DISCUSSION PAPER SERIES

IZA DP No. 2488

Persistence in Infant Mortality: Evidence for the Indian States

Wiji Arulampalam
Sonia Bhalotra

December 2006

Forschungsinstitut
zur Zukunft der Arbeit
Institute for the Study
of Labor

Persistence in Infant Mortality: Evidence for the Indian States

Wiji Arulampalam
University of Warwick
and IZA Bonn

Sonia Bhalotra
University of Bristol,
CSAE, QEH and CHILD

Discussion Paper No. 2488
December 2006

IZA

P.O. Box 7240
53072 Bonn
Germany

Phone: +49-228-3894-0
Fax: +49-228-3894-180
E-mail: iza@iza.org

IZA Discussion Paper No. 2488
December 2006

ABSTRACT

Persistence in Infant Mortality: Evidence for the Indian States*

This paper investigates the high correlation in infant mortality across siblings using micro-data for each of the fifteen major states of India. The main finding is that, in thirteen of the fifteen states, there is evidence of a causal effect of a child death on the risk of death of the subsequent child in the same family (a scarring effect), which is identified after controlling for observed and unobserved heterogeneity at the family level. The two states in which evidence of scarring is weak are Punjab, the richest, and Kerala, the state that is most advanced in socio-economic terms. In the other states, scarring effects are large. Indeed, the only other covariate that has a marginal effect on mortality that is as big, or bigger, than the survival status of a preceding sibling is an indicator for mothers having attained secondary or higher levels of education. These results show that policies targeted at reducing infant mortality will have social multiplier effects through helping avoid the death of subsequent siblings. The size of the scarring effect depends upon the gender of the previous child in three states, in a direction consistent with son-preference. Comparison of other covariate effects across the states offers some new insights, there being no previous research that has compared the determinants of infant mortality across the Indian states.

JEL Classification: J1, C1, I1, O1

Keywords: persistence, siblings, infant mortality, state dependence, scarring, unobserved heterogeneity, dynamic random effects logit, gender, India

Corresponding author:

Wiji Arulampalam
Department of Economics
University of Warwick
Coventry, CV4 7AL
United Kingdom
E-mail: wiji.arulampalam@warwick.ac.uk

* We are grateful for financial support from the ESRC under Research Grant number RES-000-22-0651 and to ORC Macro International for providing us with the data. We are grateful to Arthur van Soest and Mike Veall for many helpful comments. The paper has benefited from presentation at the Universities of Toronto, McMaster, Tilburg, Essex and Southampton, the ESRC Econometric Study Group Meeting and at an international workshop we organised at the University of Bristol on Child Health in Developing Countries. We are grateful to DFID-UK for funding the workshop.

1. Introduction

While there is considerable research on the determinants of the level of infant mortality (i.e., mortality in the first year of life), and on regional and gender inequalities in its incidence, research on inequality in the risk of death between families is relatively limited. The family is an important institution, the inherent characteristics and behavioural choices of which impact upon outcomes for children (infant survival chances in this case). Indeed, data from a wide spread of regions in developing countries show strong evidence of family effects, with a small fraction of families accounting for most child deaths. Clearly, some families will have characteristics (e.g. lower level of maternal education) that predispose their children to higher death risk, and many previous studies have been concerned with identifying these characteristics. Recent demographic research has shown that, on top of differences in observed characteristics, there is substantial unobserved heterogeneity (e.g. genetic frailty, unobserved characteristics of the environment) between families (e.g. DasGupta, 1990; Curtis *et al*, 1993; Guo, 1993; Zenger, 1993; Sastry, 1997). This paper investigates whether, in addition to inter-family heterogeneity that produces a positive correlation of sibling death risks, there is a causal process at work, whereby the actual event of death of a child results in a higher risk of death for the next child in the family.

The basic idea is not new. It is the problem, well-known in labour economics, of separating state dependence (or scarring) from unobserved heterogeneity (e.g. Heckman 1981). A contribution of our work is to bring this distinction to analysis of the important problem of childhood death in developing countries. In the traditional setting, state dependence refers to the dependence of an outcome (e.g. current unemployment risk) for an individual on his or her history of outcomes (previous unemployment spells), given his or her characteristics. Given the natural sequencing of siblings in time, an

analogous model can be specified in which the outcome (e.g. mortality risk) for a child in a given family can be described as a function of previous childhood deaths in that family, given family characteristics.[1] The natural sequencing of siblings and the availability of data on the first-born child of each mother can be exploited to address the classical problems that arise in identification of endogenous effects given endowment heterogeneity.

In our earlier work, we focused on methodological issues and, in particular, on the potential bias created by left-truncation of the data, which is standard practice in demographic research (Arulampalam and Bhalotra 2006). In this paper, we are primarily concerned with the empirical question of how prevalent scarring is, and whether it decreases or disappears with socio-economic development. The investigation is therefore conducted using micro-data on infant mortality and its covariates for each of the fifteen major Indian states. These states are comparable in size to European countries. They exhibit large differences in social, demographic and economic indicators (e.g. Dreze and Sen, 1997). A second objective of this paper is to investigate whether scarring effects are larger following the death of a boy rather than a girl. This question is also best addressed by looking at state-specific estimates, given previous evidence that son-preference in India is concentrated in the North and West of the country (e.g. Miller 1981). Scarring aside, no previous research has provided comparable evidence for the Indian states that is relevant to understanding the vast differentials in infant mortality within the country (see World Bank 2004 for a survey of the relatively small body of research on infant mortality in India).

[1] Hence our borrowing of the terms state dependence and scarring from the unemployment literature. Intuitively, the death of a child *scars* or marks the survival prospects of the succeeding sibling. Alternatively, defining a state as a realisation of a stochastic process, one may think of state dependence (at the family level) in terms of the (infant) mortality risk facing a child being dependent upon the state (died in infancy or not) revealed for the previous child in the family.

India makes an interesting laboratory for the study of demographic processes. The size of India results in large sample sizes for statistical analysis of what is a rare event and, even more, it increases the relevance of the research to policy and well-being. India contains one in six of the world's people and accounts for a quarter of the under-5 child deaths in the world (Black *et al,* 2003). About 70% of under-5 deaths in India occur in infancy. Estimates for 1998/9 suggest that nearly 1¾ million Indian children died before their first birthday (World Bank 2004, Introduction). Infant mortality has been declining, having halved between the early-1970s and 2000, but the decline is less impressive than that in some (poorer) South and South-east Asian countries (e.g. Claeson *et al* 2000). It is plausible that persistence associated with scarring effects has slowed the decline in mortality.

We find that, after allowing for all inter-family differences, there is evidence of scarring in thirteen of the fifteen states. The two states with smaller and less significant scarring are Punjab, the richest, and Kerala, which is most advanced in social and demographic indicators. A ranking of the other states by the size of the scarring effect confirms that scarring depends upon both social and economic development. The scarring effects are large. Indeed, the only other covariate that has a marginal effect on mortality that is as big as or bigger than the survival status of a preceding sibling is an indicator for mothers having attained secondary or higher levels of education. The important role of mother's education in improving outcomes for children is now a well-established finding (e.g. Strauss and Thomas 1995). In contrast, scarring is a neglected idea in the literature on mortality. The other main finding is that, in three of the fifteen states, the scarring effect is significantly larger when driven by the death of a boy as opposed to a girl. These states are Punjab, Rajasthan and the North-Eastern cluster, and son-preference is known to be strong in Punjab and Rajasthan. Insignificant gender

differences in scarring in the remaining twelve states are consistent with the fact that gender mortality differentials are more marked in the age-range 1-5 years than in the age range of 0-1 year that is considered here. Comparison of the effects of other covariates across the states provides some new insights. For instance, we find that the only state with higher infant mortality risk for girls, other things equal, is Punjab, the richest state. It is well known that low-caste individuals have suffered historical disadvantage in India but hard evidence on the extent and location of this disadvantage is relatively limited. We find that children of scheduled castes and tribes and other backward castes are significantly more likely to experience infant mortality in three of the fifteen regions, namely, Punjab, Uttar Pradesh and Rajasthan. In the other states, conditional upon other covariates like parental education, the lower castes are not significantly more likely to suffer infant death.

Section 2 sets the paper in the context of some related research and it outlines mechanisms that might explain scarring. The data and the econometric model are described in Section 3. The variables used in the analyses are described in Section 4 and the results are presented in section 5. Sensitivity of estimated scarring effects to various model extensions are considered in Section 6, and conclusions are presented in Section 7.

2. Background

Related Literature

Like demographers, economists have tended to identify the observed correlation of sibling outcomes with family background (i.e. inter-family heterogeneity); e.g. Solon *et al* (1991). Several studies have used sibling data to difference out unobservable elements of family background, with the aim of identifying behavioural effects (see, for

example, Behrman and Wolfe, 1984; Neumark and Korenman, 1994; Rosenzweig and Wolpin 1994; Altonji and Dunn, 1996).[2]

We are aware of two previous studies that take data on correlated outcomes within the family and seek to disentangle genuine state-dependence from unobserved heterogeneity. Heckman, Hotz and Walker (1985) start with the observation, made in a previous demographic literature, that the successive birth interval durations of women are positively correlated, or that a long preceding birth interval predicts a long subsequent birth interval for the same woman. They show that this "well-noted empirical regularity" (their words) vanishes, at least for married women, once controls for unobservables at the woman-level are introduced. As explained in section 1, we also start with an empirical regularity noted in the demographic literature, namely, that the death of a child predicts a higher risk of death for the subsequent child in the same family. In line with Heckman *et al*, we seek to separate the structural from the "spurious" explanations of this finding. In contrast to Heckman *et al*, we find evidence of genuine state dependence in infant mortality. The other study that is similar in spirit is Oettinger (2000). He looks to identify causal effects of an individual's schooling on the schooling (attainment) of his or her younger sibling, after allowing for shared traits amongst siblings. There are other studies that analyse the effects of sibling characteristics like gender on outcomes for subsequent siblings (e.g. Butcher and Case, 1994; Kaestner, 1997). However, gender is an exogenous variable, and we are interested here in causal effects flowing from endogenously determined outcomes.

[2] The causal influence of interest is usually a parental choice or a policy-amenable input, for example, parental education, teenage motherhood, school years, or school quality. Outcomes studied in this way include school attainment or achievement, birth weight and foetal growth, the returns to education, wages and socio-economic status.

There is an interesting parallel, in this sense, with the literature on social networks and neighbourhoods. It is commonly observed that people who share residential location, race or ethnicity have correlated outcomes. These are often associated with exogenous effects that reflect similarity of characteristics and constraints, or else that define group membership. Recent research attempts to separate from these exogenous effects any endogenous effects arising from the propensity of an individual to behave in a certain way, resulting in a causal influence on the behaviour of other members of the group (see, for example, Moffitt, 2004; Aizer and Currie, 2004). This is similar to the problem in this paper once the group is defined as a group of siblings.

An interesting feature of the analysis when the group is a group of siblings is that the reflection problem that plagues analysis of correlated effects in neighbourhoods and peer groups (Manski, 1995) can be avoided by virtue of the natural sequencing of siblings by birth order. This allows us to re-cast the problem in terms of a dynamic model with unobserved heterogeneity where the endogenous effect is represented as a first-order Markov process, running from the survival status of a child to the survival chances of the subsequent child.

In summary, this paper addresses a question of immense and immediate importance that appears not to have been addressed before. The structure of the problem and the methods employed intersect with research in economics on unemployment persistence, on the importance of family background in determining child outcomes, and on endogenous effects in groups/neighbourhoods.

Scarring Mechanisms

This section illustrates the sorts of mechanisms that might drive state dependence in infant mortality. In other words, why would the death of a child lead to a higher risk of

death for the next child of the same mother, once all observed and unobserved differences between mothers are held constant? One mechanism is that which operates by the death of a child shortening the time to the next birth. This process may be set off in either of two ways. One possibility, which we refer to as *the fecundity hypothesis,* is that the death of an infant results in the mother ceasing to breastfeed and, thereby, being able to conceive sooner than otherwise (e.g. Bongaarts and Potter 1983, Kennedy et al. 1992). An alternative is the *replacement hypothesis,* according to which the death of a child leads parents to (intentionally) conceive sooner in a desire to "replace" their loss (e.g. Preston 1985). In both cases, it is a short preceding birth interval for the index child that causes an elevation of her death risk. There is plenty of evidence in the demographic and medical literature that short preceding birth intervals result in higher death risk (e.g. Cleland and Sathar 1984, Stephansson et al. 2003), especially amongst poor women (e.g. Rawlings et al. 1995). This is thought to be because it takes time for an under-nourished mother to recover physiologically from a birth, and to replenish her stock of the nutrients essential to support the next pregnancy. Many studies have suggested that a minimum interval of about 24 months is needed (e.g. daVanzo and Pebley, 1993) but, in India, almost 40% of birth intervals are shorter than 24 months (see Appendix Table 1).

A further possibility is that a child death leaves the mother depressed, as a result of which her subsequent child's health is compromised, both in the womb and in early infancy; we call this the *depression hypothesis*. It is especially interesting as it can explain scarring for a given duration of the preceding birth interval. The possibility that maternal depression is a causal factor in childhood mortality in high-mortality environments has not been previously recognised in the demographic literature but there is supporting evidence in the medical literature (e.g. Steer et al. 1992, Rahman et al.

2004).[3]

The discussion so far indicates positive scarring effects. In fact there may be learning effects that result in the mortality risk of the index child *falling* on account of the death of the preceding sibling. For instance, if an older sibling dies of diarrhoea, the mother may rush to learn how to prevent diarrhoea-related infant death. Any positive degree of scarring (state dependence) that is identified is then net of learning effects. Although it is of policy significance to establish which mechanisms underlie state dependence and there is little definitive research in this area, this paper is concerned primarily with the prior task of identifying whether there are any state dependence effects after controlling for observed and unobserved heterogeneity.[4] In Section 6, we investigate the hypothesis that birth-spacing drives scarring, albeit with some qualifications.

3. Methodology and Empirical Analysis

The Data

The data used are from the Indian National Family Health Survey (NFHS) of 1998-99, which interviewed 92300 ever-married women aged 15-49 at the time of the survey (see IIPS and ORC Macro, 2000). They contain a complete retrospective history of births for each mother, together with a record of child deaths.[5] We are therefore able to construct

[3] The analyses of death clustering that we refer to are restricted to data from developing countries where infant mortality risks are high, and where family and sibship sizes are large. It is nevertheless interesting to consider the relation of this research to recent research on multiple infant deaths within families in richer countries, associated with Sudden Infant Death Syndrome (SIDS) (see Firstman and Talan, 1997), for example). SIDS is, almost by definition, a phenomenon in which the cause of death cannot be identified. Where families have experienced multiple infant deaths, mothers have been implicated in a number of cases, especially in the UK. Media coverage of these events is consistent with a depression story.

[4] See Bhalotra and van Soest (2006) for an attempt at quantifying the birth-interval related mechanisms by endogenising birth spacing.

[5] The NFHS is one of a series of comparable Demographic Health Surveys (DHS), available for about sixty-nine low and middle-income countries. The ideas and methods introduced in this paper are therefore immediately applicable to other regions.

(unbalanced) panel data for mothers in which the length of the panel corresponds to the number of births. The width of the panel (number of mothers) varies between 9370 (the North Eastern states) and 2340 (Kerala); see Appendix Table 1.

Descriptive Statistics

Table 1 reports mortality rates by state, which are averages over the data sample, including births across the four decades, 1961-1999.[6] Of every 1000 children born in India over this period, 82 died in infancy. There is remarkable variation across the Indian states. For example, the large backward state of Uttar Pradesh (UP) in Central India had a mortality rate (in 1000) of 116, while the Southern state of Kerala, known for its relative success in human development, had a rate of 36.[7]

The raw data probability of infant death conditional on whether the preceding sibling died in or survived infancy is displayed in Columns 2 and 3 of Table 1. The difference of these conditional probabilities (column 4) is a measure of the extent of death clustering. This ranges from about 0.09 in Punjab and Maharashtra, to 0.18 in Bihar. These are enormous increases in risk, given an average mortality rate of 0.082 in India. Column 5 contains an alternative presentation of the data. The relative odds of a child dying in infancy if the previous sibling died rather than survived infancy lie between 2.9 and 4.8.[8] Overall, there is a remarkable degree of death clustering in India, and also huge variation in this across the states. These, however, are simply the observed tendencies in the data. Estimation of the statistical model allows us to

---

[6] For this reason, they are larger than recently published UN figures that refer to recent years; and the ranking of states is not the same as that for mortality rates today. Estimates of infant mortality for 1998/9 reported in World Bank (2004) are 68 for the country as a whole, ranging from 14 in Kerala to 96 for Orissa.

[7] The relative position of the states with respect to relevant indicators is discussed in the Appendix to the working paper version (Arulampalam and Bhalotra 2004). Appendix Table A1 in this paper presents some illustrative descriptive statistics.

[8] The relative odds ratio reported here corresponds to the coefficient on the previous child's survival status in a logit regression of the survival status of the index child on an intercept and the survival status of his or her previous sibling.

disentangle clustering effects into correlated risks amongst siblings (inter-family heterogeneity) and, conditional upon this, a causal effect of the death of one sibling on the risk of death of the next sibling (state dependence or scarring).

The Econometric Model

Let there be n_i children in family *i.* For child *j* (*j=2,…,*n_i) in family *i* (*i=1,2,…, N*), the unobservable propensity to experience an infant death, y_{ij}^*, is specified as

$$y_{ij}^* = \boldsymbol{x_{ij}}'\beta + \gamma y_{ij-1} + \alpha_i + u_{ij} \qquad (1)$$

where $\boldsymbol{x}$ is a vector of strictly exogenous observable child and family specific characteristics that influence y_{ij}^* and β is the vector of coefficients associated with $\boldsymbol{x}$. A child is observed to die when his or her propensity for death crosses a threshold; in this case, when $y_{ij}^* > 0$. The model has a random intercept α_i, to account for family-specific unobserved characteristics. This picks up any correlation of death risks among siblings arising, for example, from shared genetic characteristics or from the innate ability of their mother. The model also includes the observed survival status of the previous sibling, y_{ij-1}, the coefficient on which picks up scarring. The null of no scarring implies $\gamma=0$.[9] Equation (1) reflects the first-order Markov assumption common in models of this type (see Zenger (1993), for example). This is that, conditional on $y_{ij-1,}$ $\boldsymbol{x_{ij}}$ and α_i, the survival status of other older children has no impact on y_{ij}^*. If child (*j-2*) died then, in our model, this would affect the risk of death of child (*j-1*) and, *thereby*, affect the risk of death of child *j*.[10] A model restricted to first-order effects is consistent with the

[9] The estimated parameter γ should be interpreted as the 'average' effect of scarring over the time period considered.

[10] This is plausible since we are conditioning on α_i, and any risk factors common to the siblings, j-2, j-1 and j will be captured by α_i.

mechanisms that we suggest might drive scarring (see Section 3 above, and Zenger 1993).

Since (1) is a dynamic model, it faces the initial conditions problem (e.g. Heckman, 1981; Wooldridge, 2002). The problem is that y_{ij-1} and α_i are necessarily correlated. This problem would become unimportant if the number of children per mother were to tend to infinity but this, clearly, is not the case. In standard applications of this model, it is unusual that the start of the stochastic process of interest coincides with the start of the sample observations. Instead, the available data are typically left-truncated. However, our data contain complete retrospective histories of fertility and mortality for each mother. We are thus able to model the initial condition of the process as a natural extension of the model given in (1). We specify the equation for the first-born child in each family as

$$y_{i1}^{*} = \boldsymbol{z_i}'\boldsymbol{\lambda} + \theta\,\alpha_i + u_{i1} \qquad i=1,....,N \ \ and\ j=1 \qquad (2)$$

where $\boldsymbol{z_i}$ is a vector of exogenous covariates. In general, equation (2) allows the vector of covariates $\boldsymbol{z}$ to differ from $\boldsymbol{x}$ in (1). However, we set the two vectors of covariates to be the same given that we observe the process from the start. We allow the effect of unobservable family characteristics to differ from that in (1) by the proportion θ. If we were to find that $\theta=0$, then we could conclude that unobserved heterogeneity does not enter (2), from which it follows that the initial conditions problem is empirically unimportant. A test of the significance of θ is presented in the Results section. A potential issue with this identification strategy is that, if the first conception is a miscarriage, then the first-born (live) child is not a good proxy for the initial condition of the process. In other words, the data may be *implicitly* left-truncated. This problem cannot be directly addressed or assessed because the data do not record miscarriages. However, in our earlier work, we show that the bias associated with left-truncated data

is largely redressed by modelling equation (1) jointly with a reduced form equation like (2) for the first-*observed* child in the sample (see Arulampalam and Bhalotra, 2006).

Equations (1) and (2) together specify a complete model for the infant survival process. In this way, the endogeneity of the "lagged dependent variable", that is, the previous child's survival status, is taken into account. To the extent that this is driven by shorter birth-spacing or higher fertility, we are implicitly allowing for the effects of reproductive behaviour on mortality. The estimated model may be thought of as the reduced form of a full structural model of fertility and mortality. Explicit estimation of a structural model would involve making what are likely to be arbitrary identifying assumptions, it being difficult to find valid instruments.

We assume that u_{ij} is independently distributed as a logistic distribution (Λ), and that the family-specific unobservables, α_i, are independent and identically distributed as normal (density φ).[11,12] Marginalising the likelihood function with respect to α_i gives, for family *i*

$$L_i = \int_{-\infty}^{\infty} \Big(\prod_{j=2}^{n_i} \Lambda[(\boldsymbol{x_{ij}}'\boldsymbol{\beta} + \gamma y_{ij-1} + \sigma_\alpha \tilde{\alpha})(2y_{ij} - 1)\Big]$$

$$\Lambda[(\mathbf{z}_i'\boldsymbol{\lambda} + \theta\, \sigma_\alpha\, \tilde{\alpha})\ (2y_{i1} - 1)]\ \Big)\ \varphi(\tilde{\alpha})\, d\tilde{\alpha} \qquad (3)$$

where, $\tilde{\alpha} = \alpha/\sigma_\alpha$. The log likelihood function is maximized using a routine written by the authors in Stata (2000).

[11] We investigated the probit model and found that the results were not sensitive to functional form.

[12] In dynamic models where the index *j* represents time, one might wish to allow for serial correlation in u_{ij} to capture any persistence in the effects of shocks. This is often done by writing $u_{ij}=\rho u_{ij-1}+\varepsilon_{ij}$, where ρ is the persistence parameter. This is not appropriate here since the index *j* refers to child number *j* and there is no reason to believe that ρ is constant across children in a family.

Previous analyses of dynamic models with unobserved heterogeneity have shown potential sensitivity of the estimates to the assumption made about the distributional form for unobserved heterogeneity, α_i (e.g., Heckman and Singer, 1984). A weakness of the normality assumption is that it may not be flexible enough to account for the fact that some families never experience any child deaths and that, in some families, all children die (the mover-stayer problem). Our sample does not contain any families in which all children die in infancy. However, there are many families that experience no infant deaths, and this is accommodated by allowing for a single (empirically determined) mass at minus infinity: a very large negative value for α_i gives a very small value for y_{ij}^*, and hence a very small probability of observing death of the index child.[13] The modified likelihood for family *i* is given as,

$$L_i^* = \frac{\psi_0}{1+\psi_0}\left[\prod_{j=1}^{n_i}(1-y_{ij})\right] + \frac{L_i}{1+\psi_0} \qquad (4)$$

where L_i is given by equation (3) and ψ_0 is the unknown end-point parameter. The estimated proportion of families who will have a very small α_i is given by p, where,

$$p = \frac{\psi}{1+\psi} \qquad . \qquad (5)$$

In order to ensure the non-negativity of ψ, it was parameterised as $exp(\kappa)$, and κ was estimated.

In addition to mother-specific unobserved heterogeneity, community-level random effects were allowed in order to correct the standard errors for community-level clustering in the sample design. The community effect is treated as a nuisance parameter because we cannot interpret a time-invariant community-level effect in any meaningful manner: Children of the same mother, born at different dates, will experience different

[13] See Narendranathan and Elias (1982) for an application of this distributional assumption in the context of modelling individual unemployment.

community-level effects, especially where development of community infrastructure is rapid.

4 The Empirical Model

The dependent variable and the survival status of the preceding child were both coded as binary variables that are unity if the child dies before the age of 12 months and zero otherwise. Since the data show some age-heaping at 6 month intervals, we investigated sensitivity of the estimates to altering the definition to include deaths at 12 months. As the results were similar, they are not reported. Children who were younger than 12 months at the time of the survey were dropped from the sample because they had not had 12 months exposure to mortality risk.

Child-specific regressors in the model include birth-order, gender, an indicator for whether the child is one of a multiple birth (twin, triplet, etc) and the age of the mother at birth of the index child. For birth-order, we use a set of dummies so as to allow for a non-linear pattern. This may be expected given the evidence that mortality risk tends to be higher amongst first-borns than for subsequent siblings, and to then rise amongst higher birth-orders, possibly on account of maternal depletion. Maternal age at birth is, of course, related to birth order. Other things equal, higher birth-order children will be born when the mother is older. Inclusion of this variable "purges" the estimated birth-order effects of maternal age effects. We include a quadratic in maternal age to allow for higher risk at younger and older ages. This is especially relevant in poor countries, where many women are teen mothers. Family-specific covariates included are the educational attainment of each of the mother and father, religion and caste. These are all included as dummy variables to allow flexibility. Cohort effects were modelled by including indicator variables for year of birth of the mother during 1948-1959, 1960-1969 and 1970-1984. These are expected to pick up any secular decline in death risks

over time, other things equal, and are especially important since our strategy involves using retrospective histories that go back several years in time. Notice that, since the model also includes the age of the mother at birth of the index child, these variables effectively control for the date of birth of the child.[14]

Information on household assets, immunization, prenatal care, access to piped water and relevant community-level variables are not used because they are time-inconsistent. These data are available at the time of the survey, while exposure to the risk of infant death in these data spans about three decades. The same holds for breastfeeding. If we had information on breastfeeding for every child, it would help illuminate the mechanisms underlying scarring, but this information is available only for recently born children. In order to incorporate these variables we would have to left-truncate the data. We would then not only lose the first-born child of most mothers (see section 3), but we would also be left with too few children per mother to be able to identify both the within-family dynamics that create scarring and mother-level unobserved heterogeneity. A further problem with incorporating these variables in the model is that they are potentially endogenous. For example, families will tend to simultaneously decide what resources to allocate to the purchase of a bicycle or a TV and what resources to allocate to attending immunization clinics to reduce the risk of child mortality. Access to facilities like piped water will be endogenous if selected families migrate to regions with these facilities, or if governments place these facilities in regions with worse health indicators (see Rosenzweig and Wolpin 1986). Breastfeeding is also endogenous to the extent that children who are unhealthy at birth are often unable to suckle.

[14] To see this, consider a woman who was born in 1940 and gave birth to the index child in 1960 so that the age of the mother at birth of the child is 20. Since the model includes "20" and "1940", it implicitly includes "1960".

The mother-level random effect that we include in the model will control for the time-invariant component of these omitted variables, for example for the fact that some mothers are more prone to breastfeed than others. Maternal age at birth and maternal cohort (maternal year of birth) will tend to capture unobservable trends in the data. Suppose, though, that these controls are inadequate, and that the omitted variable is correlated with the previous child's survival status. Then, our estimates of the scarring parameter *are* subject to bias. However, it seems plausible to argue that this bias will be downward. Suppose that the previous child wasn't immunized (or the mother did not seek prenatal care for that child's pregnancy) and that this child died (so $y_{ij-1}=1$). It seems likely that the mother will then seek immunization for her next child. Since immunization of child j is positively correlated with the mortality status of child j-1 (y_{ij-1}) and negatively correlated with the mortality status of child j (y_{ij}), omitting immunization from the model will create a downward bias on the coefficient on y_{ij-1} in a model of y_{ij}. In sum, if there is an omitted variable bias, the estimates of scarring that we provide are conservative estimates.

5 Results

Unobserved heterogeneity

As discussed in Section 2, economists have studied the extent to which the socio-economic outcomes of siblings are correlated in order to understand the force of family background and, thereby, the perpetuation of inequality across the generations (e.g. Solon *et al* 1991). Demographers have interpreted family-level effects in mortality equations as a measure of the importance of genetic traits (e.g. Sastry 1997) or, occasionally, other variables like maternal ability (e.g. DasGupta 1990).

Column [9] of Table 2 presents the proportion of the error variance attributed to unobserved heterogeneity. This ranges between 0.019 in Haryana and 0.212 in West

Bengal. The estimates reject the null of no family-level unobservables in twelve states, at the 10% significance level but, in the relatively developed states of Punjab, Haryana, Tamil Nadu and Kerala, unobservables have limited power to explain death clustering. Previous research in demography has tended to over-estimate the contribution of fixed family traits by virtue of neglecting scarring. In the following section, we show how our estimates of scarring levels in each state, and of the differences in scarring between states, would change if we ignored unobserved heterogeneity.

Scarring

Refer Table 2. For ease of reference, column 1 repeats our measure of the extent of death clustering or persistence in the *raw* data; this was first displayed in Table 1. Estimates of scarring are then presented under two different conditions. First, we report estimates from a model that ignores unobserved heterogeneity (Panel 1) and then we report estimates from the preferred model that allows for this (Panel 2). The *marginal effect* associated with $\hat{\gamma}$, the coefficient on the previous child's survival status, is computed as the difference between the sample averages of the probability of death predicted by the estimated model when y_{ij-1} =0 and when y_{ij-1}=1.[15] This is reported in columns 2 and 6.

The main result is that death in infancy of a previous sibling raises the probability of infant death for the index child in every state, and this result remains after controlling for a number of child and family-specific characteristics and for all unobserved differences between families. In Panel 2, the scarring effect is significant, at the 5% level, in 13 states. In the other two states, it is significant at the 13% level (Punjab) and at the 9% level (Kerala). Comparison of columns 2 and 6 shows that

[15] This is approximately equivalent to the first partial derivative of the conditional probability of death of the index child (the conditional expectation of y_{ij}) with respect to the covariate.

failing to control for unobserved heterogeneity can result in marginal effects that are as much as two or three times as large as the correct effect. Using the preferred estimates in column 6, consider how the marginal effect associated with a previous sibling's death compares with the marginal effect of other influences on mortality (see Tables 2 and 5). For singleton births, no variable other than mother's having achieved secondary or higher education has a comparable effect! In India as a whole, 10% of mothers in the sample for India as a whole have this level of education, and 9.6% of children have a preceding sibling who died in infancy.

Punjab and Kerala have the lowest mortality rates. Kerala has human-development and demographic indicators that put it in a different league from the rest of India. Punjab is the richest of India's states. If the fecundity and replacement mechanisms discussed in section 2 were the driving force in scarring, then we might expect these two states, which have the smallest scarring effects, to have the longest birth intervals. It is therefore striking to find that exactly the opposite is the case in Punjab. This state has the highest fraction (21%) of births with a preceding birth interval of less than 18 months (see Appendix Table A1, column 5). The apparent paradox is resolved if one recognizes that the birth interval is a choice variable, and that the impact of short birth intervals on mortality risk is increasing in poverty (e.g. Rawlings et al. 1995). For instance, it is clear that a birth interval of less than 18 (or 24) months is not as risky in OECD countries as it is in developing countries. The reason for this is that mothers are less well-nourished in developing countries and so their bodies need a longer time to recoup from the demands of pregnancy and birth (e.g DaVanzo and Pebley 1993). So, if women in Punjab (and Kerala) are, on average, strong enough to endure shorter birth intervals then this could explain both the observation that birth

intervals are shorter, and also the observation of a low degree of scarring in these states.[16]

Column 7 shows the percentage of raw persistence (or clustering) that is explained by the within-family process of scarring. This lies between 16.4% (in West Bengal) and 61.8% (in Haryana). Consistent with this finding, column 9 shows that the percentage of the error variance that is attributable to unobserved heterogeneity is smallest in Haryana, and largest in West Bengal. Notice that, in column 4, the contribution of scarring to raw persistence in these two polar cases is inflated to 72.5% in West Bengal and 85% in Haryana. So, a model that fails to allow for unobserved heterogeneity not only over-estimates the level of scarring in every state, but it also seems to under-estimate differences across the states in scarring.

Comparing the model predicted probability of death with the predicted probability of death when scarring is set equal to zero offers an estimate of the reduction in mortality that would be achievable if scarring were eliminated. This is a useful expression of its significance. The estimates in column 8 suggest that, in the absence of scarring, mortality rates would fall by between 3.1% (in Punjab) and 10.8% (in Madhya Pradesh). The estimates exclude the probability of death attached to first-borns. These results are consistent with the expectation that the impact on infant mortality of eliminating scarring will tend to be smaller in states that have lower fertility levels [state-level data on fertility are in Appendix Table A1]. As expected, the corresponding estimates flowing from a model with no unobserved heterogeneity (see column 5) are larger. In addition, they produce a different state ranking, which further underlines the importance of a correct model specification.

[16] Further discussion of the pattern of results across the states is in the working paper version of this paper, Arulampalam and Bhalotra (2004).

Overall, these results have strong implications for policy, as discussed in Section 1. Scarring, unlike inter-family heterogeneity, involves responsive behaviour, which may be amenable to policy, while the latter involves largely untreatable factors like genes or fixed behavioural traits.[17]

Diagnostics

A test of the null hypothesis that $\theta=0$ in equation (2) is reported in Table 3, column 1. The null is rejected in 6 of the 15 states at a significance level of 10% or less. This underlines the importance of addressing the initial conditions problem, as it is a test of the hypothesis that the outcome for the first child within a family can be treated as exogenous. To see this, observe that if $\theta=0$, then unobservables in the equation for the first observation are uncorrelated with unobservables in the [dynamic] equations for subsequent observations (see Section 3). The model then collapses to a simple random effects model.

Estimates of the parameter p, the mass point at minus infinity, are in column 2 of Table 3. This is the estimated proportion of families with a very large negative value of α_i, which would be consistent with having an extremely low probability of infant deaths in the family. It is significant in four of the fifteen states (at less than 10% level), demonstrating the practical relevance of allowing for departures from normality in some cases. In line with the discussion above, this number is large in Haryana, and much smaller in high-incidence states like Uttar Pradesh.

[17] A similar distinction between alterable behaviour (such as parenting style) and unalterable family-specific traits (for example, as captured in genotypes) is central to the nurture-nature debate (e.g. Pinker 2002). Twin studies have played a critical part in effecting the separation between nurture and nature in analyses motivated by this debate. In this paper, the objective is not to identify the importance of genotypes; instead, we define all characteristics that siblings share by virtue of belonging to the same family (mother), as inter-family heterogeneity. We then seek to identify behavioural effects stimulated by an infant death on the risk of infant death for the subsequent child in the same family.

Other Covariates

In considering the effects of the other covariates (**x**), reported in Table 5, it is important to remember that these are conditional effects, obtained from a model that includes the previous child's mortality status (y_{ij-1}) and a family-level random effect (α_i). In particular, these are marginal effects of, say, caste or parental education, on the current risk of death of the individual child given their family (mortality) history. They are not strictly comparable with previous results in the literature obtained from estimation of simple reduced form logit or probit models of mortality.

The indicator variable for girls is insignificant in most states, though there are some notable exceptions. Most striking is the finding that the only state in which girls suffer excess mortality in infancy is Punjab, which is the richest of the states![18] The probability that a girl born in Punjab dies is 0.028 points higher than the probability a boy dies. This is almost half of the average risk of infant death in Punjab, which is 0.06! In West Bengal and in the North-Eastern states, girls have a significant survival advantage in infancy. Multiple births suffer significantly higher risks in all states, and the marginal effects are huge. This is a common finding in the demographic literature. Infant death risk is non-linear in birth-order. Our estimates suggest that, relative to first-borns, there is an increase in risk at birth-orders 3, 4, or higher. Higher education amongst fathers and mothers shows significant effects in about half the states. The pattern of effects shows no obvious relation with the socio-economic position of the state. Maternal age at birth of the index child has a U-shaped effect that is significant in every state. The disadvantaged castes suffer higher mortality risk only in the states of

[18] Girls are born with relatively good survival chances, which appear to be gradually eroded as the role of environmental factors increases with age. So if the dependent variable were defined as death risk conditional on survival till the age of six months, a relative disadvantage for girls would be likely to show up in more of the regions.

Punjab, Uttar Pradesh and Rajasthan.[19] Children in Muslim households enjoy lower death risks in four states: Andhra, Maharashtra, Uttar Pradesh and Rajasthan. The risk of infant death has declined over time. We estimate that children of mothers who were born before 1960 were between 2 and 7 percentage points more likely to die than children of mothers born after 1970, *ceteris paribus.*

6. Alternative Specifications

In this section we investigate alternative model specifications. First, we investigate whether scarring is driven entirely by birth-spacing (see Section 2). Then we consider whether son-preference shows up in terms of a larger scarring effect following the death of a boy as opposed to a girl. We go on to consider the sensitivity of the estimated scarring effects to the age-composition of mothers, and to the estimator. Refer Table 4, where we present only the scarring coefficients; full results are available from the authors. The basic model presented in Panel [2] Table 2 is the reference point, and the scarring coefficient estimates from this model are displayed again in Column 1 of Table 4. This model takes a long time to converge, primarily because it allows a probability mass at negative infinity. We therefore first estimate the model excluding this mass point. The results from this more restrictive model (with ψ=0) are in Column [2] of Table 4.

The mass point is insignificant in 11 of the 15 states (see Table 3, column 2). Unsurprisingly, the results are insensitive to exclusion of the mass point in these states. In the four states in which the mass point is significant, namely Orissa, Rajasthan, Haryana, and Karnataka, the estimates of scarring in column 2 are a bit larger than in

[19] Scheduled castes (SC) are the lowest caste group in India, so called because of their listing in a schedule appended to the Constitution of India. Scheduled tribes (ST), enumerated in another schedule of the constitution, fall outside the Hindu caste system, but their members are, like the ST, among the poorest in society (Government of India 2001). In India as a whole, SC account for about 18 and ST for about 8 per cent of the population.

column 1, but the differences are not so large as to change our qualitative conclusions. So the rest of this section estimates variants of the main model applying the restriction that there is no mass point.

For the reasons detailed in section 2, short birth spacing is a potential mechanism driving scarring. Since birth-spacing is amenable to policy interventions such as extension of contraception provision, it is useful to confirm this speculation. We do not include the preceding birth interval in the preferred model for two reasons. First, it is endogenous and there is no evident instrument, although see Bhalotra and van Soest (2006), who estimate a more structural model with the restriction that the risk period is neonatal. Second, we are interested in the full impact of scarring and there may be more to scarring than birth-spacing. Refer column 3 of Table 4, which presents estimates obtained after including dummy variables indicating that the length of the preceding interval is less than 18 months, 18-23, 24-29 months, or longer. These categories were chosen by reference to results in the demographic literature which suggests relevant thresholds, and by reference to the distribution of birth intervals in our data. The coefficient on previous sibling's survival status is now smaller. Additionally, in the two states of Punjab and Kerala and to a lesser extent in the three states of West Bengal, Maharashtra and Andhra Pradesh, it seems that birth-spacing is the main causal mechanism in scarring. In the other states, it seems that some other causal processes are at work. Although we do not allow for endogeneity of the birth-interval, this finding is consistent with the findings of Bhalotra and van Soest (2006) for the state of Uttar Pradesh, which are got after accounting for endogeneity. An example of a causal process that may operate independently of birth interval length is maternal depression (see section 2).

We next explore the possibility that gender preference in India shows up in the scarring coefficient. Suppose that the previous child died in infancy, and was a boy. The idea is that, under son-preference, parents will hasten to have another child more quickly than if the lost child were a girl. So, to the extent that scarring reflects short birth-spacing, we should see a larger scarring effect when the previous death is of a boy. This is investigated by interacting the previous child's survival status with a dummy that is one if the previous sibling is a girl. The results are in Table 4. Column 4 shows the base coefficients (for boys), while column 5 shows the coefficient on the interaction terms. The interaction is significant in three of the fifteen states, in each case, with the expected (negative) sign. While Rajasthan and Punjab are known to be regions with particularly high son-preference, the significance of this effect for the North-East is somewhat unexpected.

As discussed earlier, in a dynamic model with unobserved heterogeneity, ignoring left-censoring can bias the equation parameters and, in particular, the scarring coefficient. Intuitively, the history of births and deaths within a family, impacts upon the risk of death of the index child. We have argued that this bias generally goes unrecognised and is easily addressed. To avoid the problem of right-censoring of the mortality data, we have dropped children born in the twelve months before the survey date. So, all children in our sample have full exposure to the risk of infant death.

There remains a possible issue with right-censoring of fertility, that is, not all women have, at the time of the survey, completed their fertility. There is no direct mechanism by which future births of the mother will impact the mortality risk of the index child, but the age-composition of mothers in the sample will be affected by right-censoring. This problem is at least partly addressed by conditioning upon maternal age at birth. It is possible that an additive control is not adequate. However, the direction of

any remaining effects is difficult to pin down because age, cohort and time effects are involved. In particular, the extent of scarring may depend upon calendar time as well as upon the age-composition of mothers in the sample.

As a check on whether the main conclusions of the paper are affected by right-censoring of fertility, we have re-estimated the model restricting the sample to women who are beyond childbearing age (40-49 years) and can be assumed to have completed their fertility. We would expect the scarring coefficient to be different for this sample. If it turned out to be close to zero and insignificant, then we would need to be concerned that failing to correct fully for right-censoring in fertility is driving our main results. To avoid small sample sizes and also to avoid displaying 15 state-specific equations, this check is done after pooling the data for the 15 states and conditioning upon state fixed effects. The estimated scarring coefficient is 0.823. This is significant and, in fact, larger than the coefficient of 0.652 obtained when an identical model is estimated on all mothers. The detailed results are available from the authors on request.

So far we have assumed that family or mother-specific unobservables are captured by a random effect (α_i), and that the covariates (**x**) are strictly exogenous. An alternative that relaxes the assumption of strict exogeneity of **x** is to estimate the model parameters using conditional maximum likelihood (CML), which involves conditioning upon a set of sufficient statistics for the elimination of α_i. A set of sufficient statistics can only be found in the case of a dynamic model under the restrictions that the covariate effects are zero ($\beta=0$), that the error (u_{ij}) distribution is logistic, and there are at least four children per mother ($j>=4$). In this restricted model, it is possible to obtain a consistent estimate of the scarring parameter (Chamberlain 1985; see Narendranathan

and Elias 1993 for an application). [20] Although the CML estimator has the advantage of not requiring an assumption about the distribution of α_i, against it is the fact that it involves a considerable loss of information. For example, only the subset of families who have experienced at least one death contribute to the CMLE. In addition, it also has the disadvantage that it does not yield marginal or average partial effects (see Wooldridge, 2005). We therefore investigate robustness of our estimates to the random effects assumption by including in the model the time average of the time-varying covariates (see Chamberlain, 1984). The only covariate which varies across siblings in our model is the age of the mother at the birth of the index child. We have included the average of this across children, and the square of the average in the model to capture any correlation between the **x** and α_i, The results from this are reported in column 6 of Table 4. There are no qualitative changes to the results.

7. Conclusions

Development progress is now widely measured with reference to the Millennium Development Goals (MDG), one of which is to reduce under-5 mortality by two-thirds by 2015, relative to its level in 1990 (see UNDP 2003). This has resulted in renewed interest in research and policy design in this area.[21] This paper contributes new insights into the determinants of infant mortality.

Across the developing world, where fertility (sibship size) and childhood mortality rates are high, some families experience multiple child deaths while others experience none. In attempting to explain this phenomenon, this paper proposes and

[20] This method is sometimes called 'fixed-effects' method because no assumption regarding the exogeneity status of the regressors **x** is needed. Treating the α_is as fixed effects (i.e. parameters) and estimating the them along with the other parameters of the model gives rise to the incidental parameters problem (Neyman and Scott 1948).

[21] See, for example, various issues of the Lancet (2003) on Child Survival.

investigates the hypothesis that the event of a child death creates a dynamic that makes further children of the same mother more vulnerable to early death. Separating causality from correlation in this area has important implications for policy and for research on the inter-relations of family behaviour, mortality and fertility.

We find a great deal of persistence in mortality within families in each of the fifteen Indian states for which data are analysed. Simple unconditional probabilities show that a child whose previous sibling died in infancy is three to four times as likely to experience infant death as compared with a child whose previous sibling survived. Using data on 223,702 children spread across the 15 major states of India, this paper estimates the size of scarring effects and family-level unobservables across the states. Sizeable scarring effects are identified and these are significant in 13 of the 15 states. In a model that controls comprehensively for inter-family heterogeneity, the relative odds of infant death conditional upon the preceding sibling dying in infancy range between 1.4 and 2.5. The percentage of the observed clustering of sibling deaths that is explained by scarring is large, ranging between 16% in West Bengal and 62% in Haryana. Previous research on death clustering has erroneously tended to equate sibling death clustering with inter-family differences, ignoring these huge intra-family effects. We estimate that eliminating scarring would reduce infant mortality rates among second and higher-order children, by between 3.1% (Punjab) and 10.8% (Madhya Pradesh). In view of the fact that the rate of decline of child mortality in 1990-2001 was 1.1% per annum and, during 1960-90 it was 2.5% per annum (see Black, Morris and Bryce, 2003), these are large potential changes.

Our results highlight the role of short birth-intervals and, possibly, maternal depression as causal processes. The finding of scarring effects suggests a higher payoff to interventions designed to reduce mortality than previously recognised. In the

language of Manski (1995), a social multiplier is activated. In this context, this is because reducing the risk of death of a child automatically implies reducing the risk of death of his or her succeeding siblings.

The pattern of results across the 15 states is complex, indicating that both economic and social development matter. In particular, there is no clear linear association of state-level GDP with either the level of infant mortality or its distribution across families. Scarring effects are weak in Punjab, which is richest, and in Kerala, which is socially the most advanced. Interestingly, the data indicate that Punjab has the highest proportion of births with preceding intervals shorter than 18 months. This is argued to be consistent with the mortality-raising effects of short birth intervals being smaller in wealthier societies, where women are healthier and better able to regenerate the resources needed for the next pregnancy. We find evidence of son-preference in the scarring coefficients for three states. Two of these states, Punjab and Rajasthan, are known to have strong son-preference. Our results also reveal some interesting differences in the effects of standard demographic covariates of infant mortality across the fifteen regions.

References

Aizer, A. and Currie, J. (2004), Networks or neighborhoods? Correlations in the use of publicly-funded maternity care in California", *Journal of Public Economics*, forthcoming.

Altonji, J. G. and Dunn, T. A. (1996), Using siblings to estimate the effect of school quality on wages, *The Review of Economics and Statistics*, Vol. 78, No. 4, 665-671.

Arulampalam, W. and Bhalotra, S. (2006), Sibling death clustering in India: State dependence *vs* unobserved heterogeneity, Journal of the Royal Statistical Society, Series A, Vol. 169, Part 4, pp. 829-848..

Arulampalam, W. and Bhalotra, S. (2004), Inequality in Infant Survival Rates in India: Identification of State-Dependence Effects, Discussion Paper 04/558, Department of Economics, University of Bristol.

Behrman, J. R. and Wolfe, B. L. (1984), The socioeconomic impact of schooling in a developing country, *The Review of Economics and Statistics*, Vol. 66, No. 2, May, pp. 296-303.

Bhalotra, S. and Van Soest, A. (2006) Birth-spacing, fertility and neonatal mortality in India: Dynamics, frailty and fecundity. IZA Discussion Paper 2163, June.

Black, R.E., Morris, S. S. and Bryce, J. (2003), Where and why are 10 million children dying every year?, *Lancet*, 361, 2226-34.

Bolstad, W. M. and Manda, S. O. (2001), Investigating child mortality in Malawi using family and community random effects: A Bayesian analysis, *Journal of the American Statistical Association,* Vol. 96, 12-19.

Bongaarts, J. and Potter, R.G. (1983), *Fertility, Biology and Behaviour: An Analysis of the Proximate Determinants*. New York: Academic Press.

Butcher, K. F. and Case, A. (1994), The effect of sibling sex composition on women's education and earnings, *The Quarterly Journal of Economics*, Vol. 109, No. 3, Aug., pp. 531-563.

Chamberlain, G. (1984), Panel data, in S. Griliches and M. Intriligator, eds., *Handbook of Econometrics*, North-Holland, Amsterdam, 1247-1318.

Claeson, M., E.R. Bos, T. Mawji, and I. Pathmanathan (2000), Reducing child mortality in India in the new millennium, *WHO Bulletin*, Vol.78, no.10, 1192-1199.

Cleland, J. and Sathar, Z. A. (1984), The effect of birth spacing on childhood mortality in Pakistan, *Population Studies*, 38, 401-418.

Curtis, S.L., Diamond, I. and McDonald, J.W. (1993), Birth interval and family effects on post-neonatal mortality in Brazil, *Demography*, 33(1): 33-43.

DasGupta, M. (1990), Death clustering, mothers' education and the determinants of child mortality in Rural Punjab, India, *Population Studies*, 44, 489-505.

DaVanzo, J. and Pebely, A. R. (1993), Maternal depletion and child survival in Guatemala and Malaysia, Labor and Population Program Working Paper 93-18, RAND.

Dreze, J. and Sen, A. (1997), *Indian Development: Selected Regional Perspectives*, Oxford: Clarendon Press.

Firstman, R. and Talan, J. (1997), *The Death of Innocents*. New York: Bantam Books.

Government of India (2003), *Economic Survey: 2002-2003*.

Guo, G. (1993), Use of sibling data to estimate family mortality effects in Guatemala. *Demography*, 30 (1): 15-32.

Heckman, J. J. (1981), The incidental parameters problem and the problem of initial conditions in estimating a discrete time-discrete data stochastic process, in *Structural Analysis of Discrete Data with Econometric Applications*, eds. C. F. Manski and D. McFadden, 114-178, Cambridge: MIT Press.

Heckman, J., V.J. Hotz and J.Walker (1985), New evidence on the timing and spacing of births, *American Economic Review*, 75: 179-84.

Heckman, J. J. and Singer, B. (1984), A method for minimizing the impact of distributional assumption in econometric models for duration data, *Econometrica*, 271-320.

IIPS and ORC Macro (2000), *National Family Health Survey (NFHS-2) 1998-9*: India. Mumbai: International Institute for Population Sciences (IIPS). Calverton: ORC Macro.

Kennedy, K.I., C.M. Visness (1992), Contraceptive efficacy of lactational amenorrhoea, *Lancet*, 339, pp. 227-230.

Lancet (2003), Special issues on Child Survival, volumes 359-363.

Lawless, J. F. (1987) Negative binomial and mixed Poisson regression. *Canadian Journal of Statistics*, 15, 209-25.

Manski, C. F. (1995), *Identification Problems In The Social Sciences*, Cambridge, Mass:, Harvard University Press.

Miller, B. (1981), *The Endangered Sex: Neglect of Female Children in Rural North India*. Ithaca: Cornell University Press.

Moffitt, R. (2004), Policy interventions, low-level equilibria and social interactions, in Durlauf, S. and Young, P. (eds.) *Social Dynamics* (Cambridge MA: MIT Press).

Narendranathan, W. and Elias, P. (1993), Influences of past history on the incidence of youth unemployment: Empirical findings for the UK, *Oxford Bulletin of Economics and Statistics*, 55 (2), 161-185.

Neumark, D. and Korenman, S. (1994), Sources of bias in women's wage equations: results using sibling data, *The Journal of Human Resources*, Vol. 29, No. 2, Spring, 379-405.

Neyman, J. and Scott, E. (1948), Consistent estimates based on partially consistent observations , *Econometrica*, 16, pp. 1-32.

Oettinger, G. S. (2000), Sibling similarity in high school graduation outcomes: causal interdependency or unobserved heterogeneity?, *Southern Economic Journal*, 66(3), 631-648.

Pinker, S. (2002), The Blank Slate: The modern denial of human nature, Allen Lane: The Penguin Press.

Preston, S. H. (1985), Mortality in childhood: lessons from WFS. In: Cleland, J.G. and Hobcraft, J., editors. Reproductive Change in Developing Countries. Oxford: Oxford University Press, 46-59.

Rahman, A., Iqbal, Z., Bunn, J., Lovel, H. and Harrington, R. (2004) Impact of maternal depression on infant nutritional status and illness: A cohort study. *Archives of General Psychiatry*, 61: 946-952.

Rawlings, J.S., V. B. Rawlings and J.A. Read (1995), Prevalence of low birth weight and preterm delivery in relation to the interval between pregnancies among white and black women, *The New England Journal of Medicine*, Volume 332(2), January 12.

Rosenzweig, M. R. and Wolpin, K.I. (1994), Are there increasing returns to the intergenerational production of human capital? maternal schooling and child intellectual achievement, *The Journal of Human Resources*, Vol. 29, No. 2, Spring, 670-693.

Sastry, N. (1997), Family-level clustering of childhood mortality risk in Northeast Brazil, *Population Studies*, Vol. 51, Issue 3, 245-261.

Solon, G., Corcoran, M., Gordon, R., and Laren, D. (1991), A longitudinal analysis of sibling correlations in economic status, *Journal of Human Resources*, 26, 509-534.

Stata 8 (2004), *Stata Statistical Software,* Stata Corportation.

Steer, R. A., Scholl, T.O., Hediger, M. L. and Fischer, R. L. (1992), Self-reported depression and negative pregnancy outcomes, *Journal of Clinical Epidemiology*, 45(10): 1093-9.

Stephansson, O., P. Dickman and S. Cnattingius (2003), The influence of interpregnancy interval on the subsequent risk of stillbirth and early neonatal death, Obstet Gynaecology, 102: 101-108.

Strauss, J. and Thomas, D. (1995), Human resources: Empirical modelling of household and family decisions, Chapter 34 in Jere Behrman and T.N. Srinivasan (eds.), *Handbook of Development Economics*, Volume III, Amsterdam. North-Holland Publishing Company.

UNDP (2003), Human Development Report: *Millennium Development Goals: A Compact Among Nations To End Human Poverty*.

Wooldridge, J. M. (2002), Econometric Analysis of Cross Section and Panel Data, Cambridge, Massachusetts: The MIT Press.

Wooldridge, J. M. (2005), Simple solution to the initial conditions problem in dynamic, nonlinear panel data models with unobserved heterogeneity. *Journal of Applied Econometrics*, 20, pp.30-54.

World Bank (2004), Attaining the Millennium Development Goals in India: How likely and what will it take to reduce infant mortality, child malnutrition, gender disparities and hunger-poverty and to increase school enrolment and completion?, Human Development Unit, South Asia Region, April.

Zenger, E. (1993), Siblings' neonatal mortality risks and birth spacing in Bangladesh. *Demography*, 30(3): 477- 88.

Table 1 –Probabilities Of Infant Death

	Raw Data					Preferred Model
STATE	Probability of Death [1]	Probability of death given previous sibling's **death** [2]	Probability of death given previous sibling's **survival** [3]	Death clustering [2]-[3] [4]	Relative Odds Ratio [5][1]	Relative Odds Ratio [6]
Central						
Madhya Pradesh	0.113	0.223	0.085	0.138	3.09	2.05
Uttar Pradesh	0.116	0.241	0.092	0.150	3.15	1.94
East						
Orissa	0.105	0.226	0.083	0.143	3.22	1.89
Bihar	0.080	0.240	0.061	0.178	4.83	2.36
West Bengal	0.076	0.194	0.060	0.134	3.79	1.54
North						
Rajasthan	0.100	0.211	0.080	0.131	3.06	1.94
Haryana	0.066	0.202	0.053	0.149	4.56	2.50
Punjab	0.060	0.143	0.055	0.088	2.85	1.43
West						
Gujarat	0.085	0.187	0.070	0.117	3.07	1.97
Maharashtra	0.059	0.138	0.048	0.090	3.15	1.66
South						
Andhra Pradesh	0.092	0.190	0.075	0.115	2.89	1.43
Karnataka	0.076	0.190	0.062	0.128	3.57	1.58
Tamil Nadu	0.071	0.160	0.060	0.099	2.96	2.11
Kerala	0.036	0.125	0.029	0.096	4.78	1.99
North-East	0.061	0.166	0.052	0.114	3.64	1.71

Notes: All probabilities are for death in *infancy*. [1]This is calculated as the ratio of column [2]/(1-column[2)] to column [3]/(1-column[3]). This is the exponential of the estimated scarring coefficient in a simple logit model that has an intercept and the survival status of the previous sibling. Column [6] reports the equivalent numbers from the preferred model estimates in Panel [2] of Table 2 which controls for the effects of other covariates and unobserved family specific effects.

Table 2: Results from the Random Effects Logit Regressions

		Panel 1: MODEL WITHOUT UNOBSERVED HETEROGENEITY			Panel 2: MODEL WITH UNOBSERVED HETEROGENEITY			
STATE	Raw data death clustering[2] [1]	Estimated Marginal Effects[3] [2]	Raw Clustering Explained by Scarring [2]/[1] % [4]	Reduction in infant mortality if no scarring % [5]	Estimated Marginal Effects[3] [6]	Raw Clustering Explained by Scarring [2]/[1] % [7]	Reduction in infant mortality if no scarring % [8]	Estimated intra-family correlation coefficient, ρ [9]
Central								
Madhya Pradesh	0.138	0.117	84.4	10.76	0.067	48.6	10.77	0.084
Uttar Pradesh	0.150	0.125	83.3	11.68	0.063	42.0	9.97	0.104
East								
Orissa	0.143	0.123	86.2	10.53	0.065	45.5	8.64	0.077
Bihar	0.178	0.157	87.9	13.75	0.062	34.8	9.59	0.148
West Bengal	0.134	0.097	72.5	9.88	0.022	16.4	5.39	0.212
North								
Rajasthan	0.131	0.112	85.3	10.19	0.067	51.2	8.67	0.055
Haryana	0.149	0.127	85.0	11.52	0.092	61.8	9.08	**0.019**
Punjab	0.088	0.064	72.2	6.05	0.022	25.0	3.07	**0.082**
West								
Gujarat	0.117	0.100	85.0	9.05	0.051	43.6	8.53	0.105
Maharashtra	0.090	0.060	66.9	6.28	0.026	28.9	5.10	0.087
South								
Andhra Pradesh	0.115	0.086	75.0	7.97	0.026	22.6	4.30	0.119
Karnataka	0.128	0.103	80.4	9.42	0.033	25.8	4.65	0.152
Tamil Nadu	0.099	0.074	74.8	6.78	0.056	56.6	8.79	**0.045**
Kerala	0.096	0.059	61.7	6.18	0.028	29.2	4.99	**0.072**
North-East	0.114	0.093	81.2	8.36	0.028	24.6	4.84	0.174

Notes:

1. The equation is jointly estimated with the initial condition of the process (see section 3). The marginal effect associated with scarring is significant at the 5% level in all states, except in Kerala where it is significant at 9% and Punjab, where it is significant at 13%. For four states, the residual variance attributable to unobserved heterogeneity is insignificant, and these are indicated by bold coefficients in the final column.
2. This is Column [4] Table 1.
2. The marginal effect is computed as the difference between the sample averages of the probability of death predicted by the estimated model when $y_{ij-1}=0$ and when $y_{ij-1}=1$ (excluding the first born). This is approximately equivalent to the first partial derivative of the conditional probability of death of the index child with respect to the covariate, y_{ij-1}.
3. This is calculated as the difference between the predicted probability of death from the estimated model, and the predicted probability of death from the model when $\gamma=0$ is imposed after estimation, and excluding first- born children.

Table 3: Model Diagnostics [P-values]

STATE	θ [1]	Estimated mass point at minus infinity $\psi/(1+\psi)$ [2]
Central		
Madhya Pradesh	0.399 [0.169]	0.025 [0.562]
Uttar Pradesh	0.695 [0.000]	0.000 [0.955]
East		
Orissa	0.723 [0.139]	0.145 [0.051]
Bihar	1.060 [0.000]	0.082 [0.300]
West Bengal	0.944 [0.001]	0.000 [0.396]
North		
Rajasthan	0.258 [0.640]	0.132 [0.023]
Haryana	-0.341 [0.942]	0.325 [0.001]
Punjab	1.288 [0.225]	0.131 [0.524]
West		
Gujarat	0.901 [0.046]	0.000 [0.991]
Maharashtra	0.714 [0.223]	0.051 [0.624]
South		
Andhra Pradesh	1.300 [0.010]	0.055 [0.569]
Karnataka	0.445 [0.139]	0.216 [0.008]
Tamil Nadu	0.263 [0.864]	0.043 [0.793]
Kerala	1.795 [0.541]	0.153 [0.639]
North-East	0.775 [0.000]	0.174 [0.404]

Notes: θ is defined in section 5 and $\psi/(1+\psi)$ is defined in section 6.

Table 4: Estimated Scarring Coefficients (γ) from Model Extensions

STATE	Basic Model allowing for mass point at -∞ (Table 2: Column [6])	Exclude mass point at -∞	Add dummies for preceding birth intervals	Allow previous child's survival status to be gender-specific: Baseline effects	Allow previous child's survival status to be gender-specific: Interaction effects	Add mother's age at birth of index child averaged over all children, and its square
	MODEL 1 [1]	MODEL 2 [2]	MODEL 3 [3]	MODEL 4 [4]	MODEL 4 [5]	MODEL5 [6]
Central						
Madhya Pradesh	0.719**	0.732**	0.561**	0.751**	-0.043	0.702**
Uttar Pradesh	0.662**	0.663**	0.530**	0.629**	0.072	0.671**
East						
Orissa	0.639**	0.671**	0.489**	0.743**	-0.169	0.658**
Bihar	0.859**	0.862**	0.705**	0.931**	-0.145	0.834**
West Bengal	0.429**	0.407**	0.304*	0.324	0.188	0.335**
North						
Rajasthan	0.663**	0.691**	0.516**	0.799**	-0.257**	0.710**
Haryana	0.916**	0.987**	0.844**	1.147**	-0.350	0.918**
Punjab	0.360	0.348	0.167	0.670**	-0.727**	0.197
West						
Gujarat	0.676**	0.681**	0.587**	0.591**	0.195	0.633**
Maharashtra	0.506**	0.542**	0.323*	0.393*	0.319	0.439**
South						
Andhra Pradesh	0.360**	0.373**	0.270*	0.300	0.143	0.292*
Karnataka	0.457**	0.482**	0.384**	0.461**	0.046	0.460**
Tamil Nadu	0.746**	0.748**	0.639**	0.825**	-0.175	0.663**
Kerala	0.687*	0.719*	0.592	0.324	0.814	0.683
North-East	0.535**	0.531**	0.391**	0.715**	-0.463**	0.515**

Notes: The equation is jointly estimated with the initial condition of the process (see section 3). **, * indicate that the scarring coefficient γ was significant at 5 and 10% significance level. Models 3-5 are extensions of Model 2. Model 3 includes binary indicators for preceding birth intervals for less than 18 months, 18 to 23 months and 24 to 29 months. Model 4 includes an interaction term between a dummy indicating that the previous child was a girl, and the previous child's survival status. Results are reported in columns 4 and 5. Model 5 investigates the estimator.

Table 5 – Marginal Effects of the Covariates

	MadhP	UttP	Orissa	Bihar	WBeng	Rajast	Haryan	Punjab	Gujar	Mahar	AndhP	Karnat	TamilN	Kerala	NE
All children															
Female	-0.0069	0.0036	-0.0061	0.0071	**-0.0149**	0.0030	0.0089	**0.0281**	-0.0116	-0.0004	-0.0021	-0.0035	-0.0030	-0.0131	**-0.008**
Ma's year of birth – after 1970 (base)															
pre 1959	**0.0428**	**0.0678**	**0.0354**	**0.0184**	**0.0640**	**0.0450**	**0.0316**	**0.0196**	**0.0407**	**0.0277**	**0.0483**	**0.0359**	**0.0198**	**0.0527**	**0.020**
1960-1969	**0.0193**	**0.0258**	0.0037	**0.0159**	0.0167	**0.0238**	0.0151	0.0102	0.0130	0.0123	**0.0258**	**0.0260**	**0.0017**	**0.0385**	0.008
Religion – Hindu (base)															
Muslim	-0.0113	**-0.0323**	-0.0202	-0.0038	-0.0032	**-0.0230**	-0.0122	-0.0243	-0.0125	**-0.0206**	**-0.0447**	-0.0161	-0.0172	0.0018	0.009
Other	0.0013	-0.0306	0.0058	-0.0055	0.0016	**-0.0704**	-0.0043	-0.0081	-0.0674	-0.0023	-0.0297	-0.0068	-0.0135	0.0145	0.002
Ethnicity- other (base)															
Scheduled caste/tribe	0.0127	**0.0120**	-0.0099	0.0111	0.0116	**0.0078**	-0.0040	**0.0191**	-0.0013	0.0032	0.0153	-0.0047	-0.0384	0.0180	-0.008
Other backward caste	0.0179	**0.0180**	-0.0094	0.0129	0.0245	-0.0115	-0.0028	0.0161	0.0087	0.0052	0.0078	-0.0149	-0.0328	0.0055	0.003
Ma's educ: none(base)															
Incomplete primary	-0.0108	**-0.0186**	-0.0081	**-0.0319**	-0.0054	0.0092	0.0178	0.0161	-0.0110	-0.0027	-0.0193	-0.0027	-0.0026	**-0.0227**	**-0.009**
Complete primary	-0.0207	-0.0202	-0.0018	-0.0135	**-0.0492**	-0.0072	-0.0210	**-0.0230**	-0.0183	0.0054	0.0029	-0.0219	-0.0054	-0.0089	-0.010
Incomplete secondary	-0.0061	-0.0181	-0.0190	-0.0146	**-0.0644**	-0.0224	-0.0209	0.0168	**-0.0311**	**-0.0162**	**-0.0434**	**-0.0266**	-0.0017	**-0.0196**	**-0.013**
Comp sec or higher	**-0.0891**	**-0.0749**	-0.0525	**-0.0337**	-0.0379	**-0.0369**	-0.0058	-0.0127	**-0.0488**	-0.0002	-0.0005	-0.0110	-0.0070	**-0.0347**	-0.003
Pa's educ: none (base)															
Incomplete primary	0.0036	0.0167	-0.0060	0.0043	-0.0070	0.0094	-0.0248	0.0123	-0.0121	0.0072	-0.0125	-0.0175	-0.0040	0.0044	-0.005
Complete primary	-0.0035	**-0.0186**	-0.0060	-0.0055	0.0089	-0.0039	**-0.0295**	-0.0214	**-0.0265**	0.0061	0.0041	**-0.0424**	-0.0098	-0.0017	-0.013
Incomplete secondary	-0.0083	**-0.0173**	**-0.0289**	-0.0089	-0.0041	-0.0089	**-0.0201**	-0.0025	-0.0063	-0.0050	-0.0203	**-0.0373**	-0.0115	-0.0095	**-0.014**
Complete secondary	-0.0052	-0.0155	-0.0252	**-0.0233**	-0.0215	-0.0165	**-0.0249**	-0.0094	-0.0252	**-0.0389**	**-0.0579**	**-0.0347**	**-0.0450**	-0.0177	**-0.017**
Higher than secondary	0.0058	-0.0155	**-0.0736**	-0.0105	**-0.0434**	-0.0139	-0.0270	-0.0140	-0.0122	**-0.0445**	**-0.0401**	-0.0225	**-0.0393**	-0.0080	**-0.032**
Mother's age = 15 yrs	**-0.0147**	**-0.0146**	**-0.0096**	**-0.0129**	**-0.0207**	**-0.0096**	**-0.0116**	**-0.0109**	**-0.0110**	**-0.0105**	**-0.0166**	**-0.0133**	**-0.0116**	**-0.0076**	**-0.009**
Mother's age = 25 yrs	**-0.0079**	**-0.0079**	**-0.0044**	**-0.0049**	**0.0003**	**-0.0058**	**-0.0050**	**-0.0059**	**-0.0048**	**-0.0049**	**-0.0054**	**-0.0052**	**-0.0041**	**-0.0013**	**-0.004**
Multiple birth	**0.2639**	**0.2127**	**0.2250**	**0.1880**	**0.1676**	**0.1833**	**0.1412**	**0.1035**	**0.1963**	**0.1403**	**0.2049**	**0.2075**	**0.1420**	**0.0782**	**0.120**
Birth order 3	0.0046	0.0034	-0.0025	0.0024	**0.0197**	-0.0019	-0.0075	0.0098	-0.0097	-0.0062	0.0113	**0.0251**	0.0041	**-0.0198**	0.000
Birth order 4	**0.0168**	**0.0217**	-0.0078	**0.0170**	**0.0209**	-0.0049	0.0156	**0.0349**	-0.0081	0.0090	**0.0242**	**0.0247**	**0.0261**	-0.0078	**0.017**
Birth order 5	**0.0291**	**0.0312**	-0.0055	**0.0353**	**0.0225**	0.0124	0.0019	**0.0304**	0.0122	**0.0255**	**0.0314**	0.0182	0.0222	-0.0065	**0.019**
Birth order >5	**0.0541**	**0.0559**	0.0114	**0.0426**	0.0196	**0.0305**	**0.0351**	**0.0719**	0.0067	0.0174	0.0188	**0.0411**	**0.0335**	0.0223	**0.030**

Notes: All variables except mother's age are binary indicator variables. Mother's age was entered as a quadratic and has been solved for two high-frequency ages. The figures in bold indicate coefficients significant at 10% or less. The model also included interactions of all regressors with a dummy for the first born child (not shown). The model for the North Eastern (NE) region included binary indicators for the individual states in this region.

Appendix Table A1 – Sample Descriptive Statistics

STATE	Number of Mothers [1]	Number of Children [2]	Number of infant deaths [3]	% Infant deaths among under 5 deaths [4]	% births with preceding birth interval<18 months [5]	% births with preceding birth interval 18-23 months [6]	% births with preceding birth interval>23 months [7]	Total fertility rate, age 15-49: 1996-1998 [8]	Mother's age at first birth in years [9]
Central									
Madhya Pradesh	5543	21403	113	67.5	17.6	19.6	62.6	2.61	15.3
Uttar Pradesh	7297	29937	116	73.1	18.1	18.6	63.1	2.88	15.7
East									
Orissa	3655	11722	105	78.1	14.8	17.3	67.7	2.19	16.8
Bihar	5629	21374	80	67.9	13.9	19.8	67.2	2.75	15.8
West Bengal	3606	10627	76	77.3	14.4	18.5	66.8	1.69	16.2
North									
Rajasthan	5424	20774	100	70.6	17.3	21.0	61.6	2.98	15.9
Haryana	2436	8105	66	72.7	16.4	21.1	62.4	2.24	17.3
Punjab	2390	7211	60	79.9	20.8	20.5	58.3	1.79	19.1
West									
Gujarat	3192	10326	85	73.7	15.7	22.0	62.1	2.33	17.1
Maharashtra	4283	12881	59	73.5	14.0	20.1	65.6	2.24	16.4
South									
Andhra Pradesh	3233	10129	92	78.4	16.7	19.6	63.4	2.07	15.0
Karnataka	3472	11174	76	71.0	13.4	23.4	63.0	1.89	16.0
Tamil Nadu	3870	10405	71	73.4	15.7	18.4	65.6	2.11	17.6
Kerala	2340	5950	36	79.6	15.0	17.5	67.2	1.51	18.9
North-East	9370	31684	61	73.1	14.5	20.3	65.1	2.08	18.1

Appendix Table A1 – Continued

STATE	Religion: Hindu % [10]	Caste: Scheduled caste/tribe % [11]	Mother's education – none % [12]	Mother's education - secondary or higher % [13]	Father's education – none % [14]	Father's education – secondary or higher % [15]	% with no Electricity [16]	% Female children [17]	Rank of state in per capita income* [18]
Central									
Madhya Pradesh	91.2	38.3	71.5	5.8	34.4	18.7	30.1	47.9	12
Uttar Pradesh	82.3	21.7	75.5	6.3	33.8	29.1	63.6	47.5	13
East									
Orissa	95.5	39.8	60.2	5.1	33.4	16.4	60.6	48.3	14
Bihar	81.3	28.5	81.2	4.9	46.2	25.5	82.8	47.9	15
West Bengal	72.7	29.0	50.3	8.7	31.2	19.1	57.1	48.5	9
North									
Rajasthan	88.1	33.2	80.9	4.0	40.5	22.5	37.3	47.8	11
Haryana	88.2	23.0	66.6	12.7	34.2	35.0	12.1	46.0	3
Punjab	43.1	31.1	46.5	22.4	27.1	37.4	3.9	45.6	1
West									
Gujarat	89.8	38.0	56.3	12.1	26.2	25.4	16.5	48.1	4
Maharashtra	73.8	22.6	41.6	14.5	19.8	30.7	13.9	47.9	2
South									
Andhra Pradesh	85.5	26.2	67.6	7.1	47.0	18.7	24.1	48.2	8
Karnataka	83.3	24.7	60.5	10.9	39.9	22.7	19.4	48.7	7
Tamil Nadu	87.2	26.5	40.5	11.9	22.1	23.8	18.4	48.4	5
Kerala	47.3	9.8	11.4	27.7	7.8	27.0	27.9	48.1	6
North-East	44.3	56.8	46.4	8.7	27.6	20.0	44.1	48.2	10

*Data from Government of India (2003): Economic Survey 2002-3, Table 1.8: per capita net state domestic product. For the North-Eastern region, the rank is based on an unweighted average of the figures for each of the individual states. Other statistics are from the NFHS.

Map of India Showing the Major States