J Popul Econ (2011) 24:601–628
DOI 10.1007/s00148-009-0290-3

ORIGINAL PAPER

Child mortality in rural India

Bas van der Klaauw · Limin Wang

Received: 12 February 2007 / Accepted: 16 October 2009 /
Published online: 2 December 2009

Abstract This paper focuses on infant and child mortality in rural areas of India. We construct a flexible duration model, which allows for frailty at multiple levels and interactions between the child's age and individual, socioeconomic, and environmental characteristics. The model is estimated using the Indian National Family and Health Survey 1998/1999. The estimation results show that socioeconomic and environmental characteristics have significantly different impacts on mortality rates at different ages. These are particularly important immediately after birth. The parameter estimates indicate that child mortality can be reduced substantially, particularly by improving the education of women, providing safe water, and reducing indoor air pollution caused by dirty cooking fuels. Finally, we still found substantial differences

Responsible editor: Junsen Zhang

B. van der Klaauw (✉)
Department of Economics, VU University Amsterdam, De Boelelaan 1105, 1081 HV Amsterdam, The Netherlands
e-mail: bklaauw@feweb.vu.nl

L. Wang
The World Bank, 1818 H Street, NW, Washington DC 20433, USA
e-mail: lwang1@worldbank.org

B. van der Klaauw
Tinbergen Institute, Amsterdam, The Netherlands

B. van der Klaauw
CEPR, London, UK

in mortality rates between states, which are associated with differences in schooling expenditures, female immunization, and poverty rates.

Keywords Duration analysis · Mortality rates · Heterogeneity

JEL Classification C41 · J13 · J18

1 Introduction

In 1999, about 2.1 million children under age five died in India, which is the highest number within a single country. Even though India experienced in the 1990s a period of relatively high economic growth, the decline in infant and child mortality rates slowed down (see Claeson et al. 1999). As an explanation for the decline in infant and child mortality rates after an increase in income, Pritchett and Summers (1996) mention that an increase in income is usually accompanied by increased public and private expenditures in health inputs. However, in India, both state and central governments suffered from severe financial pressure due to a focus on reforms in industrial and trade policies (e.g., Ahluwalia 2002). In particular, rural areas where 75% of the Indian population live saw declined growth rates in public investments in drainage, water, electricity, irrigation, infrastructure, education, and health care. Arulampalam (2007) argues that rural areas of India lag behind many developing countries in terms of health, social, and economic development.

Hobcraft et al. (1985) indicate that, at different ages, there are differential associations between mortality rates and socioeconomic characteristics. Fikree et al. (2002) show that the causes of neonatal deaths are quite different from the causes of post-neonatal deaths. Most empirical studies assume, however, that health inputs have constant impacts on child mortality over the age of the child (e.g., Guo and Rodriguez 1992; Maitra 2004; Ridder and Tunalı 1999; Sastry 1997). We contribute to the existing literature by constructing a very flexible model that allows the impacts of health inputs to change as the child gets older. We show that a hazard rate model can very easily deal with time-varying covariate effects. More precisely, we allow socioeconomic and environmental characteristics to have different effects on neonatal (first month), infant (12 months), and child (60 months) mortality, which are the most often studied health outcomes. For example, Bhalotra and Van Soest (2008) consider neonatal mortality, while Arulampalam (2007), Arulampalam and Bhalotra (2006), and Bhargava (2003) focus on infant mortality, and Makepeace and Pal (2008) investigate child mortality. Our paper thus contributes to understanding why results may differ between studies.

It is well known that, even after controlling for family-specific characteristics, there often remains unobserved frailty at the family level, which is important in explaining child mortality. Not properly taking account of these unobserved characteristics or the relation between children within a family

may lead to inconsistent and inefficient estimators (i.e., Guo and Rodriguez 1992; Ridder and Tunalı 1999; Sastry 1997; Vaupel et al. 1979). Like, for example, Arulampalam and Bhalotra (2006) and Bhalotra and Van Soest (2008), we allow for random family-specific effects (i.e., similar for all children from the same mother). However, we follow Sastry (1997) and allow for frailty on more levels. In particular, we also allow for frailty at the level of the child. To distinguish these types of frailty, we exploit that children within the same family are related, which implies that they might share some unobserved characteristics. Unlike Sastry (1997), we do not restrict the different types of frailty to be independent of each other. It should, however, be noted that identification requires unobserved components to be age-invariant.

The key objective of our model is to identify the relevant mortality determinants at different ages of a child, which is important for designing effective public policies for reducing child mortality. We are particularly interested in socioeconomic and environmental characteristics, such as mother's education, source of drinking water, sanitation facility, type of cooking fuels, access to electricity, and availability of medical services. In low-income countries, these health inputs absorb a large share of the household expenditure (Pritchett and Summers 1996). Our empirical model can be used to identify households with high child mortality risks, which can be useful for targeting resources and policy towards such households. We also provide some indication of the health benefits of possible public programs targeted at improving households' health-related resources.

We estimate the model using data from the National Family Health Survey 1998/1999 of India.[1] This is a rich, nationally representative database that covers around 90,000 households living in India. The Indian National Family Health Survey is used by, for example, Arulampalam and Bhalotra (2006), Bhalotra and Van Soest (2008), Bhargava (2003), Makepeace and Pal (2008), Maitra (2004), and Withworth and Stephenson (2002) to investigate various aspects of infant and child mortality in India. We focus on the rural areas.

We find that a formal statistical test rejects the standard hazard rate model with constant impacts of socioeconomic and environmental characteristics over ages against our model, which allows for differential age impacts. However, when comparing policy simulations, both models have similar results. Both models show that child mortality at all ages can be reduced significantly by improving the mother's education. The effect of mother's education increases, however, substantially after the first month. Maitra (2004) shows that mother's education has a strong effect on demand for health care. Also, safe water and reducing indoor air pollution by using clean cooking fuels or having a separate kitchen can substantially reduce child mortality. We also find that, after controlling for socioeconomic differences between households, there are

[1] We only focus on a single wave of the survey. The survey does not have a panel structure, so additionally using the very recently made available 2004/2005 wave does not help in controlling for unobserved heterogeneity.

still significant differences in mortality rates between states. Arulampalam (2007) argues that there is clustering of mortality rates within states. Pritchett and Summers (1996) mention the importance of public expenditures on health inputs for reducing infant and child mortality. Since a substantial share of public expenditures are determined at the state level, we provide a decomposition of the state-level fixed effects. The decomposition shows that there are observed state-level characteristics, which are important in explaining the state-level fixed effects. We find that child mortality is negatively associated to schooling expenditures and female immunization rates. This provides guidance that state-level public policies are important.

The paper is organized as follows. Section 2 presents the statistical model. Section 3 describes the data. Estimation results and sensitivity analyses are presented in Section 4. Section 5 concludes.

2 Model

We focus on children that are born alive, and model their mortality probabilities until reaching age five. We extend the commonly used hazard rate models by allowing covariates to have different impacts on mortality at different ages of the child. The model has a random-effects specification. In particular, we allow for two types of frailty, at the level of the family and at the level of the child. As covariates in our model, we use variables that Mosley and Chen (1984) refer to as socioeconomic determinants. Within their framework, socioeconomic determinants affect proximate determinants such as maternal factors, environmental contamination, nutrient deficiency, injuries, and disease control. These proximate determinants are the intermediate variables between the socioeconomic determinants and the sickness and mortality risk. The socioeconomic determinants can be considered as predetermined with respect to a child's survival (see Ridder and Tunalı 1999; for an extensive discussion on this issue). Making such an assumption with respect to proximate determinants like breastfeeding or vaccination is much more complicated. The decision to breastfeed can be made after the child is born and unobserved child-specific characteristics are observed. For vaccinations, the problem might even be more problematic as some children might already have died prior to the moment of vaccinating. Our model estimates the effect of socioeconomic determinants such as parental education, but does not provide information on possible pathways via proximate determinants. So, for example, the effect of mother's education might include that higher-educated mothers are more aware of the importance of breastfeeding and vaccinating, which again might affect child mortality rates. Our empirical analysis should thus be considered as a reduced-form analysis.

We observe J families, which are denoted by $j = 1, \ldots, J$. As family, we consider all children born of the same mother. Family j has some specific characteristics that are described by a vector z_j, which includes, for example, religion of the household, wealth, parental education, and availability of

sanitation. A parameter s_j describes the fixed effect of the state in which the family is living. Obviously, two families living in the same state have the same state-fixed effect. Furthermore, we allow for additional heterogeneity v_j that describes unobserved family-specific characteristics. Both observed and unobserved family-specific characteristics do not vary over time. During the observation period, I_j children were born in family j. We allow children in the same family to differ in both observed and unobserved characteristics. Observed child-specific characteristics, such as gender and birth order, are captured in a vector x_{ij}. Additionally, there is a component w_{ij} that accounts for child-specific frailty. We assume that all covariates are exogenous, i.e., the joint distribution of v_j and w_{ij} does not depend on z_j, s_j, and x_{ij}.

For a child, we can distinguish two possible observations, (1) the child is observed to die during the observation period before reaching age five and (2) the child reached its fifth birthday alive or is still alive at the end of the observation period. In the first case, we observe that child i in family j died in some age interval $(\underline{t}_{ij}, \bar{t}_{ij})$. In the second case, $\underline{t}_{ij}$ equals 60 months or the age of the child in months at the end of the observation period if the child did not reach its fifth birthday (and $\bar{t}_{ij} = \infty$). We introduce a dummy variable d_{ij} that takes the value 1 if the child died within the observation period (the first case) and the value 0 otherwise. All observations are thus artificially right-censored at age five.

We define T_{ij} as the continuous random variable (with 1 month as the unit of time) that describes the age at which child i in family j dies. A common way to model these types of random variables is to specify the hazard rates $\theta_{ij}(t)$, which describes the intensity at which a child dies at age t given that the child survived until this age. The hazard rate $\theta_{ij}(t)$ of child i in family j at age t is specified as

$$\begin{aligned}\theta_{ij}(t|z_j, x_{ij}, v_j, w_{ij}) = \lambda(t) \exp\big\{ & I(t \le 1)(z_j\gamma_1 + x_{ij}\beta_1) + I(1 < t \le 12)(z_j\gamma_2 + x_{ij}\beta_2) \\ & + I(12 < t \le 60)(z_j\gamma_3 + x_{ij}\beta_3) + s_j \big\}\, v_j w_{ij} \end{aligned} \quad (1)$$

The hazard rate can be decomposed into three parts. The baseline hazard $\lambda(t)$ captures age dependence and is similar for all children. The second part is a regression function, where $I(\cdot)$ is the indicator function taking the value 1 if its argument is true and 0 otherwise. Both family-specific covariates z_j and child-specific covariates x_{ij} are allowed to have different impacts on the hazard rate at three different age intervals, (1) during the first month after birth, (2) from the second month until the first birthday, and (3) after the first year until reaching age five. The main motivation for taking these three age intervals is that it describes neonatal (first month), infant (first year), and child (under age five) mortality rates, which are the most often studied measures for infant and child mortality. Our specification implies that we are very flexible on all three outcome measures within the same model. The impact of the state fixed effect s_j does not vary over the child's age, although this could easily be relaxed. The third part $v_j w_{ij}$ accounts for the effects of unobserved heterogeneity.

So far, no empirical study of child mortality allowed the effects of socio-economic and environmental covariates to vary over the child's age. Guo and Rodriguez (1992) mention that covariate effects might be dependent on the age of the child, but they do not explicitly model it. The identification of the model is straightforward. The mortality rates $\theta_{ij}(t)$ satisfy the mixed proportionality assumption, and the time varying regressors $I(\cdot)z_j$ and $I(\cdot)x_{ij}$ change values exogenously. The frailty components are assumed to be mean one random effects, for which the joint distribution is independent of the regressors. The identification of the family-specific frailty comes from observing mothers that have multiple children, which share the same frailty component. The identification of the distribution of child-specific frailty requires some structure on the model and, in particular, the assumption of having a mixed proportional mortality rate. These conditions ensure that the model is identified (e.g., Van den Berg 2001).

The probability that a child survives until age t is given by the survivor function

$$S(t|z_j, x_{ij}, v_j, w_{ij}) = \Pr(T_{ij} > t|z_j, x_{ij}, v_j, w_{ij}) = \exp\left(-\int_0^t \theta_{ij}(s|z_j, x_{ij}, v_j, w_{ij})ds\right) \tag{2}$$

To estimate the parameters, we use maximum likelihood. The loglikelihood function equals

$$\log \mathcal{L} = \sum_{j=1}^{J} \log \left\{ \int_v \left(\prod_{i=1}^{I_j} \int_w \left(S\left(\underline{t}_{ij}|z_j, x_{ij}, v, w\right) - d_{ij} S\left(\bar{t}_{ij}|z_j, x_{ij}, v, w\right) \right) dG(w|v) \right) dG(v) \right\}, \tag{3}$$

where $G(v)$ is the marginal distribution function of the family-specific frailty term and $G(w|v)$ is the conditional distribution of child-specific frailty given the family frailty.

The baseline hazard $\lambda(t)$ describes how the mortality rate changes with the age of the child. This is parameterized as a piecewise constant function

$$\lambda(t) = \begin{cases} \exp(\lambda_1) & 0 < t \leq 1 \\ \exp(\lambda_{2-6}) & 1 < t \leq 6 \\ \exp(\lambda_{7-12}) & 6 < t \leq 12 \\ \exp(\lambda_{13-24}) & 12 < t \leq 24 \\ \exp(\lambda_{25-36}) & 24 < t \leq 36 \\ \exp(\lambda_{37-48}) & 36 < t \leq 48 \\ \exp(\lambda_{49-60}) & 48 < t \leq 60 \end{cases} \tag{4}$$

As will be discussed in Section 3, the older the child was when it died, the less precise the recorded age of death is. This specification of the duration dependence pattern follows this feature of the data.

Both family-specific and child-specific frailty are modeled using distributions with discrete mass-points, which are flexible and attractive from a computational point of view. Such distributions allow easily for dependence between frailty components. Sastry (1997) uses two frailty terms, but restricts these to be independent and to follow gamma distributions.

The family-specific component has the distribution

$$\Pr(V_j = v^k) = \frac{\exp(p_k)}{\sum_{k=1}^{K} \exp(p_k)} \qquad k = 1, \ldots, K \tag{5}$$

We normalize the probabilities by setting $p_K = 0$. The mass-point locations should all be positive, $v^k > 0$, and V_j should have mean one (a normalization necessary since the location of $\lambda(t)$ is unrestricted). Therefore, we specify

$$v^k = \exp(\mu_k) \qquad k = 1, \ldots, K-1, \tag{6}$$

and

$$v^K = 1 + \sum_{k=1}^{K-1} \exp(p_k)(1 - \exp(\mu_k)) \tag{7}$$

The child-specific frailty can be interrelated with the family-specific frailty. We specify

$$\Pr(W_{ij} = w^{lk} | V_j = v^k) = \frac{\exp(q_{lk})}{1 - \exp(q_{lk})} \qquad l = 1, \ldots, L; k = 1, \ldots, K \tag{8}$$

Again, we normalize the probabilities by setting $q_{kL} = 0$ for all $k = 1, \ldots, K$. To ensure that all mass-point locations are positive and the conditional random variables $W_{ij}|V_j = v^k$ have mean one, we impose the following specification:

$$w^{lk} = \exp(\eta_{lk}) \qquad l = 1, \ldots, L-1; k = 1, \ldots, K \tag{9}$$

and

$$w^{Lk} = 1 + \sum_{l=1}^{L-1} \exp(q_{lk})(1 - \exp(\eta_{lk})) \qquad k = 1, \ldots, K \tag{10}$$

It should be noted that, for different family-specific frailty terms, the child-specific frailty terms can take different values, and also the weights on these values can differ. The joint distribution of the frailty components is thus described by the sets of parameters p_k, μ_k, q_{lk}, and η_{lk}, which are estimated along with the other parameters when optimizing the loglikelihood function.

3 Data

In the empirical analyses, we use data from the second National Family and Health Survey (NFHS), undertaken in 1998 and 1999. The survey consists of three separate questionnaires, a village questionnaire, a household questionnaire, and a woman questionnaire covering all ever-married women in a

household between age 15 and age 49. The response rate of the interviews is over 95%.

We are interested in child mortality in the rural areas. Therefore, we consider all 61,800 households living in rural areas, which include 62,248 ever-married women between 15 and 49 years old who are interviewed. In the rural areas, over 95% of the women between age 25 and 30 have been married. As we are interested in recent birth histories, we restrict the sample to women who gave birth to at least one live born child since January 1993. We only consider singletons, and, therefore, drop all multiple births. This reduces the sample to 32,069 women, who, in total, gave birth to 53,201 children born alive since January 1993. The data show that, in rural areas, around 92.6% of the pregnancies result in a live birth.

At the moment of the survey, all qualified women were asked about their birth history. For each live born child, the month and year of birth are recorded, and whether or not the child is still alive at the moment of the interview. If a child died during the observation period, the age at which the child died is asked. The age of death is observed within intervals, in case a child died within a month after birth, the age of death is recorded in days, if the child died between 1 month and 2 years, it is recorded in months, and otherwise, it is recorded in years. Because we are only interested in child mortality until age five, we artificially right-censor at this age. Right-censoring can also occur if a child is alive at the moment of the interview and younger than 5 years old.

The data are recorded retrospectively and can, therefore, suffer from misreporting; for example, a child who died at a very young age might not be reported. One may argue that this problem of misreporting becomes larger when the time interval between the moment of the interview and the date at which a child was born or died increases. If this is actually the case, we can get some indication of retrospective misreporting by comparing child mortality rates computed from the NFHS 1992/1993, with those computed from the

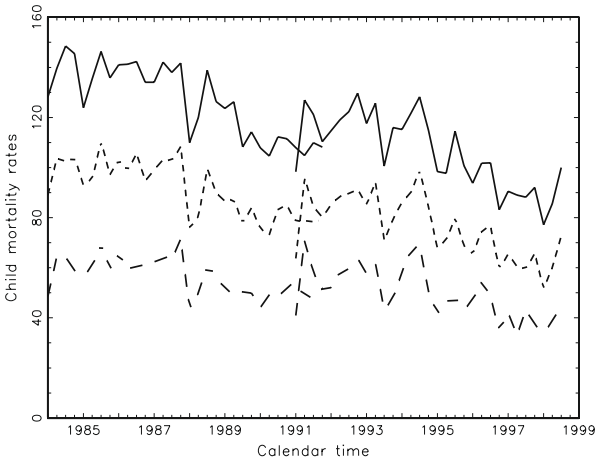

Fig. 1 Rural child mortality rates over calendar time. Explanatory note: The *lower line* is the mortality rate in the first month after birth, the *middle line* the mortality rate during the first year after birth, and the *upper line* the mortality rate during the first 5 years after birth. The *left-hand side of the figure* is based on the NFHS 1992/1993, the *right-hand side* on the NFHS 1998/1999. For 1991 measures, using both surveys are available

NFHS 1998/1999. This is shown in Fig. 1. In case there would be substantial retrospective misreporting, the estimated mortality rates (in the overlapping period) based on the 1998/1999 survey should be lower than those based on the 1992/1993 survey. As can be seen, mortality rates based on both waves are roughly similar. Furthermore, it can be seen that infant and child mortality rates dropped somewhat during the 1990s.

Table 1 provides, for each state, the number of women and the number of children in the sample, and neonatal (first month), infant (first year), and child (first 5 years) mortality rates. In rural India, out of every 1,000 live-born children, almost 100 die before reaching age five, and almost half of them die within the first month after birth. The table shows a large variation in child mortality rates across the different states. The ranking of the states by child mortality rates largely coincides with this ranking in 1960 (see Ravallion and Datt 2002).

Table 1 Cumulative child mortality rates stratified by state (for rural areas)

Region	Women	Children	Mortality		
			≤ 1	≤ 12	≤ 60
India	32,069	53,201	44.4	70.1	99.0
North					
Delhi	92	162	30.7	44.2	44.2
Haryana	1,034	1,691	33.7	57.9	76.4
Himachal Pradesh	937	1,483	22.2	33.7	43.2
Jammu & Kasmir	1,127	1,921	35.9	59.3	73.9
Punjab	812	1,332	35.5	61.7	80.5
Rajasthan	2,960	5,159	49.2	79.4	115.8
Central					
Madhya Pradesh	2,820	4,896	57.4	92.0	148.6
Uttar Pradesh	4,432	7,753	53.2	87.0	124.5
East					
Bihar	3,575	5,823	46.3	70.4	102.6
Orissa	1,726	2,724	44.2	76.5	99.8
West Bengal	1,143	1,704	34.1	50.2	69.1
Northeast					
Aranachal Pradesh	616	1,037	36.4	57.9	93.4
Assam	1,391	2,228	38.2	64.5	86.1
Manipur	592	1,039	19.1	38.3	59.5
Meghalaya	580	1,147	52.6	88.0	121.3
Mizoram	294	536	20.8	46.2	65.8
Nagaland	424	853	23.9	47.1	73.6
Sikkim	540	824	28.8	46.8	72.5
Tripura	510	780	35.8	52.8	61.9
West					
Goa	301	410	24.2	31.8	36.0
Gujarat	960	1,637	42.8	72.6	84.5
Maharashtra	968	1,641	38.4	52.3	67.6
South					
Andhra Pradesh	1,181	1,737	42.7	68.4	89.7
Karnataka	1,231	2,053	36.2	54.7	74.5
Kerala	776	1,073	12.0	16.1	17.3
Tamil Nadu	1,047	1,558	37.4	50.7	74.2

Explanatory note: "Women" provides the number of women in the sample with at least one child born since January 1993. "Children" is the total number of live born children in the observation period. The mortality rates are per 1,000 live born children, the number of children that die within 1 month, within 1 year, and before reaching age five

Table 2 Cumulative child mortality rates stratified by child specific characteristics (for rural areas)

	Share	Mortality		
		≤ 1	≤ 12	≤ 60
Child characteristics				
Girl	0.48	42.5	70.0	105.2
Boy	0.52	46.1	70.1	93.2
Age mother at birth −19 years	0.24	57.9	88.3	113.5
Age mother at birth 20–25 years	0.44	39.2	61.4	88.6
Age mother at birth 26–31 years	0.22	35.2	60.6	90.5
Age mother at birth 32+ years	0.10	50.4	79.8	127.0
Preceding birth interval < 2 years	0.22	66.0	104.6	152.8
Preceding birth interval 2+ years	0.51	30.6	50.8	76.0
First born and girl	0.13	45.1	70.6	89.6
First born and boy	0.14	60.9	86.2	101.7

Explanatory note: "Share" is the fraction of children that has a particular characteristic. "Mortality" gives the mortality rates per 1,000 live born children within the specified age (in months)

We can distinguish three types of explanatory variables, child-specific covariates, household-specific covariates, and village-specific covariates. Table 2 gives the child-specific characteristics, which shows that slightly more boys were born than girls and that, during the first month, boys have a higher probability of dying than girls. After 1 month, this changes and mortality rates of girls become higher, which leads to higher under-5-years child mortality rates among girls than boys. Most children are born when the mother is between 20 and 25 years old. Child mortality rates are substantially higher for young mothers (below 20 years at birth) and for old mothers (above 31 years at birth). Finally, child mortality rates decrease with the length of the preceding birth interval.[2]

Table 3 provides information on household-specific characteristics. Among Hindu households, child mortality rates are much higher than among households with other religions. Approximately one-third of the households belong to a scheduled caste or tribe, and one-third to other backwards classes.[3] Child mortality in scheduled castes and tribes is higher than in other backward classes, where it is, in turn, higher than among households who do not belong to any backward caste. To some extent, religion and castes are spatially correlated.[4]

India is well known for its high female illiteracy (e.g. Drèze and Sen 1995). In our sample, less than one-third of the mothers finished primary school, while

[2]In the empirical analyses, preceding birth interval will be treated as a continuous variable instead of a dummy variable.

[3]Before the 1980s, only the scheduled castes and tribes were included in the minorities that have been designed by the government of India as socially and educationally backward and in need of protection from social injustice. During the 1980s, the other backward classes were also included in these minorities. However, individuals from other backward classes have, on average, better social indicators than individuals from scheduled castes and tribes.

[4]The NFHS reports only the religion and caste and tribe of the head of the household.

Table 3 Cumulative child mortality rates stratified by household specific characteristic (for rural areas)

	Share	Mortality		
		≤ 1	≤ 12	≤ 60
Family background characteristics				
Hindu	0.77	46.1	72.8	102.3
Muslim	0.13	37.4	59.2	86.5
Other religion	0.10	35.4	54.8	79.6
Scheduled caste/tribe	0.36	48.8	78.4	117.4
Other backward class	0.29	46.7	73.3	98.3
No backward class	0.35	37.7	58.6	81.3
Mother and husband				
Mother finished primary school	0.31	30.1	43.3	55.2
Mother did not finish primary school	0.69	49.5	79.7	114.3
Mother separated/widowed	0.02	69.2	106.0	132.6
Married but husband staying elsewhere	0.06	35.5	57.4	80.1
Husband living in household	0.92	44.5	70.2	99.6
Husband finished primary school	0.55	39.0	59.9	81.4
Husband did not finish primary school	0.45	50.4	81.4	118.1
Household wealth				
First decimile	0.09	51.8	88.3	130.2
Second decimile	0.10	49.1	83.3	127.6
Third decimile	0.10	55.9	86.5	128.8
Fourth decimile	0.09	44.8	73.8	103.1
Fifth decimile	0.10	47.7	76.5	109.0
Sixth decimile	0.10	42.8	67.4	96.6
Seventh decimile	0.10	43.3	60.6	82.5
Eighth decimile	0.10	37.0	58.0	74.0
Ninth decimile	0.11	36.5	53.2	64.4
Tenth decimile	0.11	27.8	39.5	46.5
House characteristics				
Pucca	0.18	35.0	52.7	65.0
Semi-Pucca	0.41	44.4	69.6	101.2
Kachha	0.41	46.1	73.3	105.2
Kitchen and cooking fuels				
Separate room as kitchen and clean cooking fuels	0.06	25.1	36.4	46.8
Separate room as kitchen and no clean cooking fuels	0.42	41.4	62.2	80.8
No separate room as kitchen and clean cooking fuels	0.03	38.8	56.6	72.1
No separate room as kitchen and no clean cooking fuels	0.49	48.3	78.9	116.7
Toilet facility				
Flush toilet/pit toilet/latrine	0.24	29.3	44.7	57.2
No facility	0.76	47.3	75.0	107.1
Electricity supply				
Irregular supply	0.27	41.4	63.5	86.4
Regular supply	0.22	35.0	54.6	69.2
No supply	0.51	50.9	81.5	118.0
Water source				
Piped water (private)	0.11	38.0	52.0	69.6
Piped water (public)	0.14	36.8	60.1	84.7
Handpump water (private)	0.18	45.4	69.3	95.3
Handpump water (public)	0.25	46.7	75.4	109.2
Well water	0.24	47.4	75.4	105.8
Other	0.08	39.3	68.4	105.1
Purifies piped water	0.10	43.6	65.3	87.0
Does not purify piped water	0.15	33.6	52.3	74.2
Purifies handpump water	0.06	47.2	68.8	93.2
Does not purify handpump water	0.37	46.0	73.3	104.5

Table 3 (continued)

	Share	Mortality		
		≤ 1	≤ 12	≤ 60
Purifies well water	0.08	38.3	61.6	83.4
Does not purify well water	0.16	51.5	81.7	115.8
Purifies other water source	0.03	33.2	66.5	95.2
Does not purify other water source	0.04	43.0	69.6	111.2
Time to get water				
≤ 5 min	0.52	42.8	64.9	90.2
5–10 min	0.16	46.0	76.2	106.4
10–20 min	0.14	46.3	75.4	109.3
20–30 min	0.11	44.7	72.1	105.7
> 30 min	0.06	48.7	83.5	121.5

Explanatory note: Share is the fraction of families that has a particular characteristic. Mortality gives the mortality rates per 1,000 live born children within the specified age (in months)

this is the case for slightly over half of the husbands. Child mortality rates are much higher in families with uneducated parents, which is particularly pronounced for mother's education. Education is strongly correlated with the type of work. Education can, therefore, be considered as a measure for earnings capacity. A disadvantage of the NFHS is that it does not provide any information on household income or consumption expenditures. However, the NFHS is very rich on asset ownership, such as car, radio, television, refrigerator, etc., and the NFHS provides information on livestock and land ownership. In low-income countries, where household income is often difficult to measure (particularly in rural areas), consumption expenditures are often used in determining poverty (e.g., Deaton 1997). Although asset ownership is less sensitive to short-term fluctuations than consumption expenditures, asset ownership and consumption expenditures are strongly correlated. Filmer and Pritchett (2001) use asset ownership variables to construct a wealth index. Note that, for example, access to water and sanitation are included as assets in the wealth index. The ranking of households according to this index is found to largely coincide with a ranking based on standard welfare measures such as consumption or earnings. Therefore, we compute for each household the wealth index, and construct dummy variables for the decimiles of the distribution of the index in our sample.[5] Infant and child mortality is negatively correlated with the household wealth index.

Housing characteristics are important in explaining differences in child mortality rates. Child mortality among households living in houses built of high-quality material (Pucca) is lower than among households living in houses of low-quality material (Kachha). Child mortality is lower when a separate room is used for cooking, and when the house has electricity and some type of toilet facility. Regardless of whether the household has a separate kitchen

[5]The cutoff points for the decimiles are based on the distribution of the wealth index for all households living in rural areas.

Table 4 Child mortality rates stratified by village specific characteristic (for rural areas)

	Share	Mortality		
		≤ 1	≤ 12	≤ 60
Village characteristics				
Nearest town −10 km	0.49	42.1	69.0	99.0
Nearest town 10–25 km	0.35	44.9	68.2	94.5
Nearest town 25+ km	0.16	50.9	78.5	110.7
−250 households in village	0.44	45.9	74.1	108.1
250–1,000 households in village	0.42	44.0	68.9	95.4
1,000+ households in village	0.14	41.5	63.2	86.5
Primary school in village	0.89	43.5	69.0	97.5
No primary school in village	0.11	50.7	78.4	110.5
Drainage in village	0.45	43.8	69.7	97.8
No drainage in village	0.55	44.9	70.3	100.0
Health facility −5 km	0.76	42.5	66.4	92.9
Health facility 5–10 km	0.16	50.0	81.2	115.5
Health facility 10+ km	0.08	51.0	82.6	125.1
Hospital −5 km	0.35	40.5	64.9	93.8
Hospital 5–10 km	0.23	45.5	70.9	98.7
Hospital 10+ km	0.42	47.0	74.1	103.9
Doctor available	0.49	42.7	67.2	91.6
No doctor available	0.51	46.2	73.2	107.3

Explanatory note: Share is the fraction of families that lives in a village with a particular characteristic. Mortality gives the mortality rates per 1,000 live born children within the specified age (in months)

for cooking or not, child mortality rates are lower if the household uses clean cooking fuels instead of wood, crop residues, or dung cakes.

Piped water is usually considered as safer than other sources of drinking water. Child mortality rates are indeed substantially lower in households that have access to piped water. Among households with access to piped water, child mortality rates are not lower for households that purify water. For any other source of water, child mortality rates are lower among the households that purify water. Furthermore, child mortality is higher for households that need more than 5 min to get to the water source, but beyond 5 min, it does not increase with the time to get to the water source.

Finally, Table 4 provides information on the village-specific characteristics. The mortality rates are lower in villages that are closer to a town, that are larger in terms of families living in the village, that have a primary school, drainage, and doctor, and that are closer to a health facility or hospital.

4 Empirical results

4.1 Parameter estimates

The parameter estimates are presented in Table 5. Since the model contains many parameters, we will mainly focus on policy-relevant parameters. The model allowed for both child-specific frailty and family-specific frailty. However, we did not find any significant and substantial child-specific frailty, i.e., when optimizing the loglikelihood function, the mass-point locations converged to each other. This implies that we cannot find differences between children born in the same family that are not captured by observed

Table 5 Estimation results

Unobserved heterogeneity						
p_1	3.61**	(0.41)				
μ_1	−0.16**	(0.034)				
State fixed effects						
Delhi	−0.071	(0.42)				
Haryana	0.11	(0.14)				
Himachal Pradesh	−0.24	(0.17)				
Jammu & Kasmir	0.17	(0.14)				
Punjab	0.30*	(0.16)				
Rajasthan	0.23**	(0.11)				
Madhya Pradesh	0.42**	(0.11)				
Uttar Pradesh	0.31**	(0.11)				
Bihar	0.054	(0.11)				
Orissa	0.13	(0.12)				
West Bengal	−0.23	(0.15)				
Arunachal Pradesh	0.0095	(0.17)				
Assam	0.063	(0.13)				
Manipur	−0.22	(0.20)				
Meghalaya	0.27	(0.16)				
Mizoram	−0.24	(0.23)				
Nagaland	−0.16	(0.20)				
Sikkim	−0.058	(0.20)				
Tripura	−0.22	(0.19)				
Goa	−0.41	(0.30)				
Gujarat	0.088	(0.14)				
Maharashtra	−0.15	(0.14)				
Andhra Pradesh	0					
Karnataka	−0.15	(0.13)				
Kerala	−1.08**	(0.27)				
Tamil Nadu	−0.078	(0.14)				
Duration dependence						
λ_1	−2.37**	(0.20)				
λ_{2-6}	−4.67**	(0.25)				
λ_{7-12}	−5.20**	(0.25)				
λ_{13-24}	−5.57**	(0.28)				
λ_{25-36}	−5.93**	(0.28)				
λ_{37-48}	−6.21**	(0.29)				
λ_{49-60}	−6.32**	(0.30)				
	age ≤ 1		$1 <$ age ≤ 12		$12 <$ age ≤ 60	
Child specific characteristics						
Being girl	−0.035	(0.053)	0.23**	(0.067)	0.46**	(0.078)
Age Mother at birth 20–25	−0.19**	(0.057)	−0.23**	(0.080)	−0.16*	(0.093)
Age Mother at birth 26–31	−0.22**	(0.075)	−0.15	(0.097)	−0.057	(0.11)
Age Mother at birth 32+	0.16*	(0.088)	−0.0040	(0.11)	0.32**	(0.13)
Preceding birth interval (in years)	−0.26**	(0.017)	−0.24**	(0.023)	−0.34**	(0.033)
Being first child of mother and boy	−0.18**	(0.079)	−0.48**	(0.11)	−0.91**	(0.16)
Being first child of mother and girl	−0.45**	(0.086)	−0.69**	(0.12)	−1.32**	(0.15)
Household specific characteristics						
Muslim	−0.12	(0.079)	−0.10	(0.10)	−0.011	(0.12)
Other religion	−0.14	(0.11)	0.044	(0.13)	0.017	(0.15)
Scheduled caste/tribe	0.0075	(0.061)	0.028	(0.081)	0.22**	(0.10)
Other backward class	0.10	(0.062)	0.067	(0.081)	−0.025	(0.10)

Table 5 (continued)

	age ≤ 1		1 < age ≤ 12		12 < age ≤ 60	
Mother and husband						
Mother finished at least primary school	−0.21**	(0.064)	−0.42**	(0.091)	−0.47**	(0.13)
Mother separated	0.22	(0.14)	0.13	(0.20)	−0.067	(0.23)
Husband staying elsewhere	−0.19*	(0.11)	−0.11	(0.14)	0.0047	(0.16)
Husband finished at least primary school	−0.15**	(0.050)	−0.13**	(0.064)	−0.23**	(0.078)
Household wealth						
Second decimile	−0.025	(0.092)	−0.025	(0.12)	0.030	(0.13)
Third decimile	0.12	(0.093)	−0.078	(0.12)	0.14	(0.13)
Fourth decimile	−0.079	(0.10)	−0.073	(0.12)	−0.30**	(0.15)
Fifth decimile	−0.011	(0.11)	−0.072	(0.14)	−0.033	(0.15)
Sixth decimile	−0.024	(0.11)	−0.12	(0.14)	−0.18	(0.17)
Seventh decimile	0.021	(0.12)	−0.34**	(0.16)	−0.34*	(0.20)
Eighth decimile	−0.027	(0.13)	−0.24	(0.17)	−0.49**	(0.22)
Ninth decimile	0.026	(0.14)	−0.40**	(0.19)	−0.54**	(0.25)
Tenth decimile	−0.11	(0.17)	−0.51**	(0.23)	−0.90**	(0.37)
House characteristics						
Pucca	0.058	(0.090)	0.040	(0.12)	−0.038	(0.17)
Semi-Pucca	−0.027	(0.059)	−0.046	(0.075)	0.14*	(0.086)
Kitchen and cooking fuels						
No separate kitchen & clean fuel	0.23	(0.15)	−0.27	(0.26)	−0.68*	(0.40)
Separate kitchen & clean fuel	−0.24	(0.15)	−0.25	(0.23)	−0.12	(0.35)
Separate kitchen & dirty fuel	0.019	(0.053)	−0.075	(0.071)	−0.29**	(0.087)
Toilet facility						
Flush toilet / pit toilet / laterine	−0.080	(0.075)	0.019	(0.097)	−0.21*	(0.13)
Electricity supply						
Electricity available	−0.16**	(0.063)	0.0070	(0.080)	−0.076	(0.095)
Water source						
Piped water (private)	0.037	(0.15)	−0.55**	(0.20)	−0.097	(0.22)
Piped water (public) ≤ 10 min	0.098	(0.14)	−0.32*	(0.18)	−0.14	(0.22)
Piped water (public) > 10 min	0.067	(0.14)	0.076	(0.18)	0.076	(0.21)
Handpump water (private)	0.16	(0.13)	−0.20	(0.15)	−0.13	(0.17)
Handpump water (public) ≤ 10 min	0.16	(0.12)	−0.025	(0.15)	−0.014	(0.17)
Handpump water (public) > 10 min	0.15	(0.12)	−0.14	(0.15)	0.025	(0.17)
Well water ≤ 10 min	0.22*	(0.12)	−0.087	(0.15)	−0.12	(0.17)
Well water > 10 min	0.25**	(0.13)	−0.015	(0.15)	0.018	(0.17)
Other water source ≤ 10 min	−0.088	(0.19)	0.0082	(0.21)	0.037	(0.25)
Other water source > 10 min	0		0		0	
Household purifies water						
Piped water	0.24**	(0.11)	0.37**	(0.14)	0.15	(0.17)
Nonpiped water	−0.039	(0.067)	−0.065	(0.086)	−0.083	(0.10)
Village characteristics						
Drainage in village	0.0097	(0.049)	0.12*	(0.062)	0.056	(0.076)
Doctor available	−0.042	(0.048)	0.0049	(0.063)	−0.25**	(0.076)
Health facility 5+ km	0.083	(0.052)	0.011	(0.067)	0.12	(0.080)
Hospital 10+ km	−0.011	(0.049)	0.020	(0.066)	−0.11	(0.076)
Nearest town 10–25 km	0.039	(0.052)	−0.16**	(0.069)	−0.15*	(0.081)
Nearest town 25+ km	0.14**	(0.067)	−0.040	(0.088)	−0.031	(0.10)
< 250 families in village	−0.17**	(0.079)	0.017	(0.11)	−0.14	(0.13)
250–1000 families in village	−0.12*	(0.073)	−0.037	(0.10)	−0.25**	(0.12)
Primary school in village	−0.057	(0.069)	−0.069	(0.092)	0.053	(0.11)

Explanatory note: Standard errors in parentheses
*Significant at 10%
**Significant at 5%

child-specific effects. The parameter estimates thus come from a model with only family-specific frailty. Also, family-specific frailty is not very dispersed, with over 97% of the probability mass being assigned to a mass point that is almost 7.5 times lower than the other mass point. Allowing for a third mass point does not improve the estimation results. It may be important to stress that the child-specific frailty is identified from imposing the mixed proportional hazard structure on the model. Identification of family-specific frailty relies on less strong assumptions because children born in the same family share the family-specific frailty (see Ridder and Tunalı 1999 for a more extensive discussion on this issue).

The state fixed effects differ significantly from each other, with a p value for a Wald-test for joint significance being almost equal to 0. This implies that, after controlling for child-, family-, and village-specific effects, there are still significant differences in child mortality rates between states. This is consistent with the fact that state governments take the major responsibility for social spending, including health, education, nutrition, and social development programs. Therefore, there are, for example, substantial differences in immunization rates across states. As a result, living and environmental conditions vary largely across states (e.g., Datt and Ravallion 2002). We return to this issue below when we decompose the estimated state fixed effects into observed state characteristics.

The baseline hazard declines with age of the child, which implies that mortality rates decrease as a child gets older. The decreases in duration dependence are most substantial for the youngest, and become smaller as the child gets older. However, this duration dependence is not the only source of age dependence because observed child characteristics and the household's socioeconomic and environmental characteristics are allowed to have different impacts on mortality rates at different ages. Also, this source of age dependence in mortality rates is significant; a Wald test rejects the null hypothesis that covariate effects do not change over age (the p value of this test is almost 0). This implies that commonly used specifications in empirical research of child mortality, which impose that observed characteristics have the same effect on the mortality rate at all ages, are not sufficiently flexible to capture all relevant changes in child mortality rates.

We have investigated at which age socioeconomic and environmental characteristics are important. Therefore, we have computed for each child $\exp(z_j\hat{\gamma}_k + x_{ij}\hat{\beta}_k)$ for $k = 1, 2, 3$, and we have used these to compute the standard deviation within the population of children. A large standard deviation of $\exp(z_j\hat{\gamma}_k + x_{ij}\hat{\beta}_k)$ indicates that the covariates are relatively important in explaining mortality rates. The computed standard deviations equal 0.0169, 0.0014, and 0.0007, for $k = 1, 2, 3$, respectively. The standard deviation is, thus, largest in the first month after birth and decreases afterwards, indicating that differences in child mortality rates between children are particularly pronounced in the first month after birth.

The first set of parameter estimates summarized in Table 5 concerns child-specific covariates. Within the first month after birth, boys have higher

Table 6 Simulated marginal effects; lives saved per 1,000 live born children cumulative until a particular age

	One month		One year		Five years	
Piped water (private)[a]	7.2	(4.6)	16.9	(5.7)	15.5	(7.8)
Handpump water (private)[a]	2.3	(3.4)	5.1	(4.4)	5.2	(5.6)
Electricity available	6.4	(2.5)	5.9	(3.3)	7.7	(4.3)
Mother finished primary school	8.1	(2.3)	17.3	(3.0)	27.8	(4.1)
Toilet facility	3.2	(2.9)	2.5	(4.0)	7.9	(5.2)
Separate kitchen & dirty cooking fuel[b]	−0.8	(2.2)	1.1	(2.9)	8.6	(3.6)
Separate kitchen & clean cooking fuel[b]	8.8	(5.0)	14.3	(6.9)	16.9	(12.0)
Doctor in village	1.7	(2.0)	1.5	(2.6)	8.3	(3.4)

Explanatory note: The standard errors in parentheses are computed using the Delta method. Sample weights are used to make the simulation results representative for rural India

[a]The counterfactual is having well water within 10 min (no family purifies water)

[b]The counterfactual is no separate kitchen and using dirty cooking fuels

mortality rates than girls. This difference is large and significant for the first-born child of a mother. After that, the gender of the child does not affect its mortality rates if the child is the first born of a mother. However, among children who are not the first born child, girls are significantly more likely to die than boys. Claeson et al. (1999) suggest that, due to social norms, families have preferences for sons; less money is spent on girls, girls are taken to hospital in a later stage of illness than boys, and girls are taken to less qualified doctors. Reducing such gender discrimination might substantially reduce child mortality rates. Mortality rates are higher among children whose mother was a teenager at the moment of birth, and also among children whose mother was older than 31 at birth. A longer preceding birth interval significantly reduces mortality rates.[6] Withworth and Stephenson (2002) suggest that, after a short birth interval, there is a more intense pressure on the workload of the mother and the household's resources, which increases mortality rates. Arulampalam and Bhalotra (2006) argue that short birth intervals are correlated to clustering of child mortality within a family. Maitra and Pal (2008) endogenize birth intervals, but do not find substantial differences in estimated coefficients, although they find evidence for endogeneity.

Parental educational attainment is also an important determinant of child mortality. Child mortality at all ages is significantly lower when the parents finished primary education. The effect is stronger for the mother than for the father. The effect of parental education is larger for older children, implying that parental education becomes even more important in reducing child mortality at older ages. We present marginal effects on mortality rates in Table 6. The marginal effects are cumulative lives saved until a particular age. So the interpretation of the table is such that, if all mothers would finish primary education, 27.8 under-age-five deaths out of every 1,000 live-born

[6]Interacting the length of the preceding birth interval with gender shows that the covariate effect of the preceding birth interval is similar for boys and for girls.

children could be averted compared to the counterfactual where no mother would finish primary education. To interpret the marginal effect causally, it is crucial to assume that the distribution of unobserved heterogeneity remains unaffected if more mothers finish primary education. This ceteris paribus assumption may be too strong if mother's education is correlated to some unobserved (behavioral) component, and if the size of the correlation changes if the fraction of mothers with primary education increases (e.g., Lee et al. 1997). Our results on mother's education confirm the finding of Strauss and Thomas (1995) that parental education and, in particular, mother's education is important for health outcomes of children. Educated mothers are usually healthier and give birth to healthier babies. They also provide a healthier environment to children, are more likely to have more knowledge about care-taking, and have higher demands for prenatal care (Maitra 2004). Also, Pritchett and Summers (1996) stress that parental education is often mentioned as having a strong effect on reducing infant and child mortality. In India, women have a low social status. Clearly, this has a negative impact on the number of girls that attend school, which leads to high illiteracy levels among women. In comparison with men, women often have never attended school because education was not considered to be necessary or because they were required for household work or taking care of siblings.

Using dirty cooking fuels inside the house causes indoor air pollution. Having a separate kitchen reduces child mortality significantly after the first birthday even if dirty cooking fuels like wood are used. In fact, only 9% of all families use kerosene, which is considered to be a clean cooking fuel. If all families switch to clean cooking fuels in a separate kitchen, almost 17 more children out of 1,000 live-born children survive until age five (see Table 6). It should be noted that switching to clean cooking fuels seems to be particularly relevant during the first year after birth.

Mortality rates after the first birthday are significantly lower within families with some type of toilet facility. This result is in agreement with Bhargava (2003), who finds, using data from Uttar Pradesh, that access to sanitation facilities significantly reduces infant mortality. The marginal reduction in child mortality of having a sanitation facility is about eight children out of 1,000 live-born children. Also, having access to electricity reduces child mortality rates, but this effect is only significant during the first month after birth. This finding coincides with Ridder and Tunali (1999), who find for Malaysia that having access to electricity reduces child mortality. Also, Wang (2003) finds using a cross-country analysis of DHS data that access to electricity is important in reducing child mortality rates.

Most of the covariates describing the source of drinking water do not have a significant effect on mortality rates. The p value for joint significance of the impacts of these covariates is 0.085. Jalan and Ravallion (2003) argue that access to piped water alone is not a sufficient condition for improving a child's health status. Their empirical results based on propensity score matching methods show that access to piped water does not reduce the incidence of diarrhea of children in poor families. Our model simulations show that,

compared to only having access to well water, private piped water reduces child mortality rates substantially. The marginal effect of having handpump water is much smaller. It should, however, be noted that the benefits of both these types of water sources (compared to well water) are mainly during the first year.

To purify water, most households in rural areas strain water by cloth. Water purification decreases mortality rates when the source of drinking water is not piped water, but increases mortality rates in case the household has access to piped drinking water. Even though the latter increase is insignificant, it is counterintuitive as one expects water purification to decrease child mortality rates. One could think of two possible explanations why we find that water purification in combination with piped water increases mortality rates. Most likely, households that purify water have access to lower quality piped drinking water than households that do not purify. Alternatively, it might be that the purification of piped water by the municipality is adequate, and straining water by cloth lowers the quality of the piped drinking water.

Medical facilities do not seem to be very important in reducing mortality rates, except for availability of a village doctor. The presence of a doctor reduces the mortality rates of children after their first birthday. If all villages would have a doctor, slightly over eight under-age-five deaths per 1,000 births would be avoided (compared to no doctor available in any village). Although this number is not very substantial compared to, for example, mother's education, it is significant. Neonatal mortality rates are somewhat higher in villages that are further from the nearest town. Larger villages have slightly higher mortality rates. Recall from the raw statistics that smaller villages have, on average, higher mortality rates. So obviously, the smaller villages, which have, on average, worse socioeconomic and environmental characteristics, do relatively good in reducing child mortality.

4.2 Robustness checks and sensitivity analyses

In this subsection, we investigate the fit of the model, and we perform some sensitivity analyses. The model predictions on child mortality are based on the estimated survivor functions, which we weight with sample weights to make them representative for rural India.[7] Figure 2 shows how the mortality rates predicted by our estimated model coincide with the observed data. It should be noted that the 95%-confidence interval around the model predictions is relatively tight, suggesting that the model predictions are very precise. The mortality rates computed from the actual data almost always fall within the model's confidence interval, indicating that the model is sufficiently flexible to describe the data accurately.

A good fit of the model does not guarantee that the parameter estimates are true marginal effects, i.e., it might be that some covariates account for child,

[7]We use the Delta method to construct standard errors around the predicted mortality rates.

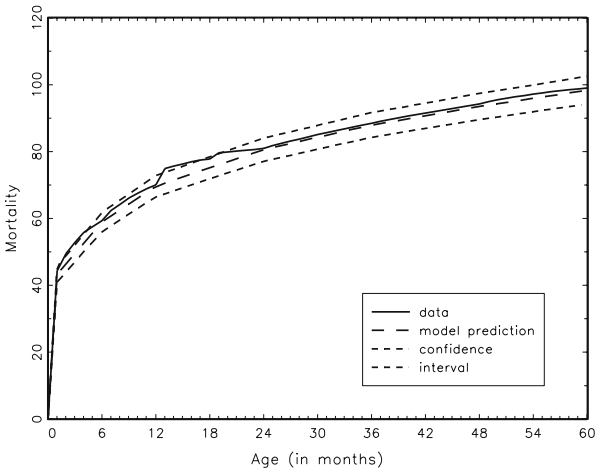

Fig. 2 Size of child mortality before a particular age (out of 1,000 live born). Both the data and the model predictions are weighted to make them nationally representative for the rural areas

family, or environmental characteristics that are not included in the model. Failing to correct for all relevant heterogeneity may lead to serious biases in estimated partial effects. Therefore, we have investigated the sensitivity of the parameter estimates with respect to including additional regressors. Obviously, such sensitivity analyses can never fully guarantee that all relevant heterogeneity is included in the model.

A particularly important assumption is that dummy variables for wealth distribution decimiles are sufficiently informative about a household's financial situation. The parameter estimates show that mortality rates are lower in wealthier households. This is particularly true after the first month. In a sensitivity analysis, we have tried to include polynomials in the wealth index. This hardly affected parameter estimates.

As mentioned earlier, rural India experienced a decline in child mortality rates during the 1990s. Therefore, we have tried to include a polynomial describing calendar time variation. A third-degree polynomial has a p value for joint significant of 0.28. Most parameter estimates are close to zero, and the polynomial does not show much variation. If any trend can be seen, true calendar time effects show even a slightly increasing trend during the observation period. Also, Arulampalam (2007) finds after controlling for socioeconomic and environmental characteristics no trend in mortality rates. One may conclude that the negative trend in rural child mortality rates are thus the result of improved living conditions during the 1990s. For example, the fraction of the population that is either illiterate or did not finish primary eduction is steadily decreasing. In the sample of ever married women, the fraction of women that is illiterate decreased from 72% in the NFHS 1992/1993 to 67% in NFHS 1998/1999. Also, more households have access to electricity, sanitation facilities, and safe water.

We have also estimated the model using both the NFHS 1992/1993 and the NFHS 1998/1999 including dummy variables for the NFHS 1992/1993. The p

value for joint significance of these dummy variables is 0.19. This implies that, after correcting for socioeconomic and environmental characteristics, child mortality rates in the NFHS 1992/1993 were slightly lower during the first month after birth, and somewhat higher after the first month after birth. The dummy variables thus do not all have the same sign.[8]

Next, we tried to include additional observed heterogeneity to the model. The variables that we added to the model did not have any significant effect on mortality rates and did not change covariate effects of other variables. In particular, we tried including additional variables describing type of work performed by the parents, additional variables describing the medical facilities in the area around the village, and village-level variables such as the percentage of households with access to piped water, electricity, and basic sanitation.

The data describe the household situation at the moment of the interview. Recall that India experienced a period of economic growth during the 1990s. However, when estimating our model, we have assumed that household situations have remained stable during the observation period. One might argue that this is a too strong assumption, for example, because asset ownership could change over time. Since the data are not informative on this, we have estimated the model again only using children born after January 1997. Because this substantially reduces the observation period, we only focus on infant mortality. In Table 7, we provide the marginal effects of this model. As a comparison, the table also gives the marginal effects for the full model for the same subsample of the data. The estimated marginal effects are fairly close to each other. In the model estimated, only on recently born children is the effect of safe water on neonatal mortality slightly lower. There are also some differences in the marginal effects of sanitation facilities and having a separate kitchen, but none of these differences are significant. This implies that there is no strong indication for biases in our model due to incorrect measurement of socioeconomic and environmental characteristics over the observation period.

In our empirical analyses, we assumed that fertility is exogenous. It may, however, be that fertility decisions depend on the family's socioeconomic characteristics or on child survival (e.g., Bhargava 2003). Pitt (1997) shows that mother's education is negatively correlated to fertility, and that fertility decisions are important in explaining child mortality. Maitra and Pal (2008) show that there are unobservables that jointly affect fertility decisions and child mortality. Their estimated covariate effects of child mortality are, however, not very sensitive to accounting for endogenous fertility. We investigate this issue further by following the approach of Pitt (1995). He assumes that, in a setting where each women has at least one child, the first-born child is not the result of self-selection. To mimic this approach, we select only first-born children whose mothers were below age 26 at the moment of birth. It

[8] A problem with combing the NFHS 1992/1993 and NFHS 1998/1999 is that some important variables for explaining child mortality, such as whether or not a household belongs to another backward caste, if the household purifies water, or if there is a doctor available in the village, are not reported in NFHS 1992/1993.

Table 7 Simulated marginal effects using only children born after 1997; lives saved per 1000 live born children cumulative until a particular age

	One month		One year	
Separate model for only recently born children				
Piped water (private)[a]	2.6	(8.2)	15.1	(9.9)
Handpump water (private)[a]	0.8	(5.5)	6.3	(8.0)
Electricity available	6.0	(4.3)	5.3	(6.2)
Mother finished primary school	8.4	(3.7)	13.5	(5.2)
Toilet facility	0.8	(5.0)	−0.8	(7.9)
Separate kitchen & dirty cooking fuel[b]	1.7	(3.7)	3.4	(5.1)
Separate kitchen & clean cooking fuel[b]	5.7	(8.7)	10.5	(9.8)
Doctor in village	1.8	(3.4)	0.5	(4.7)
Full model simulated only for recently born children				
Piped water (private)[a]	6.7	(4.2)	15.5	(5.2)
Handpump water (private)[a]	2.1	(3.1)	4.7	(4.1)
Electricity available	6.0	(2.3)	5.5	(3.1)
Mother finished primary school	7.6	(2.2)	16.0	(2.8)
Toilet facility	3.0	(2.7)	2.4	(3.7)
Separate kitchen & dirty cooking fuel[b]	−0.7	(2.0)	1.0	(2.7)
Separate kitchen & clean cooking fuel[b]	8.1	(4.7)	13.2	(6.4)
Doctor in village	1.6	(1.8)	1.4	(2.4)

Explanatory note: The standard errors in parentheses are computed using the Delta method. Sample weights are used to make the simulation results representative for rural India

[a]The counterfactual is having well water within 10 min (no family purifies water)

[b]The counterfactual is no separate kitchen and using dirty cooking fuels

should be noted that, within our data, over 90% of the women had their first child before age 26. In Table 8, we show the marginal effects from the model estimated only on the subsample of first-born children. We compare this to the marginal effects obtained from the full model on the same subsample of the data. The marginal effect of safe water is somewhat lower after the first birthday, and having electricity has a slightly larger marginal effect. However, the main difference is that the marginal effect of having a separate kitchen is much larger. For first-born children, the effect of having a separate kitchen (in the presence of using dirty cooking fuels) is larger than using the sample of all children. All other marginal effects are roughly the same. In particular, the marginal effect of mother's education does not differ substantially. If there would be endogenous fertility, we would have expected it to show up in the effects of mother's education, which is obviously not the case. This result coincides with Pitt (1997). However, since there are some differences in the marginal effects of having a separate kitchen, and the use of dirty cooking fuels, we cannot rule out that fertility decisions are endogenous with respect to child mortality.

Our model extends the usual hazard rate models for child mortality by allowing socioeconomic and environmental characteristics to have different impacts at different ages. Recall from the previous subsection that a formal statistical test rejects that covariate effects are the same at all ages. This implies that the usual mixed proportional hazard rate model is also rejected against our alternative model. It is, however, interesting to see how the predictions

Table 8 Simulated marginal effects with only first born children; lives saved per 1,000 live born children cumulative until a particular age

	One month		One year		Five years	
Model for only first born children						
Piped water (private)[a]	6.2	(8.5)	12.6	(9.2)	7.7	(11.3)
Handpump water (private)[a]	1.0	(6.1)	3.2	(6.6)	2.1	(7.5)
Electricity available	11.3	(4.7)	11.3	(4.8)	6.0	(5.6)
Mother finished primary school	10.8	(4.5)	16.8	(5.2)	27.6	(6.7)
Toilet facility	3.6	(5.6)	2.0	(5.6)	4.1	(6.2)
Separate kitchen & dirty cooking fuel[b]	4.3	(4.1)	10.2	(4.5)	12.2	(5.0)
Separate kitchen & clean cooking fuel[b]	12.5	(6.3)	21.6	(10.5)	23.9	(14.2)
Doctor in village	6.5	(3.6)	4.5	(3.4)	6.2	(4.2)
Full model simulated only for first born children						
Piped water (private)[a]	8.9	(5.7)	18.0	(6.5)	17.0	(7.5)
Handpump water (private)[a]	2.8	(4.2)	5.5	(5.0)	5.5	(5.5)
Electricity available	8.0	(3.2)	7.4	(3.8)	8.6	(4.2)
Mother finished primary school	10.2	(3.1)	19.2	(3.6)	26.6	(4.2)
Toilet facility	4.0	(3.6)	3.3	(4.5)	6.9	(5.1)
Separate kitchen & dirty cooking fuel[b]	−1.0	(2.7)	0.9	(3.2)	6.1	(3.6)
Separate kitchen & clean cooking fuel[b]	10.9	(6.3)	16.1	(7.7)	17.8	(10.2)
Doctor in village	2.1	(2.5)	1.9	(2.9)	6.6	(3.3)

Explanatory note: The standard errors in parentheses are computed using the Delta method. Sample weights are used to make the simulation results representative for rural India

[a]The counterfactual is having well water within 10 min (no family purifies water)

[b]The counterfactual is no separate kitchen and using dirty cooking fuels

of our model differ from the standard mixed proportional hazard. Therefore, in Table 9, we present the marginal effects for a standard mixed proportional hazard rate model. These marginal effects largely coincide with the marginal effects of the extended model presented in Table 6. This implies that even though the standard mixed proportional hazard rate model is rejected, it yields roughly the same policy implications.

Table 9 Simulation marginal effects of a model without age-varying effects of socioeconomic and environmental characteristics; lives saved per 1,000 live born children cumulative until a particular age

	One month		One year		Five years	
No age-varying coefficients						
Piped water (private)[a]	8.3	(3.3)	13.1	(5.2)	18.2	(7.2)
Handpump water (private)[a]	2.5	(2.5)	3.9	(3.9)	5.5	(5.5)
Electricity available	3.8	(1.9)	6.0	(3.0)	8.3	(4.1)
Mother finished primary school	11.4	(1.7)	18.0	(2.8)	25.0	(3.8)
Toilet facility	2.9	(2.3)	4.6	(3.6)	6.4	(5.0)
Separate kitchen & dirty cooking fuel[b]	2.9	(1.6)	4.6	(2.6)	6.3	(3.5)
Separate kitchen & clean cooking fuel[b]	9.4	(4.0)	14.9	(6.4)	20.7	(8.9)
Doctor in village	2.9	(1.5)	4.6	(2.3)	6.4	(3.2)

Explanatory note: The standard errors in parentheses are computed using the Delta method. Sample weights are used to make the simulation results representative for rural India

[a]The counterfactual is having well water within 10 min (no family purifies water)

[b]The counterfactual is no separate kitchen and using dirty cooking fuels

4.3 Decomposition of the state fixed effects

In Subsection 4.1, we showed that, after controlling for the family's socioeconomic and environmental characteristics, the differences in mortality rates between states remain significant. Datt and Ravallion (2002) stress that living conditions vary largely between states. Additionally, Pritchett and Summers (1996) mention the important role of public expenditures on health inputs in reducing infant and child mortality. Furthermore, Arulampalam (2007) mentions that high infant and child mortality is concentrated within a few states. In this subsection, we investigate this issue further by decomposing the state fixed effects in observed state characteristics. Our method for decomposing the state-level fixed effects is very similar to the approach of Donald and Lang (2007).

The state fixed effects are included in a vector s. Let s_x denote a matrix containing observed state characteristics (including an intercept). We decompose the state fixed effect using the regression equation

$$s = s_x\delta + \varepsilon \tag{11}$$

However, the state fixed effects s are not observed. Instead, maximum likelihood provides consistent estimators for these state fixed effects $\hat{s}$, with $\hat{s} \sim \mathcal{N}(s, \Sigma)$, where Σ is the covariance matrix of the maximum likelihood estimators (obviously, we have an estimate $\hat{\Sigma}$).

To estimate the parameters δ, we use feasible generalized least squares (GLS). In the first step we simply perform ordinary least squares on

$$\hat{s} = s_x\delta + e \tag{12}$$

The variance of e equals $\Omega = \Sigma + \sigma^2 I$, where I is the identity matrix. Therefore, we can use $\hat{e}_i^2$ and the diagonal elements of $\hat{\Sigma}$ to estimate σ^2. Once we have an estimate $\hat{\Omega} = \hat{\Sigma} + \hat{\sigma}^2 I$, we can use GLS to estimate the parameters δ.

Parameter estimates are provided in Table 10. We tried a large number of explanatory variables in the matrix s_x. However, because the vector $\hat{s}$ contains only estimated fixed effects for 26 states, there are serious data limitations. Furthermore, many explanatory variables are highly correlated with each other. Therefore, we only include three explanatory variables in s_x. Higher per capita schooling expenditures are associated to lower child mortality rates. Recall that, in the hazard rates, we already included indicators for whether or not the mother and father finished primary education. So the effect of the state schooling expenditures affects child mortality not only via the educational level

Table 10 Estimation results from the regression of the state level fixed effects on state characteristics

Intercept	1.53	(0.34)
State schooling expenditures (per capita)	−14.7	(2.4)
Fraction of women without immunization	2.30	(0.54)
Head count poverty rate	1.93	(0.55)

Explanatory note: standard errors robust against heteroskedasticity in parentheses

of parents, but also via other channels. It might, for example, be the case that the better schools also provide more facilities or information concerning public health. Furthermore, low immunization coverage among women is associated to higher child mortality. It should be noted that these immunization rates are strongly correlated with state's public health care expenditures, which are, therefore, not included separately. Finally, a higher (head-count) poverty rate within the state increases child mortality rates. GDP per capita and poverty rates are too highly correlated to include both. The hazard rate already included the wealth level of the family. This implies that there is a negative externality associated to living in a state with a high level of poverty.

We also tried to include other state characteristics. In particular, we tried including the population fraction that lives in rural areas and the fraction of households that are exposed to mass media. Rural mortality rates are lower in states with a higher fraction of the population living in rural areas and with a higher exposure to mass media. However, these variables do not add much explanatory power to the earlier mentioned state characteristics. Furthermore, we tried including the state public expenditures on water and sanitation. This did not improve the fit of the model, most likely because access to water and sanitation was already included at the household level. Also public expenditure on nutrition programs did not have a substantial effect on the state fixed effects. Measham and Chatterjee (1999) indicate that, even though malnutrition rates are high in India, public expenditures on direct nutrition programs are low. Also, regional indicator variables (North, Central, East, etc.) did not have any impact on the state fixed effects.

5 Conclusions

In this paper, we have developed a flexible parametric framework for analyzing infant and child mortality. This framework is based on widely used hazard rate models, which we have extended with two features. First, we allow for frailty at multiple levels, which can be correlated with each other. Second, the model allows individual characteristics and family's socioeconomic and environmental characteristics to have different impacts on infant and child mortality at different ages. This second extension seems most relevant. A formal statistical test rejects that socioeconomic and environmental characteristics have the same impact at all ages. However, a sensitivity analysis shows that a model with constant effects over all ages yields similar marginal effects for the policy-relevant characteristics.

The key advantage of using a model with age-varying characteristics is that policy makers cannot only target families with high child mortality risks, but targeting can also depend on the age composition of children in families. The model shows, for example, that first-born boys are exposed to high mortality risks in the first month after birth, while girls who are not first born have high mortality risks after the first month after birth. Furthermore, while there are no differences in mortality rates between castes during the first year, children born

in families from scheduled castes and tribes have significantly higher mortality rates after their first birthday. Public policy is likely to become more efficient when targeted towards families with children in these higher risk categories.

Even after controlling for socioeconomic differences between families, we still found significant and substantial differences in mortality rates between states. A decomposition of these state-level fixed effects shows that lower child mortality is associated to higher schooling expenditures, higher female immunization rates, and lower poverty levels. This provides some additional evidence that state-level expenditures (and, thus, public policy) may be important in reducing infant and child mortality. The estimation results indicate that child mortality rates can be reduced substantially by improving educational attainments of women, reducing indoor air pollution, and providing safe water. Also, the presence of a doctor in the village has a significant positive effect on child mortality survival (after the first birthday).

It is interesting to consider our estimating results in the light of the National Population Policy, which was developed in 2000. The main goal of the National Population Policy is to achieve population stabilization by 2045 and to improve the quality of life.[9] A key strategy of the National Population Policy is to reduce child mortality because it is believed that a decline in child mortality will cause fertility rates to drop. Declined fertility rates should achieve population stabilization.

An important element of the National Population Policy is the strong focus on improving health and education of women and children. Our results confirm the assumption of this policy that infant and child mortality rates can be reduced by improving the education level of women. Another element of the policy is that it induces women to start childbearing at later ages and to increase the spacing between children. Except that this has a direct effect on fertility, our estimation results show that both later childbearing and increased spacing reduce infant and child mortality rates. Whether or not the National Population Policy will succeed in achieving population stabilization depends on how child mortality affects fertility decisions. It should be stressed that this has not been investigated in this paper.

Acknowledgements We gratefully acknowledge the financial support from Swedish International Development Cooperation Agency (SIDA), TF024884. We thank Kirk Hamilton, Stephen Howes, Maarten Lindeboom and two anonymous referees for valuable comments. The findings, interpretations and conclusions expressed in this paper are entirely those of the authors. They do not represent the views of the World Bank, its Executive Directors, or the countries they represent.

[9]Despite the steady decline in fertility from around six children per mother in 1951 to slightly over three children in 1997, the Indian population is still increasing and crossed one billion in May 2000.

References

Ahluwalia MS (2002) Economic reforms in India since 1991: has gradualism worked? J Econ Perspect 16(3):67–88
Arulampalam W (2007) Changing neighbourhood and infant mortality in rural India. University of Warwick, Mimeo
Arulampalam W, Bhalotra S (2006) Sibling death clustering in India: state dependence versus unobserved heterogeneity. J R Stat Soc, Ser A 169(4):829–848
Bhargava A (2003) Family planning, gender differences and infant mortality: evidence from Uttar Pradesh, India. J Econom 112(1):225–240
Bhalotra S, Van Soest A (2008) Birth-spacing, fertility and neonatal mortality in India: dynamics, frailty and fecundity. J Econom 143(2):274–290
Claeson M, Bos E, Pathmanathan I (1999) Reducing child mortality in India, keeping up the pace. World Bank, Health, Nutrition and Population (HNP) Discussion Paper, Washington
Datt G, Ravallion M (2002) Is India's economic growth leaving the poor behind? J Econ Perspect 16(3): 89–108
Deaton A (1997) The analysis of household surveys: a microeconometric approach to development policy. Johns Hopkins University Press, Baltimore
Donald SG, Lang K (2007) Inference with difference-in-differences and other panel data. Rev Econ Stat 89(2):221–233
Drèze J, Sen A (1995) Indian development; selected regional perspectives. Oxford University Press, Dehli
Fikree FF, Iqbal Azam S, Berendes HW (2002) Time to focus child survival programmes on the newborn: assessment of levels and causes of infant mortality in rural Pakistan. Bull WHO 80(4):271–276
Filmer D, Pritchett LH (2001) Estimating wealth effects without expenditure data—or tears: an application to educational enrollments in states of India. Demography 38(1):115–132
Guo G, Rodriguez G (1992) Estimating a multivariate proportional hazards model for clustered data using the EM algorithm, with an application to child survival in Guatemala. J Am Stat Assoc 87(420):969–976
Hobcraft JN, McDonald JW, Rutstein SO (1985) Demographic determinants of infant and early childhood mortality: a comparative analysis. Popul Stud 39(3):363–385
International Institute for Population Sciences (2000) National family health survey (NFHS-2), 1998–99. Mumbai
Jalan J, Ravallion M (2003) Does piped water reduce diarrhea for children in rural India. J Econ 112(1):153–173
Lee L-F, Rosenzweig MR, Pitt MM (1997) The effects of improved nutrition, sanitation, and water quality on child health in high-mortality populations. J Econ 77(1):209–235
Maitra P (2004) Parental bargaining, health inputs and child mortality in India. J Health Econ 23(2):259–291
Maitra P, Pal S (2008) Birth-spacing, fertility selection and child survival: analysis using a correlated hazard model. J Health Econ 27(3):690–750
Makepeace G, Pal S (2008) Understanding the effect of siblings on child mortality: evidence from India. J Popul Econ 21(4):877–877
Measham AR, Chatterjee M (1999) Wasting away, the crisis of malnutrition in India. World Bank, Washington
Mosley WH, Chen LC (1984) An analytical framework for the study of child survival in developing countries. Popul Dev Rev 10(Suppl):25–45
Pandey A, Choe MK, Luther NY, Sahu D, Chand J (1998) Infant and child mortality in India. National Family Health Survey Subject Report 11. East-West Center Program on Population, Hawaii
Pitt MM (1995) Women's schooling, the selectivity of fertility and child mortality in Sub-Saharan Africa. World Bank, LSMS Working Paper 119
Pitt MM (1997) Estimating the determinants of child health when fertility and mortality are selective. J Hum Resour 32(1):129–158
Pritchett L, Summers LH (1996) Wealthier is healthier. J Hum Resour 31(4):841–868

Ravallion M, Datt G (2002) Why has economic growth been more pro-poor in some states of India than in others? J Dev Econ 68(2):381–400
Ridder G, Tunalı İ (1999) Stratified partial likelihood estimation. J Econ 92(2):193–232
Sastry N (1997) A nested frailty model for survival data, with an application to the study of child survival in northeast Brazil. J Am Stat Assoc 92(438):426–435
Schultz TP (1984) Studying the impact of household economic and community variables on child mortality. Popul Dev Rev 10(Suppl):215–235
Strauss J, Thomas D (1995) Human resources: empirical modeling of household and family decisions. In: Behrman J, Srinivasan TN (eds) Handbook of development economics, vol 3A. North-Holland, Amsterdam
Van den Berg GJ (2001) Duration models: specification, identification, and multiple duration. In: Heckman JJ, Leamer EE (eds) Handbook of econometrics, vol 5. North-Holland, Amsterdam
Vaupel JW, Manton KG, Stallard E (1979) The impact of heterogeneity in individual frailty on the dynamics of mortality. Demography 16(3):439–454
Wang L (2003) Determinants of child mortality in LDCs: empirical findings from demographic and health surveys. Health Policy 65(3):277–299
Withworth A, Stephenson R (2002) Birth spacing, sibling rivalry and child mortality in India. Soc Sci Med 55(12):2107–2119
Wolpin KI (1997) Determinants and consequences of the mortality and health of infants and children. In: Rosenzweig MR, Stark O (eds) Handbook of population and family economics, vol 1A. North-Holland, Amsterdam