |
| Bangladesh  | BGD   | 2007 | 5,975 | 67 | Jordan        | JOR   | 1997 | 5,427 | 34  |
| Bangladesh  | BGD   | 2011 | 8,442 | 53 | Jordan        | JOR   | 2002 | 4,883 | 26  |
| Bangladesh  | BGD   | 2014 | 7,665 | 47 | Jordan        | JOR   | 2007 | 4,020 | 18  |
| Benin       | BEN   | 1996 | 4,092 | 170| Jordan        | JOR   | 2009 | 3,327 | 28  |
| Benin       | BEN   | 2001 | 5,132 | 157| Jordan        | JOR   | 2012 | 4,736 | 25  |
| Benin       | BEN   | 2006 | 15,024| 123| Kazakhstan    | KAZ   | 1995 | 1,347 | 51  |
| Benin       | BEN   | 2012 | 12,879| 79 | Kazakhstan    | KAZ   | 1999 | 598   | 63  |
| Burkina Faso| BFA   | 2003 | 10,113| 179| Kenya         | KEN   | 1993 | 5,634 | 98  |
| Burkina Faso| BFA   | 2010 | 7,528 | 129| Kenya         | KEN   | 1998 | 5,263 | 112 |
| Burundi     | BDI   | 2010 | 3,566 | 90 | Kenya         | KEN   | 2003 | 5,218 | 112 |
| C. A. R.    | CAF   | 1995 | 3,840 | 149| Kenya         | KEN   | 2009 | 5,816 | 77  |
| Cambodia    | KHM   | 2000 | 4,038 | 142| Kenya         | KEN   | 2014 | 9,738 | 53  |
| Cambodia    | KHM   | 2005 | 4,093 | 94 | Kyrgyz Republic| KGZ  | 1997 | 1,886 | 64  |
| Cambodia    | KHM   | 2010 | 4,001 | 60 | Kyrgyz Republic| KGZ  | 2012 | 3,433 | 28  |
| Cameroon    | CMR   | 1998 | 3,000 | 168| Lesotho       | LSO   | 2004 | 1,743 | 119 |
| Cameroon    | CMR   | 2004 | 3,942 | 149| Lesotho       | LSO   | 2014 | 1,532 | 96  |
| Cameroon    | CMR   | 2011 | 5,809 | 130| Liberia       | LBR   | 2007 | 5,383 | 112 |
| Chad        | TCD   | 1997 | 7,029 | 200| Liberia       | LBR   | 2013 | 3,849 | 114 |
| Chad        | TCD   | 2004 | 5,229 | 173| Madagascar    | MDG   | 1997 | 5,081 | 166 |
| Chad        | TCD   | 2015 | 11,326| 136| Madagascar    | MDG   | 2004 | 5,216 | 82  |
| Comoros     | COM   | 1996 | 1,559 | 107| Madagascar    | MDG   | 2009 | 6,044 | 84  |

(continued)
Table A1. (Continued)

| Country          | ISO-3 | Year | n  | U5M | Country    | ISO-3 | Year | n  | U5M | Country    | ISO-3 | Year | n  | U5M |
|------------------|-------|------|----|-----|------------|-------|------|----|-----|------------|-------|------|----|-----|
| Comoros          | COM   | 2012 | 2,753 | 51 | Malawi     | MWI   | 1992 | 4,245 | 233 | Uganda   | UGA   | 2001 | 6,350 | 149 |
| Congo, Dem. Rep. | COD   | 2007 | 4,349 | 150 | Malawi     | MWI   | 2000 | 11,454 | 186 | Uganda   | UGA   | 2006 | 2,844 | 118 |
| Congo, Dem. Rep. | COD   | 2014 | 9,342 | 121 | Malawi     | MWI   | 2004 | 10,277 | 129 | Uganda   | UGA   | 2011 | 2,414 | 101 |
| Congo, Rep.      | COG   | 2012 | 4,789 | 68 | Malawi     | MWI   | 2010 | 6,449 | 108 | Uzbekistan | UZB   | 1996 | 2,225 | 63 |
| Ethiopia         | ETH   | 2000 | 10,541 | 168 | Malawi     | MWI   | 2015 | 5,544 | 60 | Zambia    | ZMB   | 1992 | 6,038 | 190 |
| Ethiopia         | ETH   | 2005 | 4,797 | 126 | Maldives   | MDV   | 2009 | 2,991 | 18 | Zambia    | ZMB   | 1996 | 6,942 | 196 |
| Ethiopia         | ETH   | 2011 | 11,115 | 99 | Mali       | MLI   | 1996 | 7,180 | 234 | Zambia    | ZMB   | 2002 | 6,690 | 177 |
| Ethiopia         | ETH   | 2016 | 9,952 | 80 | Mali       | MLI   | 2001 | 12,328 | 213 | Zambia    | ZMB   | 2007 | 6,282 | 116 |
| Gabon            | GAB   | 2000 | 4,004 | 75 | Mali       | MLI   | 2006 | 13,576 | 178 | Zimbabwe  | ZWA   | 1994 | 3,112 | 89 |
| Gabon            | GAB   | 2012 | 3,858 | 80 | Mali       | MLI   | 2013 | 5,260 | 110 | Zimbabwe  | ZWA   | 1999 | 3,385 | 111 |
| Gambia           | GMB   | 2013 | 3,294 | 49 | Mozambique | MOZ   | 1997 | 5,529 | 176 | Zimbabwe  | ZWA   | 2006 | 5,076 | 91 |
| Ghana            | GHA   | 1993 | 2,975 | 131 | Mozambique | MOZ   | 2003 | 9,455 | 149 | Zimbabwe  | ZWA   | 2011 | 5,176 | 85 |
| Ghana            | GHA   | 1998 | 3,152 | 104 | Mozambique | MOZ   | 2011 | 10,855 | 96 | Zimbabwe  | ZWA   | 2015 | 5,563 | 65 |
| Ghana            | GHA   | 2003 | 3,591 | 115 | Namibia    | NAM   | 1992 | 3,467 | 84 | Zimbabwe  | ZWA   | 2007 | 4,677 | 80 |
| Ghana            | GHA   | 2008 | 2,865 | 86 | Namibia    | NAM   | 2007 | 4,677 | 80 | Zimbabwe  | ZWA   | 2013 | 2,245 | 55 |

Source: DHS datasets; calculation by authors.
### A.2. Descriptive statistics covariates

| Definition covariates                                                                 | Asia and sub-Saharan Africa | Asia | sub-Saharan Africa |
|---------------------------------------------------------------------------------------|-----------------------------|------|-------------------|
| Child has older dead sibling (1 = yes)                                                | 8.43%                       | 3.25%| 10.31%            |
| Sex of child (1 = male)                                                               | 50.8%                       | 51.53%| 50.53%            |
| Place of living (1 = urban)                                                           | 28.66%                      | 34.17%| 26.67%            |
| Migration (1 = same location since birth)                                            | 51.04%                      | 35.38%| 57.87%            |
| Continent (1 = sub-Saharan Africa)                                                    | 73.4%                       | 13    | 35                |
| Asset index deviation reg. mean                                                       | -0.03                       | 0     | -0.04             |
| Age children (in months)                                                              | 26.75                       | 27.97 | 26.3              |
| Date of birth and death not imputed (1 = yes)                                        | 96.01%                      | 99.75%| 94.65%            |
| Birth order within household                                                          | 3.51                        | 2.84  | 3.75              |
| Years of education mother                                                             | 3.89                        | 5.03  | 3.48              |
| Number of people in household                                                         | 6.84                        | 6.62  | 6.92              |
| BMI mother                                                                             | 21.86                       | 21.32 | 22.06             |
| Age mother at birth (in years)                                                        | 25.36                       | 24.26 | 25.76             |
| Age mother not imputed (1 = yes)                                                      | 81.45%                      | 56.22%| 90.59%            |
| Preceding birth interval                                                              | 37.95                       | 39.34 | 38.2              |
| Z-score stunting reg. mean                                                            | -1.4                        | -1.33 | -1.42             |
| Reg. vaccination rate                                                                 | 0.54                        | 0.6   | 0.52              |
| Real per capita GDP                                                                   | 2,990.05                    | 5,005.78| 2,241.35         |

**Notes:** The sample contains 836,286 observations in 358 regions and 48 countries.

**Source:** DHS, Penn World Table 9.0; calculations by authors.
### Table A3. Model selection: differences DIC to baseline model

| Model                                                                 | Δ DIC  |
|-----------------------------------------------------------------------|--------|
| Additive model (AM) Equation (1)                                      | 3,918  |
| AM Equation (3)                                                       | 0      |
| Multilevel AM (MAM) Equation (4) omitting regional level             | −1,633 |
| MAM Equation (4)                                                      | −2,725 |
| MAM Equation (4), omitting regional level, including country random slopes | −4,300 |
| Equation (4), including country random slopes                         | −5,241 |
| MAM Equation (4), omitting regional level, including multiplicative country random effects (MRE) | −5,312 |
| MAM Equation (4), including MRE                                       | −6,232 |
| **MAM Equation (4), including MRE, omitting insignificant varying coefficient terms** | −6,223 |
| MAM Equation (4), including MRE, interaction gender|continent | −6,199 |
| MAM Equation (4), including MRE, interaction urban|continent   | −6,187 |
| MAM Equation (4), including MRE, omitting birth order                 | −5,763 |
| MAM Equation (4), including MRE, omitting age mother                  | −5,478 |

*Source: DHS; calculation by authors.*
A.4. Varying and random effects main model

Figure A1. Non-linear varying effects of the baseline hazard (top), the nutritional status (middle left), and regional immunisation rate (middle right), and the country random effects of the two-dimensional effect of the age of the mother and the birth order (bottom left) and the BMI of the mother (bottom right).

Notes: Except for the two-dimensional effect of the age of the mother and the birth order, and the BMI of the mother the panel depicts the within continent varying effects, together with 95 per cent and 80 per cent simultaneous credible intervals. Asian countries are shown to the left of the vertical dashed line, while sub-Saharan African countries are on the right.

Source: DHS; calculations by the authors.
Figure A2. Non-linear varying effects the years of education of the mother (top), the wealth of the household (row 2), the size of the household (row 3), and the year of birth (bottom).

Notes: Shown are the continent varying effects, together with 95 per cent and 80 per cent simultaneous credible intervals. Except for the birth year country random effects are shown. Asian countries are shown to the left of the vertical dashed line, while sub-Saharan African countries are shown on the right, using international country codes (ISO-3) as abbreviations.

Source: DHS; calculations by the authors.