Demography (2021) 58(1):137–163 Published online: 12 January 2021
DOI 10.1215/00703370-8931544

Decomposition of Improvements in Infant Mortality in Asian Developing Countries Over Three Decades

Toshiaki Aizawa

ABSTRACT Low- and middle-income countries in Asia have seen substantial improvements in infant mortality over the last three decades. This study examines the factors contributing to the improvement in infant survival in their first year in six Asian countries: Bangladesh, India, Indonesia, Nepal, Pakistan, and the Philippines. I decompose the overall improvement in the infant survival rate in the respective countries from the 1990s to the 2010s into the part that can be explained by the improvements in circumstantial environments in which infants develop and the remaining part that is due to the structural change in the hazard functions. This decomposition is achieved by employing the random survival forest, allowing me to predict the counterfactual infant survival probability that infants in the 2010s would have under the circumstantial environments of the 1990s. The results show that large parts of the improvement are explained by the improvement in the environments in all the countries being analyzed. I find that the reduction in family size, increased use of antenatal care, longer pregnancy periods, and improved living standards were associated with the improvement of the infant mortality rate in all six countries.

KEYWORDS Infant mortality • Survival • Random survival forest • Asia • Decomposition

Introduction

Child and infant mortality rates have been improving in many developing countries across the world over time (UNICEF 2018). Substantial progress in child survival has accelerated since 1990 after the adoption of the Millennium Development Goals of the United Nations, the fourth of which targeted a two-thirds reduction in child mortality between 1990 and 2015 (UNICEF 2017; United Nations Development Program [UNDP] 2015). In Asia, where child and infant mortality are high, under-5 mortality rates were reduced—by 62% in South-East Asia, 60% in Southern Asia, 78% in Eastern Asia, and 65% in Western Asia (UNDP 2015). These figures outnumbered the average rate found in all the developing regions of the world (53%) (UNDP 2015). Although the Millennium Development Goals did not specifically target *infant* mortality rate, its improvement was also observed internationally. The infant mortality

ELECTRONIC SUPPLEMENTARY MATERIAL The online version of this article (https://doi.org/10.1215/00703370-8931544) contains supplementary material.

rate was 64.7 deaths per 1,000 live births in 1990 and decreased to 28.9 deaths per 1,000 in 2018 (World Bank 2019). In terms of the absolute number, in 1990, 8.7 million infants died worldwide before celebrating their first birthday. By 2018, this number had shrunk to just 4 million. The Sustainable Development Goals, adopted in 2015 by the United Nations, introduced a third goal to be achieved by 2030: "Ensure healthy lives and promote well-being for all at all ages" (United Nations 2015:20).

Given that the first 12 months after birth is the critical period in terms of subsequent survival, a large volume of the literature has investigated the determinants of infant mortalities in developed and developing countries. For example, using country-level data, Sartorius and Sartorius (2014) explored risk factors associated with global infant mortality for the period 1990–2011 in 192 countries. Islam and Hyder (2016) analyzed risk factors relating to infant survival in four countries in South Asia using the Cox proportional hazard model and cross-sectional infant-level data. This study contributes to the existing research by providing new, internationally comparable evidence using repeated cross-sectional infant-level data. In contrast to the existing literature, I focus on the transition of the infant mortality rate over three decades and, through decomposition analysis, closely investigate the contributing factors to the improvement.

This study employs survival analysis rather than the binary response model, which estimates the probability of surviving the first year. The first year is such a critical period for subsequent survival that it is important to carefully consider heterogeneous risk factors and their influences throughout this critical period. In contrast to the binary response model, the survival model allows the researcher to examine heterogeneous risk factors associated with infant mortalities at each point in time over the period. It is globally observed that as infant mortality declines, the mean age at death in infancy decreases because of declining exogenous mortality risks, such as digestive and respiratory problems, suggesting that risk factors for infant mortality are not homogeneous throughout the period (Andreev and Kingkade 2015). Moreover, the survival model can be regarded as a general approach of the binary response model in the sense that the probability of surviving the first year can be expressed by hazard functions up to the first year.

I first explore the improvement in infant survival rates over the three decades from the 1990s to the 2010s in low- and middle-income countries in Asia. Specifically, I analyze six countries in South and South-East Asia: Bangladesh, India, Indonesia, Nepal, Pakistan, and the Philippines. Then I investigate the changes in survival curves in the respective countries by decomposing the changes into the part that can be explained by the observable improvement in the environments where infants grow up and the other, unexplained part in the spirit of the Oaxaca and Blinder decomposition (Blinder 1973; Oaxaca 1973). This decomposition analysis helps elucidate the factors contributing to the pronounced decline in infant mortality over the last few decades. I aim to quantify how much of the improvement is due to the improvement in the circumstantial environments, such as household characteristics, parental socioeconomic status, and use of maternal healthcare. I hope that this exploration of the reasons behind the improvements in Asia will in turn suggest important health policy implications for the other developing countries in Asia and beyond with high infant mortality rates.

The Cox proportional hazard (Cox-PH) model (Cox 1972) has been one of the most common regression modeling frameworks for survival analysis. However, in

contrast to the majority of studies on infant mortality, this study employs a fully data-adaptive machine-learning algorithm called the random survival forest (RSF) method. The RSF was developed by Ishwaran and Kogalur (2007) and Ishwaran et al. (2008) as an extension of the random forest model (Breiman 2001) to right-censored survival data. I model the survival function by the RSF and predict the counterfactual survival probability in the hypothetical scenario in which infants born in the 2010s are assumed to have the circumstantial environments of the 1990s. Using this counterfactual survival probability in the decomposition analysis, I quantify how much of the improvement is associated with the improvement in circumstantial environments.

Data

Demographic Health Survey

The Demographic and Health Survey (DHS) project is an ongoing collaboration between the United States Agency for International Development and country-specific agencies. They conduct nationally representative, household sample surveys covering a range of population health indicators in low- and middle-income countries (Corsi et al. 2012). The DHS data has been gathered on the basis of comparable nationally representative household surveys conducted in more than 85 countries worldwide since 1984. The DHS respondents are selected using a two-stage sampling process stratified by urban and rural location. Key advantages of the DHS include the national coverage and high participation rates, typically exceeding 90%. In addition, the DHS questionnaire has been standardized and pre-tested to ensure comparability across populations and over time. Standard data collection procedures and interviewer training in the DHS ensure that the data is both reliable and comparable.

This study exploits the DHS conducted in all six countries studied here: Bangladesh, India, Indonesia, Nepal, Pakistan, and the Philippines. The interview years of DHS data used in this study and sample sizes are listed in Table 1. I select these six countries because they have infant mortality data from the last three decades.[1] The DHS collected data on infants who were born no more than 60 months prior to the survey. I focus on the most recent birth of each mother because in the latest DHS data collected in the 2010s, some of the information regarding the use of antenatal care is available only for the most recent birth, and I do not have complete information for children whose mother had more than one childbirth in the five years prior to the survey.[2] I estimate the survival probability, exploiting the information on child's birthday, interview date, status of alive or deceased, and death date (if applicable). For infants alive at the time of the interview, I calculate the right-censored survival length from the record.

[1] I do not use the DHS data of other countries because of data unavailability. The DHS does not have infant mortality data for Afghanistan, Cambodia, Maldives, and Timor-Leste for the 1990s; for Kyrgyzstan, Myanmar, and Tajikistan for the 2000s; or for Thailand, Uzbekistan, and Vietnam for the 2010s.

[2] I also investigate the survival rates of all the children under 5 by imputing the missing information using the information on the recent birth with hot-deck imputation (Joenssen and Bankhofer 2012). These results are available from the author upon request.

Table 1 Sample sizes in each period

	2010s	2000s	1990s	Total	Available Data Years
Bangladesh	11,447	12,783	10,012	34,242	2014, 2011, 2007, 2004, 1999, 1996, 1993
India	32,445	36,034	63,313	131,792	2015/2016, 2005, 1998/1999, 1992/1993
Indonesia	29,992	27,988	38,109	96,089	2017, 2012, 2007, 2002/2003, 1997, 1994, 1991
Nepal	7,900	8,639	3,680	20,219	2016, 2011, 2006, 2001, 1996
Pakistan	15,115	5,402	3,848	24,365	2017/2018, 2012, 2006/2007, 1990/1991
Philippines	13,038	9,418	10,793	33,249	2017, 2013, 2008, 2003, 1998, 1993

Covariates

The DHS data contain not only infant-level demographic information but also rich information about household characteristics, parental socioeconomic status, and maternal antenatal healthcare use. This study uses the variables that may well be considered time-invariant after childbirth.[3] The covariates used in this study are based on previous research on infant mortality and child health in developing countries (e.g., Aizawa 2019; Akseer et al. 2017; Hobcraft et al. 1985; Kesterton et al. 2010; Sartorius and Sartorius 2014; Westley 2003). First, regarding infant-level information, I include infant sex, whether a child was born as a twin, birth order, whether a mother has birth spacing over 36 months, and whether a mother wanted a child when she became pregnant. A short interpregnancy interval is related to maternal mortality, stillbirth, and child mortality (Fotso et al. 2013); an interpregnancy span over 36 months is encouraged to reduce infant mortality risks (Molitoris et al. 2019). I regard it as a proxy of appropriate family planning whether a mother wanted a child when she became pregnant. In addition, I include information about the mother's age at birth, which is an important risk factor of child and infant mortality. I also include two binary variables: whether the mother's pregnancy was a teenage pregnancy, and whether the mother got pregnant after age 35. A strong relationship between child mortality and the prevalence of teenage pregnancy has been internationally observed (Finlay et al. 2011).

Second, as household characteristics, I include household location (urban or rural), drinking water source (piped water, well water, or other), roof materials (finished, rudimentary, or natural), floor materials (finished, rudimentary, or natural), wall materials (finished, rudimentary, or natural), toilet type (flushing, pit latrine, or other), electricity access, and ownership of television(s), refrigerator(s), and car(s).[4] Housing conditions and ownership of various household items are used as proxies for living standards. The association between housing conditions and child health is well established; for example, dilapidated/poor housing has been regarded as one of the important causes of child illness (Bradley et al. 2001; Marmot 1999). Household sanitary conditions are an equally important factor for child health. Drinking water source and toilet types are used as proxies for sanitary conditions, which have a sub-

[3] Using only the time-invariant covariates allows me to predict meaningful survival probabilities.

[4] For detailed information about the definitions of roof, floor, and wall materials, see Table A7 in the online appendix.

stantial influence on health (Ahsan et al. 2017; Caulfield et al. 2004; Checkley et al. 2004); for example, unsafe drinking water and poor hygienic environments are leading causes of diarrhea (Ezzati et al. 2002; Konteh 2009).

Third, for parental socioeconomic status, I use paternal and maternal highest educational levels completed (incomplete primary education, primary education, secondary education, or higher education) and occupation types (professional worker, manual worker, nonmanual worker, farmer, or not working) as binary variables. A number of studies have found correlations between mothers' educational achievements and child mortality (e.g., Mondal et al. 2009), and other research has shown an association between parental employment status and child health (Blau et al. 1996; McGuire and Popkin 1990; Ruhm 2004).

Finally, I consider maternal healthcare use and breastfeeding behaviors. I include the place of childbirth (private hospital; clinic/public hospital; or clinic/others, including at home), amount of antenatal care received, cesarean delivery, birth assistant attendance, uptake of tetanus toxoid injections,[5] and the first breastfeeding time after childbirth (within one hour or within one day). The adequate use of antenatal healthcare services is a crucial factor in successful maternal and child health outcomes (Campbell and Graham 2006), and timely use of antenatal care from healthcare professionals and midwives helps reduce the risk of preterm births (Adams et al. 2000; Medley et al. 2018; Sandall et al. 2016). Breastfeeding is encouraged to reduce the risk of undernutrition and prevent resultant infectious diseases, such as pneumonia, which is the leading cause of child mortality (Chisti et al. 2009; Lamberti et al. 2013; Troeger et al. 2018). For India, I also include castes as a country-specific potentially important risk factor of infant mortality.

Descriptive statistics in each country are provided in Tables A1–A6 in the online appendix. Over the three decades, I find longer birth spacing, higher parental educational attainment, increased institutional delivery, skilled birth attendance, use of antenatal care, uptake of tetanus toxoid injections, and improvement in breastfeeding behaviors in all six countries. The reductions in family size are observed in all examples except Pakistan, where only decreased number of children in a household is observed. Descriptive statistics also indicate that all countries experienced improvements in living standards and housing conditions. Unsurprisingly, no country experienced significant changes in the sex ratio and the proportion of twin births.

Methods

Notations

In this paper, $T \geq 0$ denotes survival time to death. For infant i, I observe $T_i = \min\{T_i^0, C_i\}$, where T_i^0 is the survival time for individual i, and C_i is the observable survival time under right-censoring. I define the binary censoring indicator as

[5] Because the information about uptake of tetanus toxoid injections is not available in the most recent data in Bangladesh, this variable is not used for Bangladesh.

$\delta_i = I\left(C_i \leq T_i^0\right)$. $\delta_i = 1$ if an individual i is right-censored at time T_i, and $\delta_i = 0$ if an individual i has died at time T_i.

$S(t) = P(T > t)$ is a survival function, and $h(t) = P(T = t | T > t - 1)$ is a hazard function. A cumulative hazard function is defined as $H(t) = \int_0^t h(u)\,du$. I define $X \in \chi$ as a set of d-dimensional time-invariant covariates related to the environments where infants grow up. Finally, $Y = \{Y_{1990s}, Y_{2000s}, Y_{2010s}\}$ is a set of year indicators. For example, Y_{1990s} equals 1 for observations in the 1990s, and Y_{1990s} equals 0 otherwise.

Cox Proportional Hazard

As a benchmark model, I consider the Cox proportional hazard model (Cox-PH) (Cox 1972). The Cox-PH has been widely used in survival analysis because of its computational simplicity. It is known as a semiparametric regression model in the sense that it does not need to specify the base hazard function to estimate the hazard ratio. However, its simplicity is largely due to its stringent assumptions. First, the Cox-PH assumes a multiplicative relationship between an underlying baseline hazard function and a log-linear function of the covariates. Second, it usually does not take into account nonlinear effects and complicated higher-order interaction effects among covariates. Also, it assumes that the hazard ratio is constant across time. In order to partially relax the second assumption, I include two-way interaction terms between the period indicators, Y, and covariates as additional regressors, which allows the Cox-PH to take into account the heterogeneous hazard ratio associated with X across the three periods: the 1990s, 2000s, and 2010s. Furthermore, I also consider the parsimonious model called the selective Cox-PH, in which only a relevant combination of regressors is used in modeling the hazard and predicting the survival function (Mogensen et al. 2012). The selection of covariates helps to achieve a parsimonious model structure and potentially enhances predictive performance.

Random Survival Forests

Overview

The main purpose of employing the random survival forest (RSF) is to estimate the conditional survival function or the conditional cumulative hazard function. The RSF is a data-driven machine-learning approach to the nonparametric estimation of the conditional survival function, which is an extension of the random forest (Breiman 2001) to the right-censored survival data (Ishwaran and Kogalur 2007). The RSF is capable of delineating nonlinear effects and high-order interactions of covariates (Ishwaran et al. 2008; Mogensen et al. 2012). In contrast to the traditional approaches, such as the Cox-PH and other parametric methods, researchers do not have to select important variables in advance through stepwise regressions. The description of the RSF in this study follows expositions by Ishwaran and Kogalur (2007), Ishwaran et al. (2008), and Mogensen et al. (2012).

The RSF is an ensemble method that introduces two forms of randomization into the tree-growing. Figure 1 summarizes the macro-level procedures. Before estimating the model, I split the entire samples into training samples and testing samples. The training samples are used to train/estimate a model, and the testing samples are used to evaluate the model's predictive performance and predict the survival probabilities. I use 60% of the entire sample as the training sample and the remaining 40% as the testing sample.[6] Evaluating model performance with out-of-sample rather than in-sample information is increasingly adopted beyond the field of data science because of the growing popularity of machine-learning approaches (Mullainathan and Spiess 2017; Varian 2014).

First, the RSF draws multiple bootstrap samples from the training samples and builds a survival tree using each bootstrap sample. Samples selected in each tree are called *in-bag samples*, and the other samples not selected to grow a tree are called *out-of-bag samples*. In-bag samples are used to grow trees, and out-of-bag samples are used for tuning model parameters at a later stage. For details, see the subsection A.3.2 in the online appendix.

Second, when constructing each tree, the RSF selects a random set of independent variables as candidate variables to split a tree. The number of candidate variables corresponds to the square root of the total number of independent variables in the data. Therefore, trees are constructed using different samples and different sets of independent variables, which is designed to encourage independence among trees and prevent overfitting (Breiman 2001).

Each tree is grown by recursively partitioning the in-bag samples based on optimization of a split rule. This study applies the log-rank splitting rule (Leblanc and Crowley 1993; Segal 1988), which is a standard split rule in the RSF. Within the candidate independent variables, the one variable x^* with an optimal split point c^* that maximizes the differences between child nodes is sought and used for node splitting. Hence, each subject in the in-bag samples is classified into two child nodes exhibiting the highest log-rank statistics. Eventually, as the number of nodes increases and dissimilar subjects become separated, each node in the tree becomes homogeneous and is populated by subjects with similar survival probabilities. A more detailed description of node splitting is provided in subsection A.3.1 in the online appendix.

The growth of a tree is continued until all the terminal nodes contain only a minimal number of unique subjects, the optimal size of which is sought via parameter tuning. The RSF sorts each observation in the in-bag samples into one unique terminal node per tree. Survival estimates for each observation at each event time are constructed within each terminal node. The forest ensemble is constructed by aggregating over the 500 random trees. Having a large number of forest trees ensures that each variable has enough of an opportunity to be included in the forest prediction process. A more detailed explanation of this ensemble estimation is provided in subsection A.3.3 in the online appendix. These ensemble algorithms lead to a more accurate out-of-sample prediction in comparison with traditional survival methods, such as the Cox-PH approach (Dietrich et al. 2016; Imani et al. 2019; Yosefian et al. 2015), and achieve parsimony of the model. The property of consistency is discussed in Ishwaran and Kogalur (2010).

[6] As a sensitivity analysis, I use 70% of the samples as the training sample and the remaining 30% as the testing samples. I obtain very similar results.

1. Form training samples (60%) and testing samples (40%).
2. Conduct parameter tuning to find a optimal terminal node size.
3. Draw B bootstrap samples with size M from the training data without replacement.
4. Develop a survival tree for each bootstrap sample $b = \{1, 2, \ldots, B\}$.
   - At each root node of the tree, randomly select $p = \sqrt{dim(X, Y)}$ candidate variables.
   - Find a set of optimal variable and its splitting value, (x^*, c^*) that maximizes the log-rank statistics.
5. Continue to grow a tree until terminal nodes for each tree should have no less than minimum unique deaths.
6. Estimate a survival function for each tree, $\widehat{S}_b(t|X, Y)$.
7. Take the average over the trees to obtain the ensemble survival function, $\widehat{S}(t|X, Y)$.

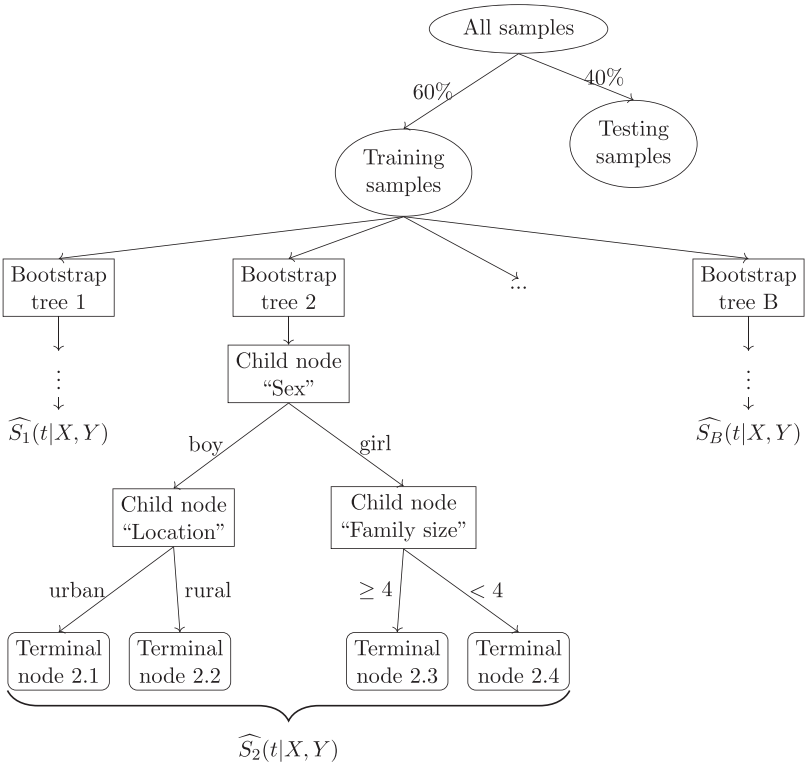

Fig. 1 Macro-level procedure

Measuring the Prediction Performance Across Models With the Brier Score

I compare the out-of-sample predictive performance across models with the Brier score. The Brier score at time t is conceptually similar to the mean squared error of prediction, which is used to evaluate predictive performance in the random forest. The expected Brier score is defined as the average of squared survival probability difference at time t between the actual observation and the prediction (Gerds and

Schumacher 2006). The Brier score deals with the right-censoring of the survival data through inverse probability weighting. For more details, see Mogensen et al. (2012).

After calculating the Brier score in each period, I calculate the integrated Brier score (IBS), which is a cumulative prediction error over time given by $IBS = 1/\max(t)\int_0^{\max(t)} BS(t)dt$. Lower values of IBS statistics indicate better overall predictive performances. An IBS score from the Kaplan-Meier estimation, in which no information regarding covariate distributions is exploited, can be used as a useful benchmark (Mogensen et al. 2012).

Variable Importance Measure

Unlike the Cox-PH and parametric survival regression approaches, the RSF does not need to explicitly specify the hazard or survival functions. Hence, there is no explicit *p* value or significance test for variable selection. Instead, the RSF calculates the variable importance (VIMP) to ascertain which variables contribute to the prediction (Breiman 2001). The VIMP computation involves "noising-up" each variable in turn. Specifically, VIMP for variable x_k is computed as the difference between (1) the prediction error when x_k is randomly permuted and (2) the prediction error under the observed values (Ishwaran et al. 2008). Intuitively, VIMP for x_k measures the change in prediction error in the test data if information regarding x_k is not available. A large positive value of VIMP implies that the corresponding variable is an important predictor. VIMP values close to 0 indicate that the variable makes little contribution to predictive accuracy.

Another alternative measurement for variable importance is *minimal depth*, which evaluates the relative influence of each variable in constructing the forest based on the timing when each independent variable is used for node splitting (Ishwaran and Kogalur 2010; Ishwaran et al. 2011). Variables with high influence on the prediction are assumed to be those that most frequently split nodes nearest to the root node, which partitions the largest samples of the data. Within each tree, node levels are numbered based on their relative distance to the root of the tree (with the root at 0). Minimal depth measures important factors by averaging the minimal depth for each variable over all trees within the forest. Smaller minimal depth values indicate that the variable separates large groups of observations and therefore has a large influence on the forest construction and prediction.

Decomposition of the Survival Curve

After modeling the conditional survival function with the RSF and the Cox-PH, I predict the survival curves for each period: the 1990s, 2000s, and 2010s. Using the testing samples in respective year periods, I obtain the predicted values for each subject *i* in each year period—that is, $\hat{S}(t \mid X_i, Y_i)$.[7] Taking averages over *X* among subjects

[7] For example, the survival curve in 1990s is estimated by $\hat{S}(t|Y_{1990s}) = \int_{\chi} \hat{S}(t|Y_{1990s} = 1, Y_{2000s} = 0, Y_{2010s} = 0)dF_{X|Y_{1990s}=1}$.

in each year period $m = \{1990s, 2000s, 2010s\}$, I obtain $\hat{S}(t \mid Y_m) = 1 / N_{Y_m} \hat{S}(t \mid X_i, Y_i)$, where N_{Y_m} is a number of subjects belonging to the year group Y_m. The improvement in the predicted survival rates at time t for the three decades is expressed by

$$\Delta(t) \equiv \hat{S}\left(t \mid Y_{2010s}\right) - \hat{S}\left(t \mid Y_{1990s}\right). \tag{1}$$

In the spirit of Oaxaca (1973) and Blinder (1973), I decompose $\Delta(t)$ into (1) the part that is associated with the difference in the covariates (explained effect), and (2) the remaining part that is associated with the difference in the hazard functions (unexplained effect). Equation (1) is therefore decomposed as follows:

$$\Delta(t) = \underbrace{\hat{S}(t \mid Y_{2010s}) - \hat{S}^{CF}(t)}_{\text{Explained effect}} + \underbrace{\hat{S}^{CF}(t) - \hat{S}(t \mid Y_{1990s})}_{\text{Unexplained effect}}, \tag{2}$$

where $\hat{S}^{CF}(t)$ is the counterfactual survival function composed of the hazard function in the 2010s and covariate distributions in the 1990s.

Counterfactual Survival Function

By the law of iterated expectations, we have

$$\hat{S}\left(t \mid Y_{2010s}\right) = \int_{\chi} \hat{S}\left(t \mid X, Y_{2010s}\right) dF_{X \mid Y_{2010s}}, \tag{3}$$

$$\hat{S}\left(t \mid Y_{1990s}\right) = \int_{\chi} \hat{S}\left(t \mid X, Y_{1990s}\right) dF_{X \mid Y_{1990s}}. \tag{4}$$

The counterfactual survival function in Eq. (2) is composed of the conditional survival function in the 2010s integrated over the covariate distribution in the 1990s. Under the common support assumption requiring that the covariate space of the 1990s is included in the 2010s, the counterfactual survival function is expressed by

$$\hat{S}^{CF}\left(t\right) = \int_{\chi} \hat{S}\left(t \mid X, Y_{2010s}\right) dF_{X \mid Y_{1990s}}. \tag{5}$$

Hence, in Eq. (2), the unexplained effect reflects the change in survival probability due to the change in the conditional survival function—that is, the difference between $\hat{S}(t \mid X, Y_{2010s})$ and $\hat{S}(t \mid X, Y_{1990s})$. On the other hand, the explained effect reflects the change in survival probability due to the change in the distribution of observable characteristics—that is, the difference between $F_{X \mid Y_{2010s}}$ and $F_{X \mid Y_{1990s}}$.

In contrast to the standard Oaxaca Blinder decomposition (Blinder 1973; Oaxaca 1973), where conditional mean functions are usually separately estimated by a linear additive model in each year group before decomposition, I estimate a single conditional survival function. Because the RSF can model interacted effects between X and Y, conditional survival functions for respective year groups can be obtained from the single RSF model.

Results

Kaplan-Meier Estimates of the Survival Probabilities

Before estimating the conditional survival functions by the RSF, I examine the non-parametric survival function estimates as descriptive results. Figure 2 illustrates the Kaplan-Meier estimates of the infant survival probabilities in the six countries. All six countries experienced substantial improvements in infant mortality over the three decades. Log-rank tests indicate evidence of significant improvements between the 1990s and 2010s ($p < .01$). A notable improvement is found in Bangladesh, where the probability of infant survival in the first 12 months rose from 97.1% (95% confidence interval (CI): [96.8%, 97.5%]) to 99.1% (CI: [98.9%, 99.3%]) between the 1990s and 2010s. The Philippines, on the other hand, exhibits the smallest improvement among the six countries, from 98.5% (CI: [98.2%, 98.7%]) to 99.2% (CI: [99.0%, 99.4%]) over the same period. The Philippines' comparatively smaller improvement, although still exceptional, can be attributed mostly to its survival probability in the 1990s, which was much greater than that of the other five countries; in this decade, only the Philippines had a survival probability of more than 98% in the first 12 months. Given that the Philippines still shows the highest survival probabilities in the 2010s, the relatively smaller progress can be attributed to the fact that there was less room for further improvement in this country.

Decomposition Results

The parameter tuning finds that $d = \{50,60,10,70,10,10\}$ are optimal terminal node sizes for Bangladesh, India, Indonesia, Nepal, Pakistan, and the Philippines, respectively.[8] Figure 3 shows the out-of-sample error rates of the Kaplan-Meier estimates (reference), Cox-PH without interaction terms, Cox-PH with two-way interaction effects, selective Cox-PH with two-way interaction effects, and the RSF. The RSF shows the smallest prediction error of all methods, although the 95% confidence intervals of each method overlap. As expected, the error of the Kaplan-Meier estimators, which does not use the covariate information to predict survival probability, exhibits the largest value. I do not find clear evidence that the Cox-PH two-way interaction model has smaller out-of-sample errors than the Cox-PH without interaction terms. Henceforth, I discuss the predicted survival probabilities and decomposition results estimated by the RSF. The estimated coefficients and survival curves predicted by the Cox-PH are available upon request.

Panel a in each of the remaining figures illustrates the predicted survival curves in the 1990s, 2000s, and 2010s, in which only the testing data are used for prediction. Panel a also shows the predicted survival probability in the counterfactual scenario in which infants born in the 2010s are assumed to have the distributions of covariates observed in the 1990s. Hence, the difference between the survival probability in the

[8] Prediction performance for each node size in the respective countries is available upon request.

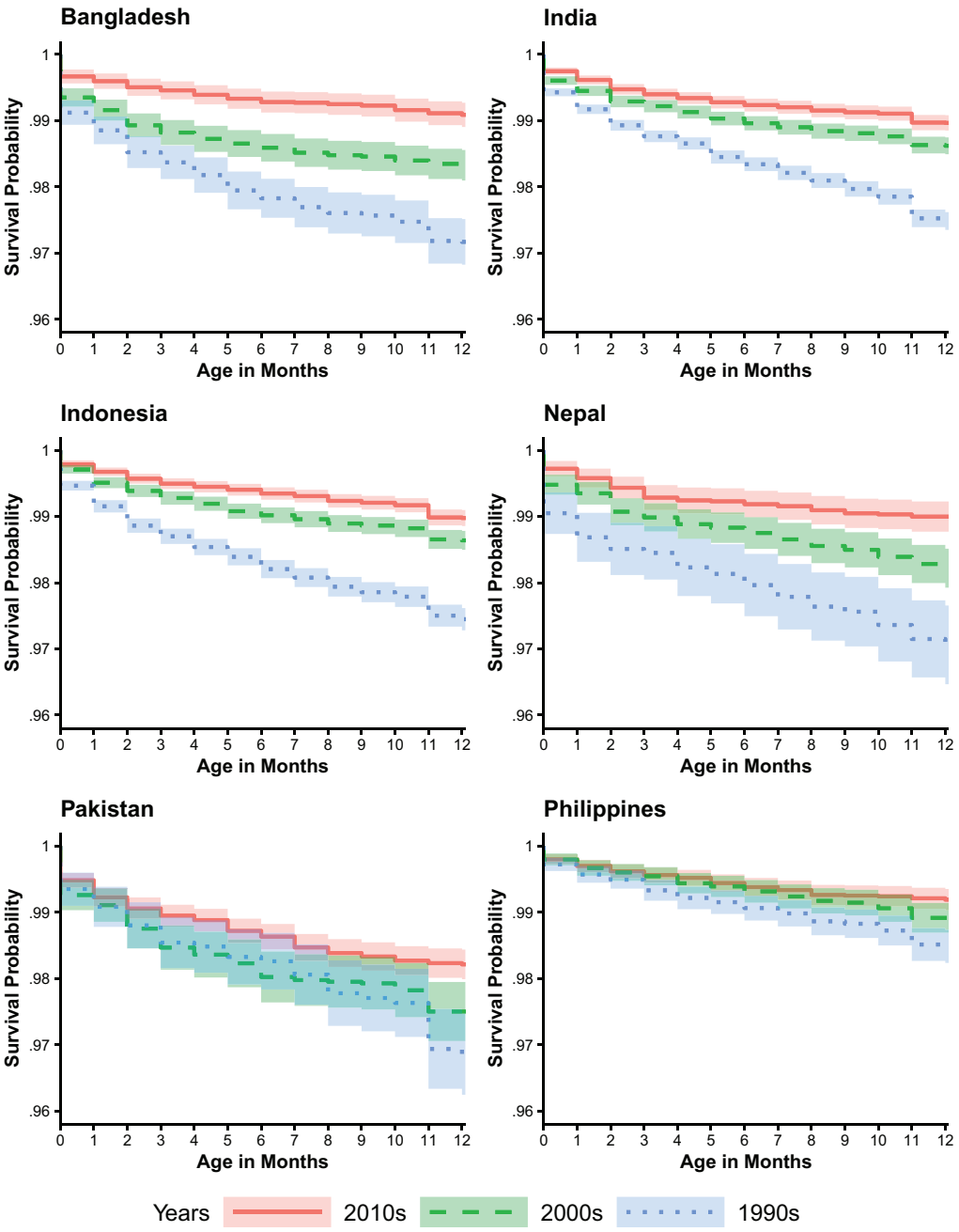

Fig. 2 Kaplan-Meier estimates, with 95% confidence intervals highlighted

2010s and the counterfactual survival probability is associated with the improvement in infant mortality due to the change in the covariate distribution (explained effect). On the other hand, the difference between the probability in the 1990s and the counterfactual survival probability is the remaining improvement in the infant mortality rate that cannot be explained by the change in the distribution of X (the unexplained effect). The unexplained effect can also be interpreted as the part of the total improvement associated with the change in the hazard function over time or the improvement triggered by the unobserved characteristics, such as the improved quality of maternal healthcare.

Bangladesh

Figure 4 shows the results in Bangladesh. Panel a highlights the widening difference in the survival curves in the 2010s and 1990s across evaluating points in time. For example, the infant survival probabilities in the first month after birth are 99.0% and 99.7% in the 1990s and 2010s, respectively—a difference of only 0.7 percentage points. On the other hand, the probabilities of surviving the first 12 months are 97.6% and 99.4%, respectively, and the difference becomes as large as 1.8 percentage points. The counterfactual survival function lies between the two survival curves in the 1990s and 2010s. The RSF predicts that if infants in the 2010s were exposed to the circumstantial environments of the 1990s, the probability of surviving the first 12 months would be 98.2%. Of the observed 1.8 percentage point difference in the survival probability in the first 12 months, 1.2 percentage points are associated with the improved circumstantial environments over time, and the remaining 0.6 percentage points are due to the change in underlying hazard functions.

Panel b in Figure 4 plots the transition of explained and unexplained effects against evaluating survival time. The explained effect is larger than the unexplained effect in all evaluating points. It consistently accounts for more than 60% of the total improvement in survival rate across time. The relative proportion of the explained effect shows its peak at the first month, which implies that the improved environments have played a major role particularly in the first month.

Panel c shows the variable importance in the RSF. The variable importance measures the contribution made to the survival prediction for each variable and the top 15 variables are shown herein. Family size, birth order, roof condition, birth spacing, and amount of antenatal healthcare use are key factors in the prediction. Hence, the changes in the distribution of these variables are likely to be the main contributors to the explained effect, accelerating the infant survival improvement over time. Panel d shows an alternative measure of the variable importance: the minimum depth. The smaller value thereof is associated with a higher influence of that variable in constructing the forest. Panel d indicates a very similar result to the one shown in panel c. Both panels indicate that year indicators are also included in the list of the top 15 important variables, suggesting that there were structural changes in survival functions conditional on the changes in the covariate distributions. Being a twin is also included among the 10 most important variables in both panels c and d. However, this interpretation should be treated with some caution given that the proportion of twins has been stable over the three decades. Being a twin is an important determinant of

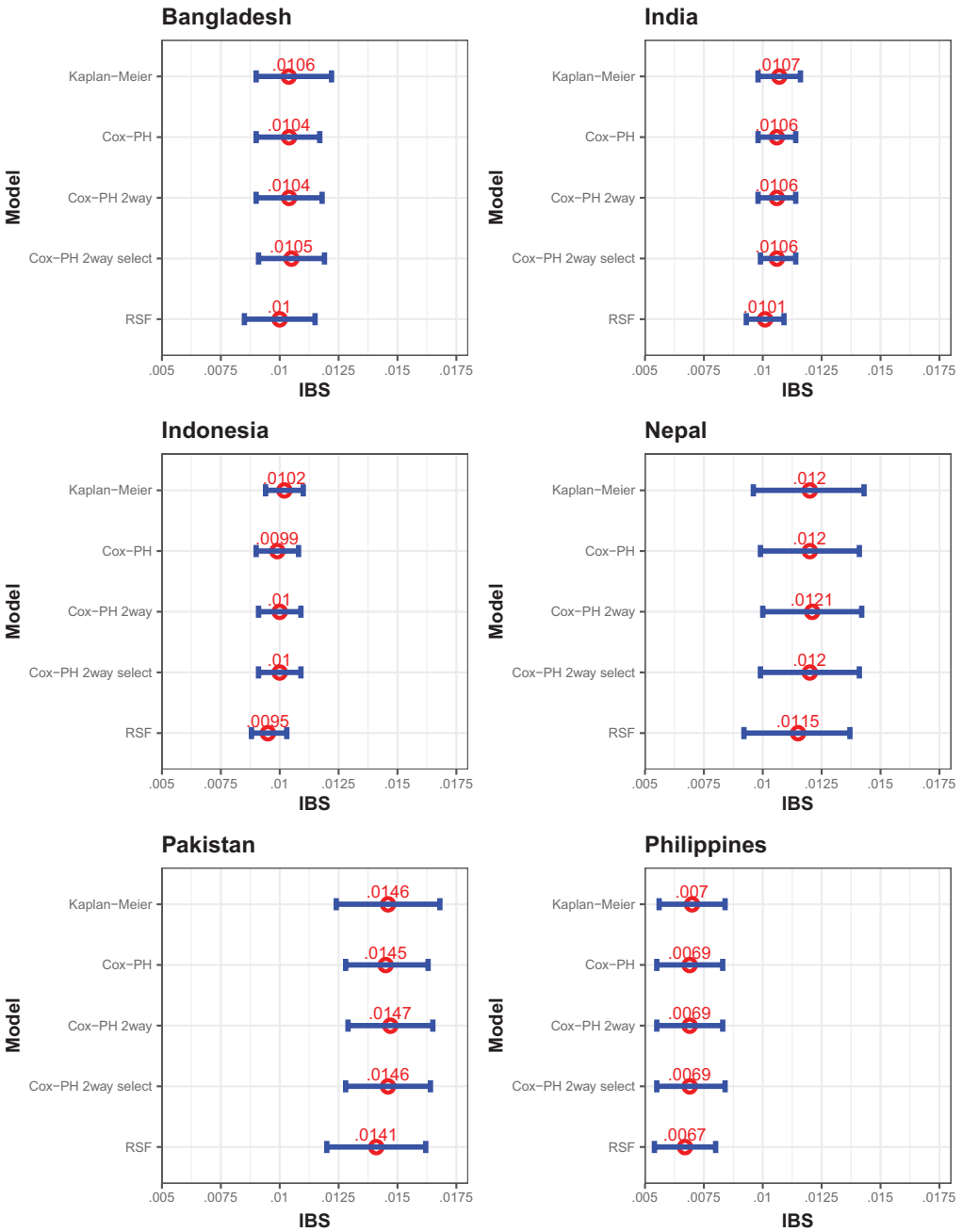

Fig. 3 Out-of-sample error rate, with 95% confidence intervals calculated from bootstrapping with 100 repetitions. Testing samples are used for prediction. IBS = integrated Brier score. Cox-PH 2way = Cox proportional hazard two-way interaction model. Cox-PH 2way select = Cox proportional hazard two-way interaction selected model.

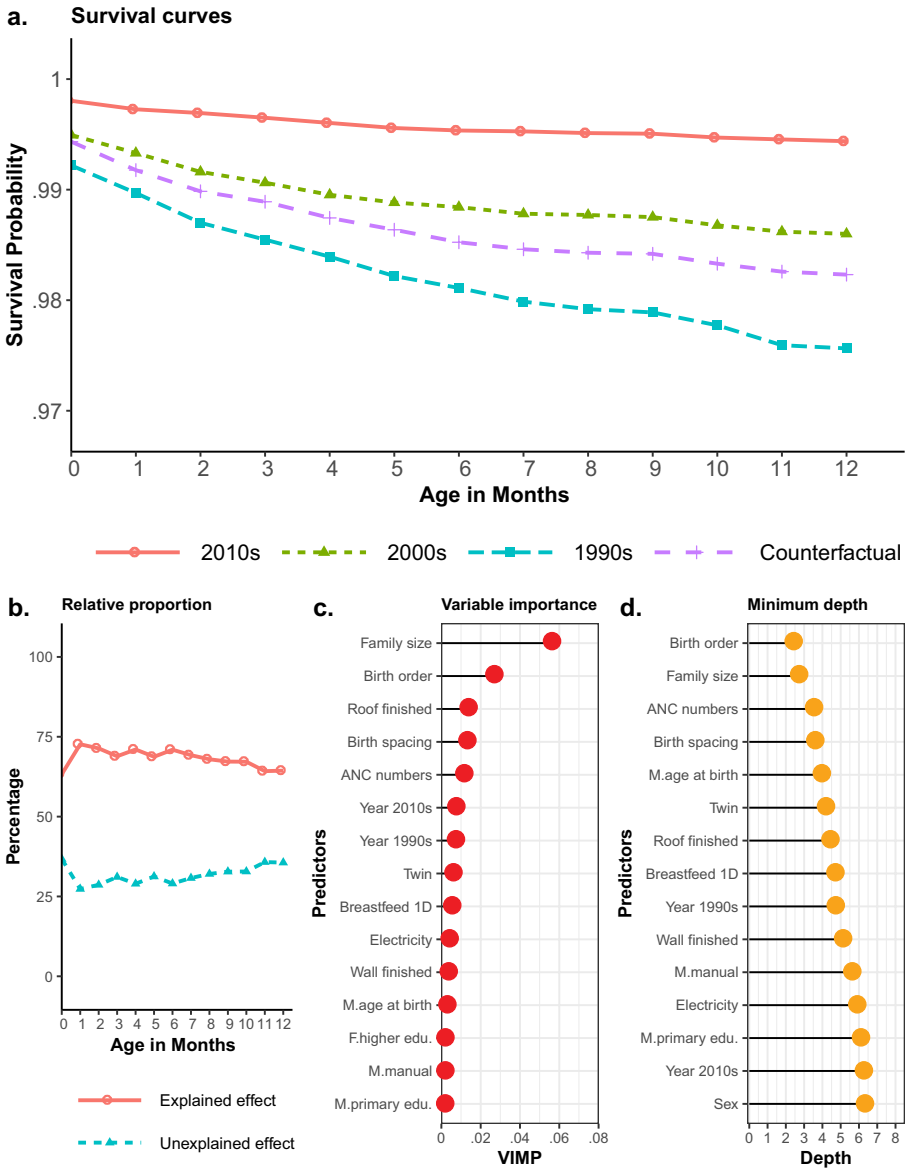

Fig. 4 Predicted survival probabilities and decomposition in Bangladesh. Testing samples are used for prediction. The counterfactual survival function is based on the health production function in the 2010s and covariates in the 1990s. ANC = antenatal care; TT = tetanus toxoid; pub.inst. = public institute; F. = father; M. = mother; 1D = one day; preg. = pregnancy; and edu. = education.

survival, but being important does not necessarily mean that it triggers the survival improvement. Given significant reductions in family size, an increase in antenatal healthcare use, and an improvement in housing conditions in Bangladesh, it is more straightforward and plausible to attribute the survival improvement to these changes across time.

India

Figure 5 shows the results in India. In panel a, the improvement in predicted survival rates is observed at every evaluating point in time. As found in Bangladesh, a growing difference exists in the two survival curves in the 1990s and 2010s across the evaluating time points. The counterfactual survival probabilities are predicted between the two survival probabilities in the 1990s and 2010s.

Panel b indicates that although the contribution of the explained effect is larger than that of the unexplained effect over the evaluation points, the relative proportions of the explained and unexplained effects to the overall improvement exhibit some variation across evaluation points in time. Panel b indicates the largest contribution of the explained effect to survival at the first two-month point.

In panel c, variable importance measures show that family size, birth order, amount of antenatal healthcare use, owning a television, and uptake of tetanus toxoid injections are among the top five important factors in predicting survival. Panel d also shows that family size, birth order, amount of antenatal care use, uptake of tetanus toxoid injections, and electricity access are influential determinants. These two panels imply that the reduction in family size as well as the increased use of antenatal care were associated with the reduction in infant mortality in India.

Indonesia

Figure 6 shows the results in Indonesia, revealing a substantial improvement in the survival probability over time. Panel a shows that the probability rates for Indonesian infants surviving the first 12 months in the 2010s and 1990s are 99.1% and 97.5%, respectively. The counterfactual survival curve exhibits a trend similar to that of the survival curve in the 1990s, thereby suggesting that the circumstantial environments are important factors in predicting survival probabilities. Hence, relatively smaller unexplained effects are observed.

According to panel b, more than 75% of the improvement in infant survival rates from the 1990s to 2010s is attributable to the improvements in the distribution of the observed covariates, thus implying that improvements in the environments strongly contributed to the reduction in infant mortality rates over the last three decades. Compared with other countries in this study, Indonesia shows a higher proportion of the relative contribution of the explained effect and a smaller variation in the relative contribution sizes of the explained and unexplained effects over the evaluation points in time.

In panel c, variable importance measures suggest that family size, birth order, amount of antenatal healthcare use, mother being a farmer, and birth spacing are the five most important variables in the forest. Minimum depth in panel d indicates that amount of antenatal care use, family size, birth order, access to electricity, and roof condition are the five most influential determinants of the survival probability prediction.

Nepal

The results for Nepal are shown in Figure 7, which shows the widening improvements in the survival probabilities for all evaluating time points. In panel a, the coun-

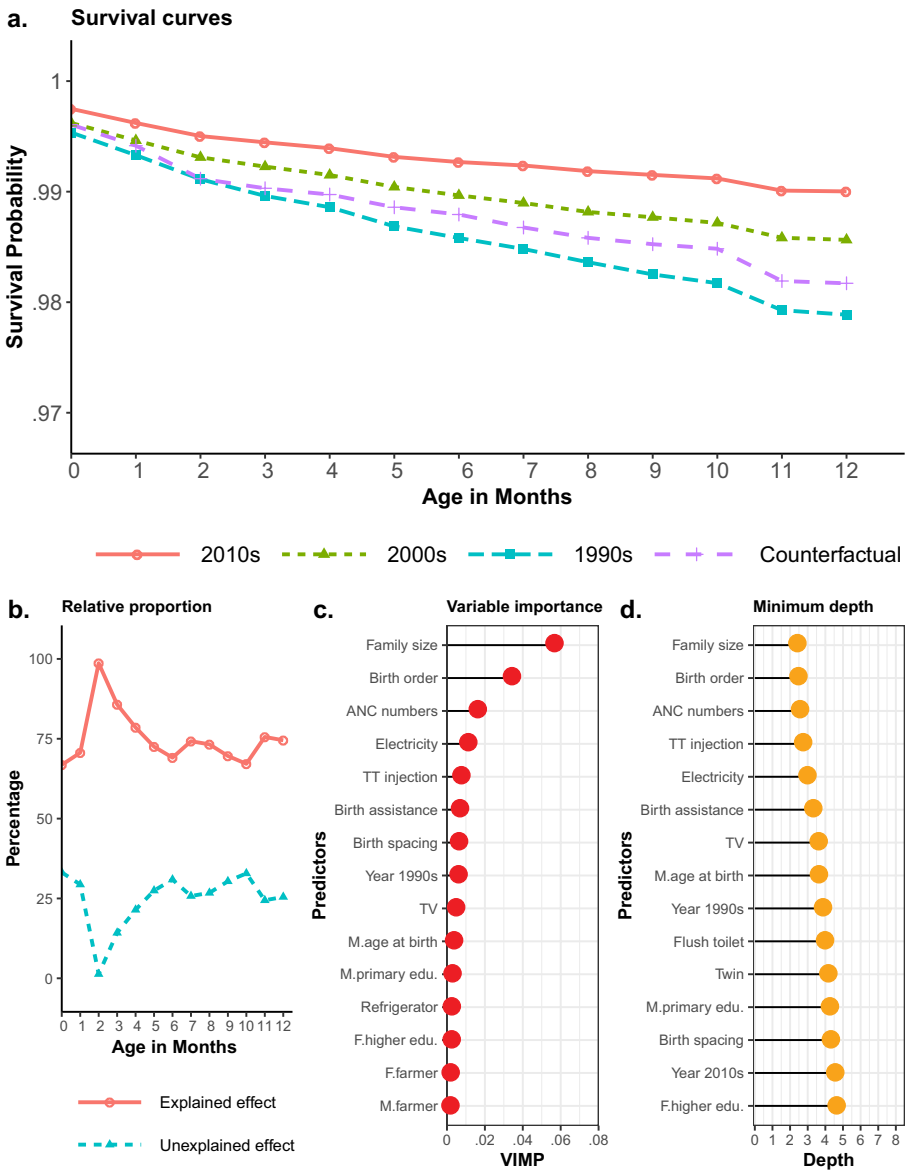

Fig. 5 Predicted survival probabilities and decomposition in India. Testing samples are used for prediction. The counterfactual survival function is based on the health production function in the 2010s and covariates in the 1990s. ANC = antenatal care; TT = tetanus toxoid; pub.inst. = public institute; F. = father; M. = mother; 1D = one day; preg. = pregnancy; and edu. = education.

terfactual survival curve is between the two survival curves in the 1990s and 2010s. The RSF predicts that the probability of surviving the first 12 months is 97.2% in the 1990s and 99.1% in the 2010s. The difference is 1.9 percentage points, of which 0.7 percentage points (equivalent to 39.7%) can be explained by the improved environments.

Panel b shows that in contrast to the other countries analyzed in this study, the unexplained effect is consistently larger than the explained effect in Nepal, suggest-

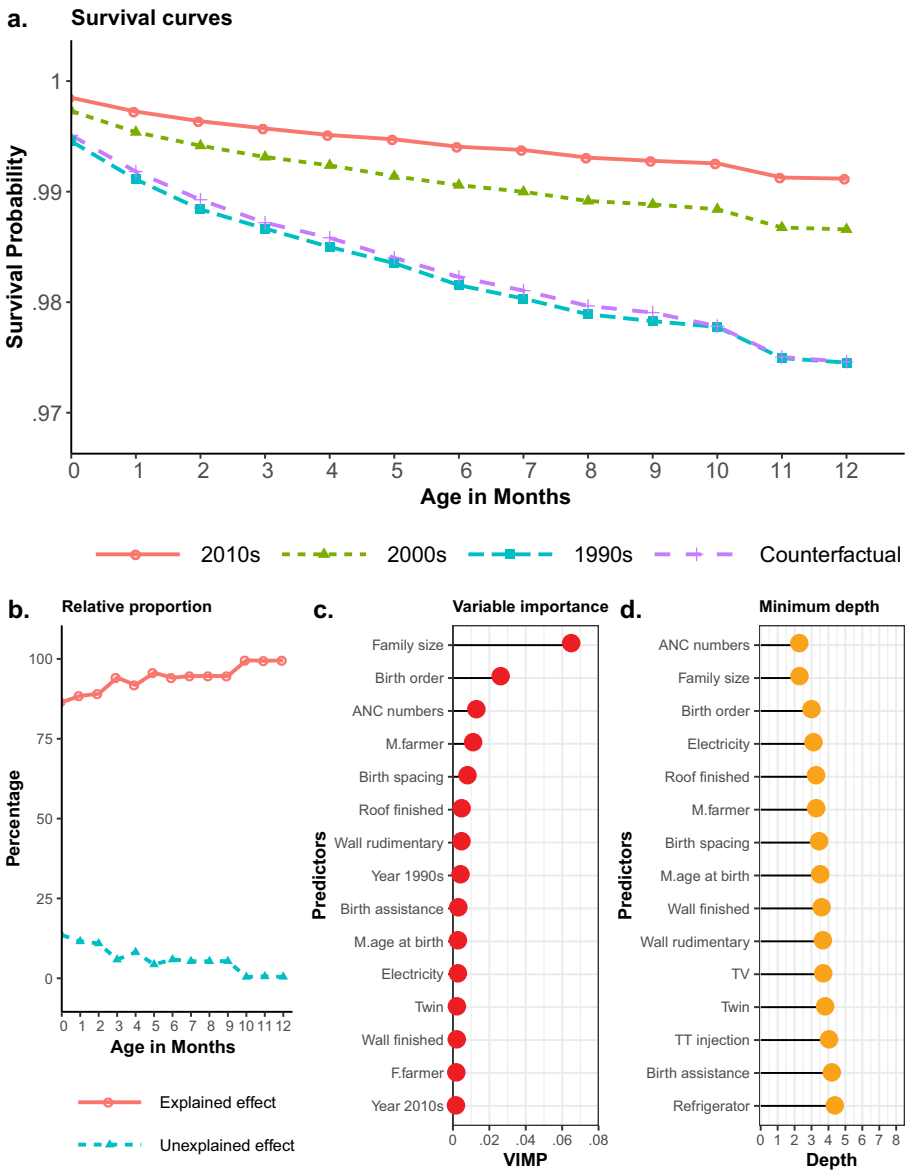

Fig. 6 Predicted survival probabilities and decomposition in Indonesia. Testing samples are used for prediction. The counterfactual survival function is based on the health production function in the 2010s and covariates in the 1990s. ANC = antenatal care; TT = tetanus toxoid; pub.inst. = public institute; F. = father; M. = mother; 1D = one day; preg. = pregnancy; and edu. = education.

ing that the improvement in the Nepalese infant mortality rate from the 1990s to 2010s is due to the changes in hazard functions rather than the changes in the distribution of observable covariates. It also suggests the possibility that the improvement in infant survival was associated more with factors that are not considered in this study, such as the improvement in the quality of maternal care. The relative contribution

made by the unexplained effect becomes smaller as the evaluation points come closer to the 12-month point.

The variance importance measurements in panel c suggest that family size, uptake of tetanus toxoid injections, amount of antenatal care use, birth order, and breastfeeding are major contributing factors. Moreover, year indicators are included in the 10 most important factors, which is in line with the finding that the unexplained effect makes a large contribution to explaining the improvement. Panel d shows that family size, amount of antenatal care use, uptake of tetanus toxoid injections, birth order, and mother's age at birth are influential determinants of survival. Panel d also indicates that year indicators are influential prediction factors.

Pakistan

Figure 8 shows the results for Pakistan. Panel a reveals substantial improvements over the three decades in the survival probabilities for every evaluating point in time. The counterfactual survival curve lies between the curves in the 1990s and 2010s. The RSF predicts that the probability of surviving the first 12 months is 97.3% in the 1990s and 98.3% in the 2010s. The improvement in initial endowments accounts for 53.3% of the difference in the survival probabilities.

Panel b indicates that up to the first six months, the size of unexplained effect is larger than that of explained effect. At age 6 months, the relative size of explained effect becomes larger than that of unexplained effect. After age 11 months, the explained and unexplained effects contribute almost equally to the improvement. These changes imply that the improvements in hazard function and those in covariate distributions contribute to improving the infant mortality rates differently over the evaluating point in time.

The variance importance measurements in panel c suggest that family size, birth order, birth spacing, the amount of antenatal care use, and access to electricity are major contributing factors. Panel d also suggests that the improved infant survival probability in Pakistan was associated with the reduction in family size coupled with increased use of antenatal care.

Philippines

Finally, in the Philippines (Figure 9), the improvement in survival rates over the three decades is smaller in an absolute term at every evaluating point in time than in the other countries. This smaller improvement is consistent with the initial observation in the Kaplan-Meier estimates (Figure 2). This result may be attributable to the country's initially relatively advantaged survival probability in the 1990s, which allowed little room for improvement. Panel a shows that larger parts of the improvement in survival probability are explained by the improvement in the covariates in the model. The probability of surviving the first 12 months after birth is 99.3% in the 2010s and 98.7% in the 1990s—a difference of 0.6 percentage points. The RSF suggests that 69.1% of this difference (corresponding to 0.4 percentage points) is explained by the improvement in covariate distributions. Panel b highlights how the relative

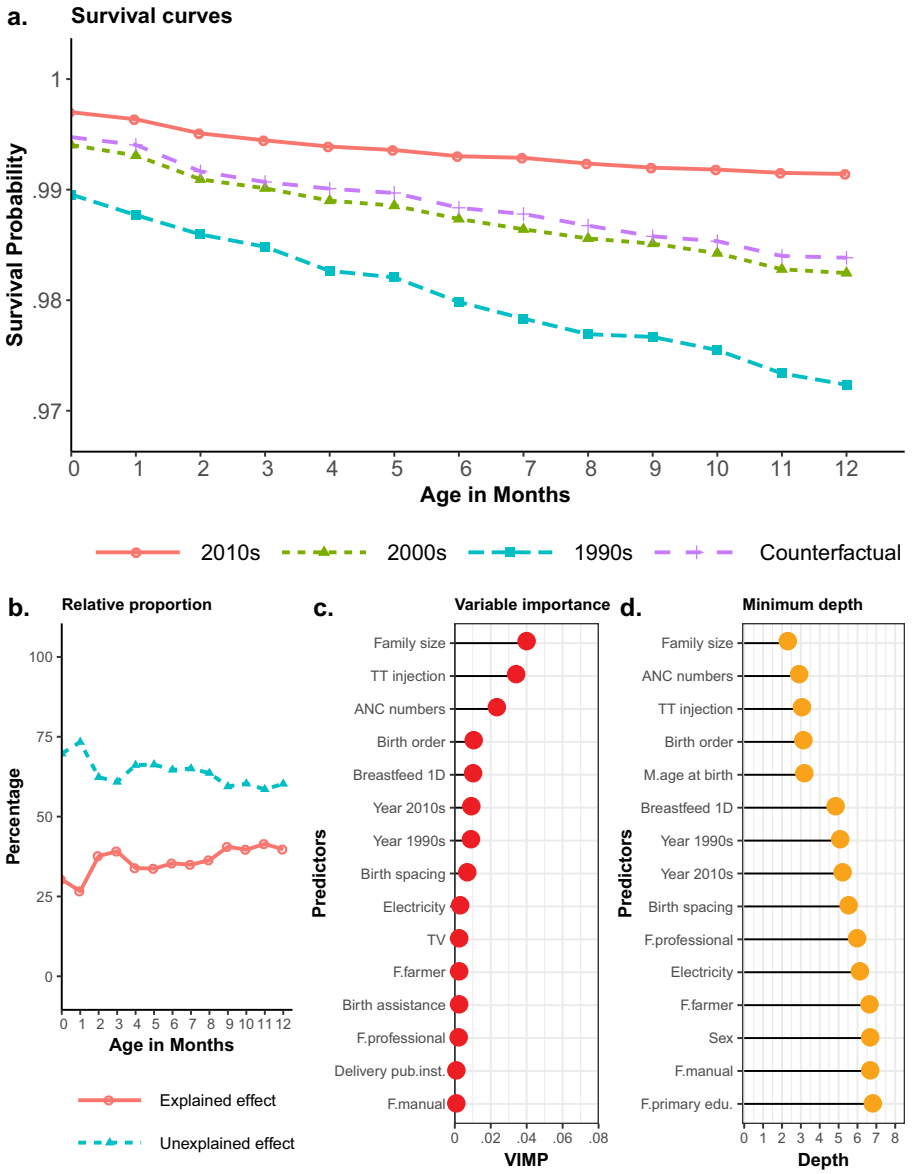

Fig. 7 Predicted survival probabilities and decomposition in Nepal. Testing samples are used for prediction. The counterfactual survival function is based on the health production function in the 2010s and covariates in the 1990s. ANC = antenatal care; TT = tetanus toxoid; pub.inst. = public institute; F. = father; M. = mother; 1D = one day; preg. = pregnancy; and edu. = education.

proportion of the explained effect is peaked at the survival probability in the first five months, implying that the improved environments have played an important role especially in the first five months, after which it shows a slower decrease over time.

In panel c, the variable importance measures indicate that family size, birth order, amount of antenatal care use, uptake of tetanus toxoid injections, and father being a

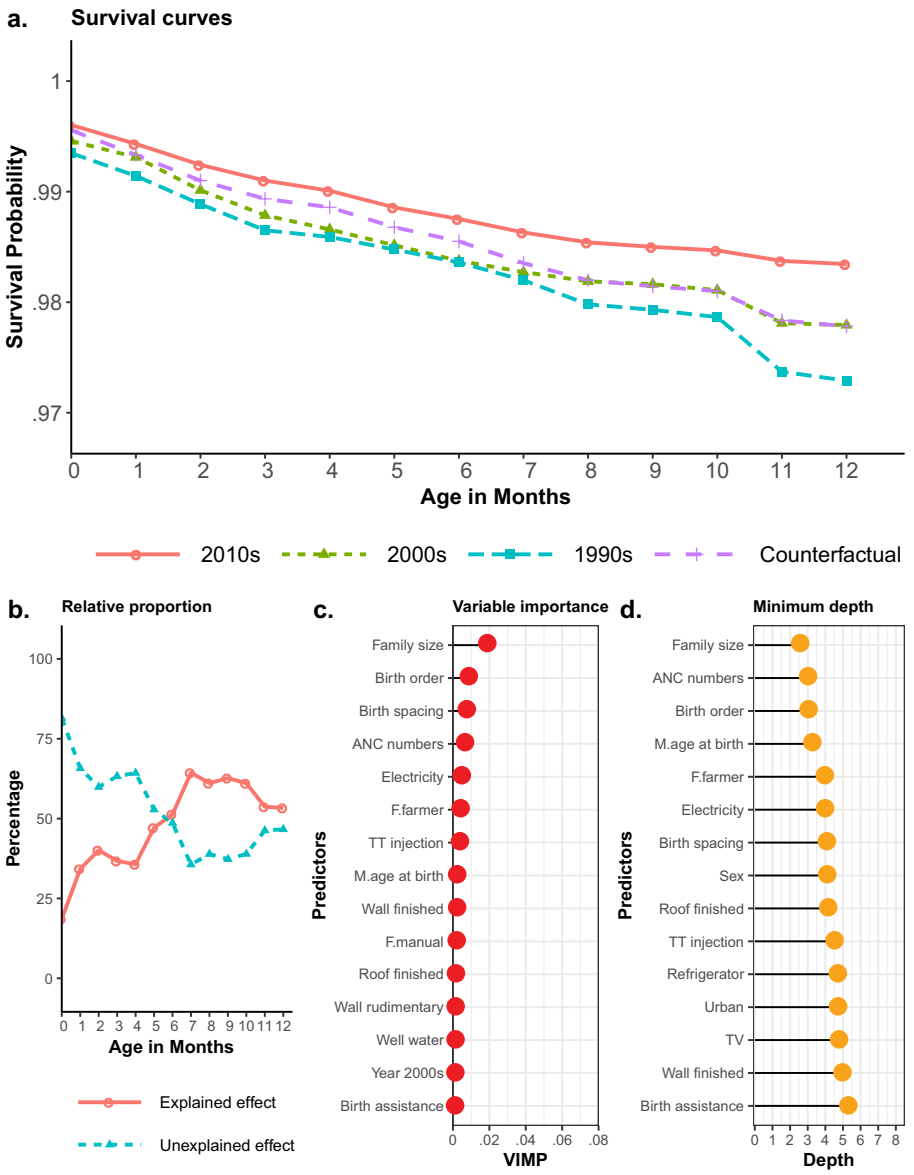

Fig. 8 Predicted survival probabilities and decomposition in Pakistan. Testing samples are used for prediction. The counterfactual survival function is based on the health production function in the 2010s and covariates in the 1990s. ANC = antenatal care; TT = tetanus toxoid; pub.inst. = public institute; F. = father; M. = mother; 1D = one day; preg. = pregnancy; and edu. = education.

farmer are important factors in the prediction. Panel d indicates that birth order, amount of antenatal care use, maternal age at birth, family size, and uptake of tetanus toxoid injections are variables with high influence in survival probability prediction. Both panels suggest that the reduction in family size coupled with increased use of antenatal care are associated with the improved infant survival probability in the Philippines.

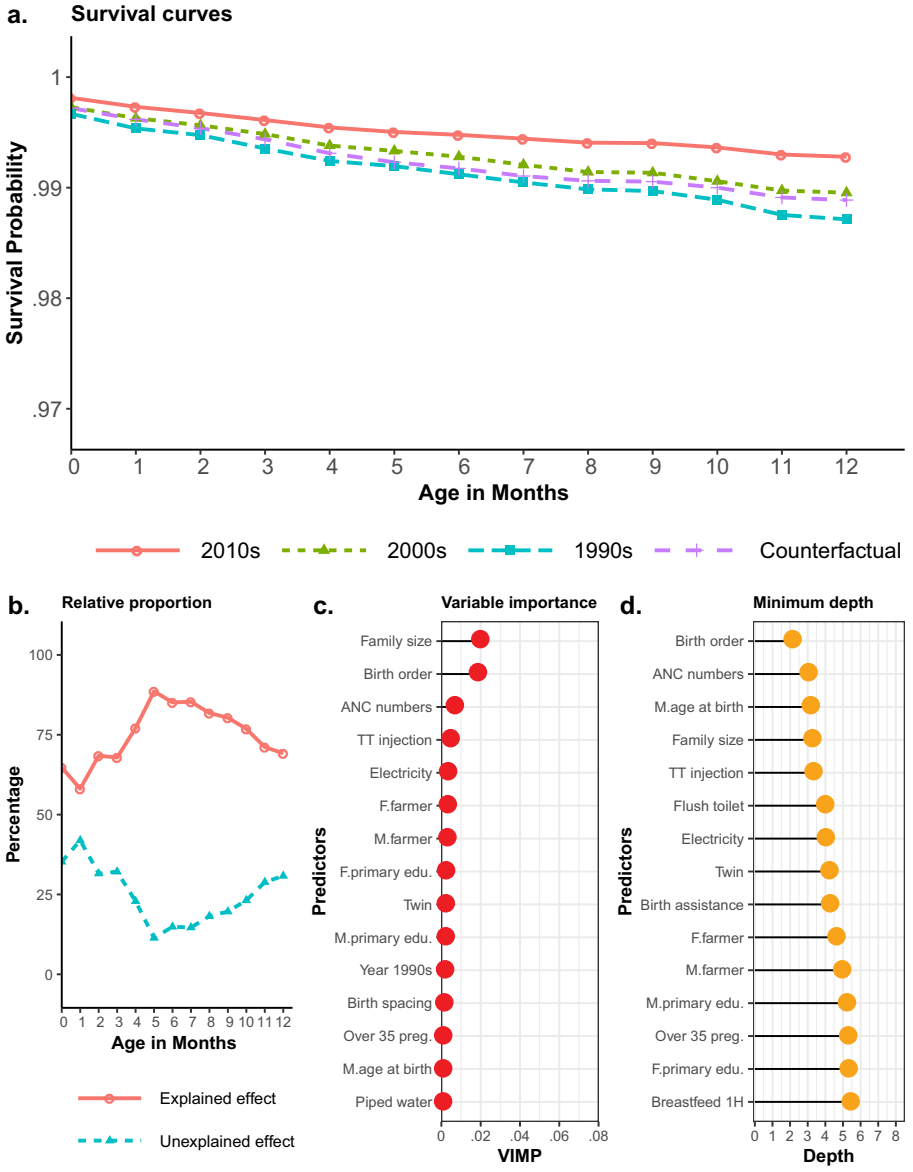

Fig. 9 Predicted survival probabilities and decomposition in the Philippines. Testing samples are used for prediction. The counterfactual survival function is based on the health production function in the 2010s and covariates in the 1990s. ANC = antenatal care; TT = tetanus toxoid; pub.inst. = public institute; F. = father; M. = mother; 1D = one day; preg. = pregnancy; and edu. = education.

Discussion and Conclusion

Low- and middle-income Asian countries have witnessed eye-catching improvements in infant mortalities over the last three decades (from the 1990s to the 2010s) (UNICEF 2013, 2017). Through decomposition analysis, this study closely exam-

ined these improvements in Bangladesh, India, Indonesia, Nepal, Pakistan, and the Philippines. I decomposed the improvements in infant survival probabilities into the explained effect associated with the improvements in infants' circumstantial environments and the unexplained effect associated with the improvement in the hazard function itself and unobservable factors, such as healthcare quality. This decomposition analysis quantifies how much of the increase in infant survival probability is due to the improvement in the circumstantial environments in the respective countries. One of the key features of this study is the use of the fully data-adaptive machine-learning method, the random survival forest (RSF), which has achieved high prediction performance in previous studies (Dietrich et al. 2016; Imani et al. 2019; Yosefian et al. 2015). I apply the RSF to model infant survival probability and predict the counterfactual probability that infants in the 2010s would face if they had the distributions of household characteristics, parental socioeconomic status, and antenatal healthcare use of the 1990s. The results show that in Bangladesh, India, Indonesia, and the Philippines, the explained effect is consistently larger than the unexplained effect, suggesting that the infant mortality reductions are largely attributable to the observed improvement in infants' environments. The results also suggest that the reductions in family size, increased use of antenatal care, and improved household living standards were strongly associated with improvements in infant survival rates.

One of the limitations of this study is that I included only the variables that were readily available in all six countries across the three decades. I was not able to include other variables of potential interest, such as mothers' intake of iron and vitamin A tablets during pregnancy, the frequency of postnatal checkups, a detailed history of child immunization uptake, use of sleeping nets to prevent malaria, maternal nutritional status, and maternal HIV history because the relevant information was available only in the recently collected DHS data. Moreover, I could not include community-level indicators, such as the degree of local air pollution (one of the leading causes of lower respiratory diseases), water quality, and heterogeneity of healthcare quality provided in local health facilities. Including information about these community-level indicators could promote more thorough research into the infant mortality rates across countries. Having said that, this study included more than 45 variables relating to circumstantial environments, and the RSF effectively takes into account their higher-order interaction effects as well. Given that the RSF is capable of including many characteristics as potential contributing factors, more detailed research should become possible once more sophisticated micro- and macro-level data become available.

Mitigating the within-country inequality in infant mortality across parental socioeconomic status is also paramount in achieving the objectives of the Sustainable Development Goals (United Nations 2015). In low- and middle-income countries in Asia, women's uptake of maternal care is strongly associated with socioeconomic status, and facilitating the adequate use of maternal care among poor and marginalized women remains a major challenge (Kesterton et al. 2010; Pathak et al. 2010). Although inequality in infant mortality across parental socioeconomic status and its transition over time have not been examined in this study because of space constraints, substantial socioeconomic inequalities in the degree of improvements within countries are also likely. If the improvements in infant mortality have not yet been enjoyed by mothers and children in socioeconomically disadvantaged households, inequality in infant and child mortalities would become a significant bottleneck that

could make sustainable improvements difficult to achieve. In this sense, enhancing the knowledge of the importance of maternal healthcare among marginalized mothers and ensuring opportunities to use it regardless of socioeconomic background will be essential in the coming years. One promising avenue would be to enhance education among poor women. The potential benefits of educational betterment would not be limited to enhanced knowledge because educated mothers are more likely to have longer pregnancy spacing and utilize antenatal and postnatal healthcare (Cleland and van Ginneken 1988; Jain 1985). Moreover, better education should lead to higher income, which could then be spent on purchasing more healthcare goods and services, further contributing to improvements in infant health. Exploring the existence of heterogeneity across various parental socioeconomic status, such as living standards, educational backgrounds, and occupation, would be a promising future research agenda.

When every mother begins to use maternal healthcare appropriately, the quality of care will become the next challenge in developing countries in Asia to mitigate the existing gap between developing and developed countries. The low quality of facilities and poor human resources, such as the high rates of absenteeism, are still being widely witnessed. Overcoming these challenges and ensuring opportunities to receive high-quality care would further improve the infant mortality rates and help close the existing gap between developing and developed countries.

Another promising future research route would be to expand on the countries and regions covered in the analysis—for example, to sub-Saharan Africa, where numerous countries still face high neonatal, infant, and child mortality rates (World Bank 2019). Some countries are gradually overcoming these problems, but most are still struggling with various political, economic, and cultural bottlenecks. Further research in the context of sub-Saharan Africa would help to elucidate necessary health policies to further reduce the mortality rates and eradicate avoidable deaths. ■

Acknowledgments This study used the Demographic Health Surveys (DHS). I acknowledge the original collectors of the data, authorized distributors of the data, survey respondents, and relevant funding agencies. They bear no responsibility for the results and interpretations in the study.

References

Adams, M. M., Elam-Evans, L. D., Wilson, H. G., & Gilbertz, D. A. (2000). Rates of and factors associated with recurrence of preterm delivery. *JAMA, 283,* 1591–1596.

Ahsan, K. Z., Arifeen, S. E., Al-Mamun, M. A., Khan, S. H., & Chakraborty, N. (2017). Effects of individual, household and community characteristics on child nutritional status in the slums of urban Bangladesh. *Archives of Public Health, 75*(9), 1–13.

Aizawa, T. (2019). Ex-ante inequality of opportunity in child malnutrition: New evidence from ten developing countries in Asia. *Economics & Human Biology, 35,* 144–161.

Akseer, N., Kamali, M., Arifeen, S. E., Malik, A., Bhatti, Z., Thacker, N., . . . Bhutta, Z. A. (2017). Progress in maternal and child health: How has South Asia fared? *BMJ, 357,* 1–6.

Andreev, E. M., & Kingkade, W. W. (2015). Average age at death in infancy and infant mortality level: Reconsidering the Coale-Demeny formulas at current levels of low mortality. *Demographic Research, 33,* 363–390. https://doi.org/10.4054/DemRes.2015.33.13

Blau, D. M., Guilkey, D. K., & Popkin, B. M. (1996). Infant health and the labor supply of mothers. *Journal of Human Resources, 31,* 90–139.

Blinder, A. S. (1973). Wage discrimination: Reduced form and structural estimates. *Journal of Human Resources, 8,* 436–455.

Bradley, R. H., Corwyn, R. F., McAdoo, H. P., & Garcia-Coll, C. (2001). The home environments of children in the United States—Part I: Variations by age, ethnicity, and poverty status. *Child Development, 72,* 1844–1867.

Breiman, L. (2001). Random forests. *Machine Learning, 45,* 5–32.

Campbell, O. M., & Graham, W. J. (2006). Strategies for reducing maternal mortality: Getting on with what works. *Lancet, 368*, 1284–1299.

Caulfield, L. E., de Onis, M., Blssner, M., & Black, R. E. (2004). Undernutrition as an underlying cause of child deaths associated with diarrhea, pneumonia, malaria, and measles. *American Journal of Clinical Nutrition, 80,* 193–198.

Checkley, W., Gilman, R. H., Black, R. E., Epstein, L. D., Cabrera, L., & Sterling, C. R. (2004). Effect of water and sanitation on childhood health in a poor Peruvian peri-urban community. *Lancet, 363,* 112–118.

Chisti, M. J., Tebruegge, M., La Vincente, S., Graham, S. M., & Duke, T. (2009). Pneumonia in severely malnourished children in developing countries—Mortality risk, aetiology and validity of WHO clinical signs: A systematic review. *Tropical Medicine & International Health, 14,* 1173–1189.

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

Corsi, D. J., Neuman, M., Finlay, J. E., & Subramanian, S. (2012). Demographic and Health Surveys: A profile. *International Journal of Epidemiology, 41,* 1602–1613.

Cox, D. R. (1972). Regression models and life-tables. *Journal of the Royal Statistical Society: Series B (Methodological), 34,* 187–220.

Dietrich, S., Floegel, A., Troll, M., Kühn, T., Rathmann, W., Peters, A., . . . Drogan, D. (2016). Random survival forest in practice: A method for modelling complex metabolomics data in time to event analysis. *International Journal of Epidemiology, 45,* 1406–1420.

Ezzati, M., Lopez, A. D., Rodgers, A., Vander Hoorn, S., & Murray, C. J. (2002). Selected major risk factors and global and regional burden of disease. *Lancet, 360,* 1347–1360.

Finlay, J. E., Ozaltin, E., & Canning, D. (2011). The association of maternal age with infant mortality, child anthropometric failure, diarrhoea and anaemia for first births: Evidence from 55 low- and middle-income countries. *BMJ Open, 1,* e000226. https://doi.org/10.1136/bmjopen-2011-000226

Fotso, J. C., Cleland, J., Mberu, B., Mutua, M., & Elungata, P. (2013). Birth spacing and child mortality: An analysis of prospective data from the Nairobi Urban Health and Demographic Surveillance System. *Journal of Biosocial Science, 45,* 779–798.

Gerds, T. A., & Schumacher, M. (2006). Consistent estimation of the expected Brier score in general survival models with right-censored event times. *Biometrical Journal, 48,* 1029–1040.

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

Imani, F., Chen, R., Tucker, C., & Yang, H. (2019). Random forest modeling for survival analysis of cancer recurrences. In *2019 IEEE 15th International Conference on Automation Science and Engineering (CASE 2019)* (pp. 399–404). Vancouver, Canada: IEEE.

Ishwaran, H., & Kogalur, U. B. (2007). Random survival forests for R. *R News, 7*(2), 25–31.

Ishwaran, H., & Kogalur, U. B. (2010). Consistency of random survival forests. *Statistics & Probability Letters, 80,* 1056–1064.

Ishwaran, H., Kogalur, U. B., Blackstone, E. H., & Lauer, M. S. (2008). Random survival forests. *Annals of Applied Statistics, 2,* 841–860.

Ishwaran, H., Kogalur, U. B., Chen, X., & Minn, A. J. (2011). Random survival forests for high-dimensional data. *Statistical Analysis and Data Mining, 4,* 115–132.

Islam, T., & Hyder, A. (2016). *A reflection on child and infant mortality in selected South Asian countries* (MPRA Paper No. 86309). Retrieved from https://mpra.ub.uni-muenchen.de/86309/1/MPRA_paper_86309.pdf

Jain, A. K. (1985). Determinants of regional variations in infant mortality in rural India. *Population Studies, 39,* 407–424.

Joenssen, D. W., & Bankhofer, U. (2012). Hot deck methods for imputing missing data. In P. Perner (Ed.), *Machine learning and data mining in pattern recognition* (pp. 63–75). Berlin, Germany: Springer.

Kesterton, A. J., Cleland, J., Sloggett, A., & Ronsmans, C. (2010). Institutional delivery in rural India: The relative importance of accessibility and economic status. *BMC Pregnancy and Childbirth, 10,* 30. https://doi.org/10.1186/1471-2393-10-30

Konteh, F. H. (2009). Urban sanitation and health in the developing world: Reminiscing the nineteenth century industrial nations. *Health & Place, 15,* 69– 78.

Lamberti, L. M., Zakarija-Grković, I., Fischer Walker, C. L., Theodoratou, E., Nair, H., Campbell, H., & Black, R. E. (2013). Breastfeeding for reducing the risk of pneumonia morbidity and mortality in children under two: A systematic literature review and meta-analysis. *BMC Public Health, 13*(Suppl. 3), S18. https://doi.org/10.1186/1471-2458-13-S3-S18

Leblanc, M., & Crowley, J. (1993). Survival trees by goodness of split. *Journal of the American Statistical Association, 88,* 457–467.

Marmot, M. (1999). Epidemiology of socioeconomic status and health: Are determinants within countries the same as between countries? *Annals of the New York Academy of Sciences, 896,* 16–29.

McGuire, J., & Popkin, B. M. (1990). Beating the zero-sum game: Women and nutrition in the third world. Part 2. *Food and Nutrition Bulletin, 12*(1), 1–9.

Medley, N., Vogel, J. P., Care, A., & Alfirevic, Z. (2018). Interventions during pregnancy to prevent preterm birth: An overview of Cochrane systematic reviews. *Cochrane Database of Systematic Reviews, 2018,* CD012505. https://doi.org/10.1002/14651858.CD012505.pub2

Mogensen, U. B., Ishwaran, H., & Gerds, T. A. (2012). Evaluating random forests for survival analysis using prediction error curves. *Journal of Statistical Software, 50*(11), 1–23.

Molitoris, J., Barclay, K., & Kolk, M. (2019). When and where birth spacing matters for child survival: An international comparison using the DHS. *Demography, 56,* 1349–1370.

Mondal, M. N. I., Hossain, M. K., & Ali, M. K. (2009). Factors influencing infant and child mortality: A case study of Rajshahi District, Bangladesh. *Journal of Human Ecology, 26,* 31–39.

Mullainathan, S., & Spiess, J. (2017). Machine learning: An applied econometric approach. *Journal of Economic Perspectives, 31*(2), 87–106.

Oaxaca, R. L. (1973). Male-female wage differentials in urban labor markets. *International Economic Review, 14,* 693–709.

Pathak, P. K., Singh, A., & Subramanian, S. V. (2010). Economic inequalities in maternal health care: Prenatal care and skilled birth attendance in India, 1992–2006. *PLoS One, 5,* e13593. https://doi.org/10.1371/journal.pone.0013593

Ruhm, C. J. (2004). Parental employment and child cognitive development. *Journal of Human Resources, 39,* 155–192.

Sandall, J., Soltani, H., Gates, S., Shennan, A., & Devane, D. (2016). Midwife–led continuity models versus other models of care for childbearing women. *Cochrane Database of Systematic Reviews, 2016,* CD004667. https://doi.org/10.1002/14651858.CD004667.pub5

Sartorius, B. K., & Sartorius, K. (2014). Global infant mortality trends and attributable determinants—An ecological study using data from 192 countries for the period 1990–2011. *Population Health Metrics, 12,* 29. https://doi.org/10.1186/s12963-014-0029-6

Segal, M. R. (1988). Regression trees for censored data. *Biometrics, 44,* 35–47.

Troeger, C., Blacker, B., Khalil, I. A., Rao, P. C., Cao, S., Zimsen, S. R. M., . . . Reiner, R. C., Jr. (2018). Estimates of the global, regional, and national morbidity, mortality, and aetiologies of lower respiratory infections in 195 countries, 1990–2016: A systematic analysis for the Global Burden of Disease Study 2016. *Lancet Infectious Diseases, 18,* 1191–1210.

UNICEF. (2013). *Improving child nutrition: The achievable imperative for global progress*. New York, NY: UNICEF.

UNICEF. (2017). *Levels and trends in child malnutrition: UNICEF / WHO / World Bank Group Joint Child Malnutrition Estimates*. New York, NY: UNICEF.

UNICEF. (2018). *UNICEF data: Monitoring the situation of children and women*. Retrieved from https://data.unicef.org/topic/child-survival/under-five-mortality/

United Nations. (2015). *Transforming our world: The 2030 agenda for sustainable development*. New York, NY: United Nations.

United Nations Development Program (UNDP). (2015). *The Millennium Development Goals report 2015*. Retrieved from https://www.undp.org/content/undp/en/home/librarypage/mdg/the-millennium-development-goals-report-2015.html

Varian, H. R. (2014). Big data: New tricks for econometrics. *Journal of Economic Perspectives, 28*(2), 3–27.

Westley, S. B. (2003). *Child survival and healthcare in developing countries of Asia* (Asia-Pacific Population & Policy, Report No. 67). Honolulu, HI: East-West Center, Population and Health Studies.
World Bank. (2019). *World Bank development indicators* [Data set]. Retrieved from https://data.worldbank.org/indicator/sp.dyn.imrt.in
Yosefian, I., Mosa Farkhani, E., & Baneshi, M. R. (2015). Application of random forest survival models to increase generalizability of decision trees: A case study in acute myocardial infarction. *Computational and Mathematical Methods in Medicine, 2015,* 576413. https://doi.org/10.1155/2015/576413

Toshiaki Aizawa
toshiaki.aizawa@aoni.waseda.jp

Waseda Institute for Advanced Study, Waseda University, Tokyo, Japan