Inequality in Infant Survival Rates in India: Identification of State-Dependence Effects

Wiji Arulampalam (Warwick)
Sonia Bhalotra (Bristol)

Discussion Paper No. 04/558

May 2004

Department of Economics
University of Bristol
8 Woodland Road
Bristol BS8 1TN
UK

Inequality in Infant Survival Rates in India: Identification of State-Dependence Effects

Wiji Arulampalam (Warwick)[+] *and Sonia Bhalotra (Bristol)*[++]

17 May 2004

Abstract

Data from a number of regions indicate that childhood deaths are unequally distributed across families. This has been identified, in previous research, with (observed and unobserved) heterogeneity between families. In this paper, we investigate whether, on top of these correlated risks, there is a causal process at work within families, whereby the death of a child elevates the risk of death of the succeeding sibling. Borrowing language from the unemployment literature, the causal process is termed state dependence or scarring. To the extent that scarring exists, a social multiplier comes into play, raising the payoff to policies that reduce infant mortality. Acknowledging scarring effects is also potentially relevant to understanding the relation of mortality and fertility behaviour within families. The analysis is conducted using data for the 15 major states of India. Large scarring effects are observed in 14 of the 15 states.

Keywords: Death clustering, infant mortality, state dependence, scarring, unobserved heterogeneity, dynamic random effects logit, India.

JEL classification: J1, C1, I1, O1

[+] Department of Economics, University of Warwick, UK. Also IZA, Bonn, Germany.

[++] Department of Economics, University of Bristol, UK. Also Research Associate, CSAE (Oxford), QEH (Oxford), CHILD (Turino).

Correspondence: Department of Economics, University of Bristol, 8 Woodland Road, Bristol BS8 1TN, U.K. Tel.: +44 (0)117 9288418; Email: s.bhalotra@bristol.ac.uk.

Inequality in Infant Survival Rates in India: Identification of State-Dependence Effects[♣]

Wiji Arulampalam and Sonia Bhalotra

Introduction

Development progress is now widely measured with reference to the Millennium Development Goals that grew out of the resolutions of world conferences organised by the United Nations in the year 2000. Selected indicators were identified that complement the formerly popular measure of income poverty (UNDP 2003). One of the goals is to reduce under-5 mortality by two-thirds by 2015, relative to its level in 1990. This has resulted in renewed interest in research and policy design in this area.[1] Since most under-5 deaths occur in infancy (70% in South Asia), this paper concentrates on infant deaths. A great deal of previous research has studied determinants of the level of infant mortality, and also regional and gender inequalities in its incidence. Here we focus on inequality in the distribution of deaths between families. The family is an important institution and both its inherent traits and its behavioural choices impact upon infant survival chances. Yet, analysis of the inter-family distribution of childhood deaths has been relatively limited. This paper uses methods familiar in labour economics to analyse the problem of death clustering amongst siblings that has interested demographers. The rest of this Section presents a quick overview of related research in demography and in economics, highlighting the contributions made in this paper.

Using data from a wide spread of regions in developing countries, recent research in demography has noted that a small fraction of families accounts for most child deaths (e.g. DasGupta, 1990; Curtis *et al*, 1993; Guo, 1993; Zenger, 1993; Sastry,

[♣] We are grateful to Arthur van Soest for many helpful comments.

1997). In this research, the clustering of deaths amongst siblings has been identified with inter-family heterogeneity. In other words, it is explained in terms of family-level characteristics common to siblings, such as maternal education (observable) and genes (unobservable), which generate a positive *correlation* of sibling mortality risks. A contribution of our research is to argue that, in addition to inter-family heterogeneity, a *causal* process may be at work, whereby the actual event of death of a child influences the risk of death of the succeeding child in the family. This is what we shall refer to as state dependence. This paper investigates the extent to which state dependence effects can explain multiple child deaths within a family. The existence of state dependence raises the returns to policy interventions that target the reduction of infant mortality; in the language of Manski (1995), a social multiplier is activated. 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 distinction pursued in this paper, between causality and correlation, is therefore important in motivating policy action.[2] Although this paper does not attempt to identify the precise mechanism driving scarring, consideration of the possible mechanisms (Section 2) shows that recognizing scarring is also important in understanding the nexus of mortality, fertility and family behaviour.

Economists have not, as yet, investigated the clustering of deaths within families, but they have exhibited interest in the observed correlation of other sibling outcomes. Like demographers, economists have tended to identify this with inter-family heterogeneity. For example, Solon *et al* (1991) study sibling correlations in economic status to assess

---

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

2 It might be clear to some that policies designed to reduce infant mortality need no further motivation. However economists and funding agencies like the World Bank and the World Health Organisation are currently interested in estimating the impact of mortality on GDP and, thereby, in estimating the cost-effectiveness of policy in this area. See, for example, Kirigia *et al* (2004).

the extent to which family background matters. Their research is motivated by the idea that, the more important is family background, the stronger is the intergenerational transmission of inequality and, hence, the stronger the case for government intervention that aims to equalize opportunity amongst children. In other research on child outcomes in economics, psychology and medicine, data on siblings have been used to difference out unobservable elements of family background (like ability or frailty), so as to identify behavioural effects (see, for example, Behrman and Wolfe, 1984; Neumark and Korenman, 1994; Rosenzweig and Wolpin, 1994, 1995; Altonji and Dunn, 1996a, b).[3] Again, the similarity of sibling outcomes has, implicitly or explicitly, been attributed to shared family traits.

A notable exception is a recent study by Oettinger (2000), which attempts to identify causal effects of an individual's educational attainment on the educational attainment of his or her sibling after allowing for shared (and unobserved) traits amongst siblings. In this respect, it is similar to the analysis in this paper. 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.

A closer parallel to this research, concerned with the disentanglement of causality and correlation in outcomes, appears in the literature on social networks and neighbourhood effects. It is commonly observed that people who share residential location, race or ethnicity have correlated outcomes. These are often associated with exogenous effects

---

[3] 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

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 except that, here, the group is defined to be 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. Borrowing language from the economics of unemployment, the endogenous effect is called state dependence or scarring (see, for example, Heckman 1981).[4] In the unemployment literature, several studies have sought to identify state dependence after controlling for unobserved heterogeneity. The observation of interest is that certain people are unemployed period after period while others are seldom unemployed: there is a "clustering" or concentration of unemployment spells amongst certain individuals. While a natural explanation of this is

studied in this way include school attainment or achievement, birth weight and foetal growth, the returns to education, wages and socio-economic status

[4] It may be useful to clarify the language for the current context. The idea is that the event of 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 mortality risk facing a child being dependent upon the state (died in infancy or not) revealed for the previous child in the family. Since time is implicit in the sequencing of children, models that include the previous child's survival status are analogous to dynamic models.

that these individuals are different from those more successful in the labour market in observed or unobserved ways [heterogeneity], an alternative explanation of interest is that being unemployed in one period causes a higher risk of unemployment in the subsequent period [state dependence]. The structure of our problem is similar to this, except that successive periods of time are replaced by successive siblings.

Application of these ideas to the question of sibling death clustering enjoys two advantages. First, a common problem in the unemployment literature is that it is difficult to disentangle duration and state dependence effects because the same unemployment spell can extend over two consecutive observation periods, and unemployment is customarily defined as the state occupied at the time of the interview. This problem does not arise in the current application because infant mortality is a discrete event, and the data are retrospective. A second problem faced in studies of unemployment is that most available data do not contain information that date back to the initial exposure of the individual to the labour market. This creates what is called an initial conditions problem (e.g. Heckman 1981). The data used in this paper have information on the first birth of every mother and this is a relatively straightforward way of overcoming the initial conditions problem. The estimation methods used signify a break from previous demographic research in this area (see Arulampalam and Bhalotra, 2004).

The analysis is conducted for each of the 15 major states of India. The high rates of childhood mortality in India and the number of affected children make this an important place to locate the investigation. The Indian data offer vast samples and the remarkable socio-economic diversity within India makes it interesting to compare results across the

15 regions. The main result is that large scarring effects are identified in 14 of the 15 regions.

In order to illustrate the nature of causal effects and to further motivate the analysis, Section 2 outlines mechanisms that might explain scarring. The data are described in Section 3, where descriptive statistics that show a remarkable degree of death clustering amongst siblings are presented. Section 4 sets out the econometric model and discusses estimation issues, indicating how common biases in parameter estimators may be avoided. The variables used in the analyses are described in Section 5 and the results are presented in section 6. Section 7 concludes.

2. Scarring Mechanisms

So as to clarify further the idea of state dependence or scarring as a causal process, it is useful to consider what sorts of mechanisms might be involved. For example, the process might operate by an infant death shortening the time to the next birth because the mother stops breastfeeding and, thereby, is able to conceive sooner than otherwise (e.g. Bongaarts and Potter, 1983). A new pregnancy requires replenishment of vital nutrients like calcium and iron that are needed to support foetal development (e.g. daVanzo and Pebley, 1993). This problem is likely to be more acute in poor societies where bio-availability of these nutrients from staples like cereal is low, and nutrient losses associated with infections challenge the capacity of women to produce healthy children. As it can, in these circumstances, take up to 24 months for the mother to recuperate physiologically from a birth, a short preceding birth interval for the index child elevates this child's mortality risk. Henceforth, this is referred to as *the fecundity hypothesis*. An alternative argument that has been suggested in the demographic literature is that the death of a child leads parents to (consciously) conceive sooner in a desire to "replace" their loss (e.g. Preston, 1985). This is the *replacement hypothesis*. A

further possibility, hitherto unrecognised in the demographic literature, 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 (see, for example, Steer *et al*, 1992). This is referred to here as the *depression hypothesis.* The depression argument is empirically distinguishable from the other two hypotheses because it does not work *via* the birth interval.

This distinction is interesting because the birth interval is an aspect of fertility that is responsive to intervention, in particular, interventions that improve access to or uptake of contraception. A vast literature in demography demonstrates that short preceding birth intervals raise mortality risk. A weakness of the demographic literature is that it neglects to allow for the fact that the birth interval is a choice variable (e.g. Cleland and Sathar, 1984; Koenig *et al,* 1990). Conscious of its potential endogeneity, we do not include the birth interval as a regressor in the mortality equations estimated in this paper. This is consistent with the objective of this paper, which is to identify the extent of scarring, if any. In work in progress that attempts to shed light on the mechanism driving scarring, the birth interval is modelled jointly with mortality risk.[5]

It is plausible that there are learning effects, which result in the mortality risk of the index child *falling* on account of the death of the preceding sibling. For instance, if the older sibling died of diarrhoea, the mother may rush to learn how to prevent diarrhoea-related infant death. Any positive degree of scarring that is identified is then net of this type of learning effect.

[5] The analyses of death clustering that we refer to from the demographic literature 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

3. Data

India makes an interesting laboratory for the study of demographic processes. One in six of the world's people lives in India and almost a quarter of the under-5 child deaths in the world occur in India (Black *et al,* 2003). In the year 1998-99, the child death rate in India was 95 (per 1000 live births), and the infant death rate was 68 (World Bank, 2004). The infant death rate implies that nearly 1¾ million children die in India before their first birthday. Despite the high levels of mortality in India and the availability of relevant data, there has not been much research on childhood mortality in this region (see World Bank, 2004).

Infant mortality has been declining in India, having halved between the early-1970s and 2000, but this rate of decline is somewhat less impressive than that seen in some other South and South-east Asian countries, including Bangladesh. The millennium development goal (MDG) translates to achievement of an infant mortality rate of 27 (and an under-5 mortality rate of 32) by 2015. Simulations suggest that attaining the MDG will be challenging but is possible (World Bank, 2004). The phenomenon investigated in this paper is relevant to policy analyses of this type. If scarring effects exist then a given mortality reduction will be attained more rapidly than implied by simulation of a model in which scarring effects are not modelled. This is because of the social multiplier effect referred to in Section 1.

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. The analysis is conducted for each of the 15 major states of India.[6] This is useful

implicated in a number of cases, especially in the UK. Media coverage of these events is consistent with a depression story.

[6] The states are listed in Table 1. The eight small North-Eastern states, which have fairly similar levels of mortality, were aggregated to create the North-East region (or the "fifteenth state"). The resulting diversity within this region is expected to be no greater than that in a

since it would be quite remarkable if we were to identify scarring effects in 15 different data samples. Also, as the states exhibit contrasting social, demographic and economic characteristics (see Dreze and Sen, 1997), the analysis provides some preliminary insight into the way in which scarring varies with socio-demographic and economic development.

The data employed in the analysis are drawn from the Indian National Family Health Survey (NFHS) of 1998-99, which interviewed 90000 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. The NFHS is one of a series of fairly comparable Demographic Health Surveys (DHS), available for about sixty-nine low and middle-income countries. This implies that the ideas and methods introduced in this paper are immediately applicable to other regions. The DHS surveys are freely available to researchers at www.measuredhs.com.

Table 1 reports mortality rates by state computed from the NFHS. These are averages over the data sample, which includes children born across the four decades, 1961-1999. 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.[7] In our data, of every 1000 children born in India, 82 die in infancy. There is remarkable variation across the Indian states. For example, the large backward state of Uttar Pradesh (UP) in Central India has a mortality rate (in 1000) of 116, while the Southern state of Kerala, known for its relative success in human development, has a rate of 36. The relative position of the states with respect to relevant indicators is

large state like Uttar Pradesh. The smaller states/union territories of Delhi, Himachal Pradesh, Goa and Jammu were excluded from the analysis due to the small cell sizes.

[7] 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.

discussed in the Appendix, and Appendix Table A1 presents some illustrative descriptive statistics.

Death clustering

Columns [2] and [3] of Table 1 show the raw data probability of infant death conditional on infant death and survival, respectively, of the preceding sibling. The difference of these conditional probabilities (Column [4]) is a measure of the extent of death clustering. The increase in the probability of death conditional on the previous child dying as opposed to surviving infancy ranges from about 0.09 in Punjab and Maharashtra, to 0.18 in Bihar. These are enormous increases, given an average mortality rate of 0.082 in India. While Bihar is, like UP, both socially and economically backward, it is economic rather than social progressivity that distinguishes Punjab and Maharashtra from the other states. Thus while the overall incidence (or level) of infant mortality appears to be more strongly influenced by socio-demographic development, the raw data indicate that clustering (or the distribution of mortality risk across families) is more strongly correlated with economic development. An alternative way of presenting the data is in Column [5]. This shows that the odds of a child dying in infancy if the previous sibling died rather than survived infancy are anywhere between 2.9 (as in Punjab and Andhra Pradesh) and 4.8 (as in Bihar and Kerala). These numbers indicate how dramatically the risk multiplies. The large odds for Kerala are striking. As we shall see, controlling for inter-family heterogeneity results in a considerable reduction in the odds ratio for this state (Section 6).

Overall, both Columns [4] and [5] indicate a remarkable degree of death clustering. These, however, are simply the observed tendencies in the data. Estimation of the statistical model discussed below will allow us to 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 (scarring).

4. Methods

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

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

where $\boldsymbol{x}$ is a vector of strictly exogenous observable child and family specific characteristics that influence y_{ij}^* and $\boldsymbol{\beta}$ 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 $\gamma=0$.[8] 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*.[9] A model restricted to first-order effects is consistent with the mechanisms that we suggest might drive scarring (Section 2).

Since (1) is a recursive model, it faces the initial conditions problem (e.g.

[8] The estimated parameter γ should be interpreted as the 'average' effect of scarring over the time period considered. In work in progress we investigate whether scarring has declined over time.

[9] This is fairly plausible since we are conditioning on α_i, which means that any risk factors common to the siblings, *j-2, j-1* and *j* are captured by α_i.

Heckman, 1981; Wooldridge, 2002). Since the data contain complete retrospective histories of fertility and mortality for each mother, the initial condition of the process can be specified as a reduced form equation for the first-born child in each family

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

where $\boldsymbol{z_{i1}}$ is a vector of exogenous covariates.[10] Equation (2) allows the vector of covariates $\boldsymbol{z}$ to differ from $\boldsymbol{x}$ in (1) and, similarly, the effect of unobservable family characteristics is allowed to differ from that in (1) by the proportion θ. In the special case, θ=1, unobserved heterogeneity has the same effect on the risks faced by first-borns as it does on the risk for later children. If we were to find that θ=0, then we could conclude that unobserved heterogeneity does not enter (2), from which it follows that specifying and estimating a separate equation for first-borns is in fact unnecessary as there is no correlation between the survival status of the previous sibling and the unobserved heterogeneity. The relevant tests on θ are presented in Section 6 below.

Equations (1) and (2) together specify a complete model for the infant survival process. We assume that u_{ij} is independently distributed as a logistic distribution (F), and that the family-specific unobservables, α_i, are independent and identically distributed as normal (density ϕ).[11] The likelihood function for family *i* is then given by

[10] In principle, there is no reason why the specification for the first-born should be the same as for later children, especially when the mortality risk for later children is represented by a conditional model, i.e., a model that conditions on the survival status of the previous sibling. However, this is a common assumption, and Heckman (1981) shows that it works quite well in applications.

[11] 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.

$$L_i = \int_{-\infty}^{\infty} \left(\prod_{j=2}^{n_i} F[(\boldsymbol{x_{ij}}'\boldsymbol{\beta} + \gamma\, y_{ij\text{-}1} + \sigma_\alpha\, \tilde{\alpha})(2y_{ij} - 1)] \right.$$

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

where, $\tilde{\alpha} = \alpha/\sigma_\alpha$. The log likelihood function is maximized using Stata (2000).

Specification Issues

The initial conditions problem; left truncation

The initial conditions problem arises because $y_{ij\text{-}1}$ and α_i are necessarily correlated and the model is recursive. As shown above, this problem is addressed by modelling the probability of death of the first-born child of each mother. This is a departure from previous research on sibling death clustering, which has consistently left-truncated the available data, restricting the sample to the 5, 10 or 15 years preceding the date of the survey. As a result, information on first-born children is lost for a number of mothers in the sample, but these studies do not seem to recognise that this can generate a bias in the coefficient estimators (see Curtis *et al,* 1993; Guo, 1993; Sastry, 1997; Bolstad and Manda, 2001). Comparison of estimates that do and do not address the initial conditions problem reveals a fairly large upward bias in the scarring effect when the problem is left unaddressed (see Arulampalam and Bhalotra, 2004). Tests of θ=0 (see equation 2) reported in Section 6 below further underline the importance of modelling the start of the process.

There is a potential problem with our identification strategy. This is that, if the first conception is a miscarriage, then the first-born (live) child is not in fact 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 (for details, see Arulampalam and Bhalotra, 2004). Given a non-linear model, identification of the scarring effect is aided by including amongst the regressors the age of the mother at birth of the index child, which is unique to each child (see Chamberlain, 1984; Hyslop, 1999).

Right-censoring

The available data are right censored because, at the time of the survey in 1998/9, some of the women who were interviewed had not completed their fertility. Since these will tend to be the younger women, the resulting problem is that we have in our sample a disproportionately large representation of children of older mothers. Mother's age is included as a regressor, or an intercept effect. However, as it is not interacted with $y_{ij-1,}$ the estimates of scarring in this paper may not be "representative" in the sense that, if we could include all children of younger mothers in the sample, then we might find a changed (probably smaller) scarring effect. A way around this might seem to be to conduct the analysis on only those mothers who have completed their fertility. However, this is not done since the implied sample selection is endogenous to the extent that fertility (and completion of fertility) is a choice variable. While future research might investigate this further by joint modelling of fertility and mortality, this is somewhat tangential to the current paper.

Distribution of unobserved heterogeneity

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. However, there are many families that experience no child deaths (see Appendix Table A1), 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.[12] 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_0, where,

$$p_0 = \frac{\psi_0}{1+\psi_{01}} \quad . \qquad (5)$$

In order to ensure the non-negativity of ψ_0, it was parameterised as $exp(\kappa)$, and κ was estimated.

Testing for the significance of inter-family heterogeneity

Let $corr(\alpha_i+u_{ij}, \alpha_i+u_{ik}) = \frac{\sigma_\alpha^2}{\sigma_\alpha^2 + \frac{\pi^2}{3}} = \rho$ say, for all $j \neq k \neq 1$. (6)

The correlation coefficient ρ gives the proportion of the total error variance that is attributed to the unobservable family effect, α_i. A test of H_0: $\sigma_\alpha^2=0$, which is a test that there are no unobservable family characteristics in the model and therefore that it collapses to a simple binary dependent variable model, is equivalent to a test of H_0: $\rho=0$ in equation (6). This can be tested as a likelihood ratio (LR) test but the test statistic will not be a standard χ^2 test since the parameter restriction is on the boundary of the parameter space. The standard LR test statistic has a probability mass of 0.5 at zero and

[12] See Narendranathan and Elias (1982) for an application of this distributional assumption in

$0.5\chi^2(1)$ for positive values. For this reason, the 10% critical value is used for the one-sided 5% significance level test.

5. 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. It is quite common to find age-heaping at certain values such as 6 months, 12 months etc. We have therefore 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. The latter is expected to capture effects of the physiological condition of the mother at the relevant time and, for this reason, is preferred to age of the mother at birth of the first child, which is the variable specified in some previous studies of childhood mortality. Family-specific covariates included are the educational attainment of each of the mother and father, religion and caste. The education variables are expected to capture socio-economic status. 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,

---

the context of modelling individual unemployment.

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.[13]

Covariates that are time-inconsistent or endogenous are not included as regressors. For example, the data contain information on household assets and on community variables such as access to piped water, which may be expected to influence survival chances. However, these data pertain to the time of the survey rather than to the dates at which children in the sample were exposed to infant death risk, and so are unlikely to be informative. For example, a woman aged 49 in 1999 may have experienced a birth and an infant death as long ago as 1969. Several previous analyses of infant mortality or fertility use variables that are time-inconsistent in this sense (e.g. Sastry, 1997; Atella and Rosati, 2000). The fact that most of these studies left-truncate the data does mitigate the time inconsistency problem by severing the retrospective information before it gets into the distant past. Nevertheless, the assumption of constancy of these variables for the time spans of 10-15 years covered in these studies deserves discussion.[14]

A further problem with some of these variables is that they are 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 reducing the risk of child mortality. Similarly, access to facilities like piped water will be endogenous if

[13] 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".

[14] There is plenty of evidence in the literature that both income mobility and geographical mobility in developing countries is considerable (see Baulch and Hoddinott, 2000; Williamson, 1998). Community infrastructure tends to grow rapidly from a low base in the process of economic development. Social norms also often change rapidly with growth and migration. Together, these facts make implausible the assumption that current household assets or current community infrastructure are a good proxy for the socio-economic status of the household at the time that the children in question were exposed to the risk of infant death.

selected families migrate to regions with these facilities, or if governments place these facilities in regions with worse health indicators.

6. Results

Table 2 presents estimates of scarring and unobserved heterogeneity. Effects of the other covariates in the model are presented in Appendix Table A2.

6.1. Scarring

Odds-ratios, calculated as $exp[\hat{\gamma}]$, are in Column [1]; a value of 1 for the odds-ratio indicates no scarring. The main result is that death in infancy of a previous sibling raises the odds of infant death for the index child in each of the 15 Indian states, and this result is obtained after controlling for a number of child and family-specific characteristics and for all unobserved differences between families. The scarring effect is significant at the 1% level in 13 states, at the 4% level in Kerala and at the 7% level in Punjab. So Punjab is the one state in which scarring is not significant at the conventional 5% level, and it has the smallest odds ratio (which is still fairly large, at 1.42). Punjab also exhibits the smallest odds ratio (2.85) in the "raw data", that is, before conditioning on inter-family heterogeneity and sibling characteristics (see Table 1, Column [5]).

It is interesting that Punjab is the richest of India's states, with the widest coverage of public infrastructure (Appendix Table A1).[15] But does this provide any hints as to the mechanism driving scarring? As discussed in Section 2, two of the candidate explanations of scarring (the fecundity and replacement hypotheses) involve the death of a child shortening the birth interval to the subsequent child. If these were the mechanisms at play then we might expect that Punjab, which has least scarring, has the longest birth intervals. It is therefore striking to find that exactly the opposite is the

[15] Recall that the deaths being analysed occurred over a span of about 30 years. Over this time, the ranking of states by per capita GDP has changed quite a lot. Punjab has nevertheless been a consistently rich state, with a successful agricultural sector.

case: Punjab has the highest fraction (21%) of births with a preceding birth interval of less than 18 months (see Appendix Table A1, Column [5]). It would, however, be incorrect to infer from these facts that the fecundity and replacement hypotheses have limited power. This is because the birth interval is a choice variable, and the impact of short birth intervals on mortality risk is increasing in poverty. 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, and 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. If Punjabi women 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 this state.[16]

The states fall into two groups in terms of the size of the estimated odds-ratio. In eight states, the odds-ratio is greater than about 1.9, or the odds of a child dying in infancy are about twice as high if the previous sibling died rather than survived infancy. These are remarkably large effects. There is a positive correlation of the size of the scarring effect with the mortality rate. The group with high scarring includes the group of contiguous states in the North-West, described in the Appendix as the backward *bimaru* region. The unexpected members of the high-scarring group are the relatively rich states of Haryana, Gujarat and Tamil Nadu. The low-scarring group includes the relatively prosperous states of Punjab and Maharashtra, but it also includes the North-Eastern-region (NE), the neighbouring Eastern state of West Bengal and the contiguous

[16] To explain the latter point further, an infant death in Punjab may stimulate the causal chain described by the fecundity or the replacement hypothesis, just as in any other region, resulting in a short interval to the next birth. However, in Punjab, the impact of a short preceding birth interval on mortality risk will be smaller if the mother is better nourished and so better able to produce a healthy child in spite of having had a shorter time in which to recover from the birth- and death- of the preceding child. Further research that estimates a model with both mortality and birth interval cast as behavioural variables is underway.

Southern states of Andhra Pradesh, Karnataka and Kerala, all of which are relatively socially progressive, though at very different levels of economic development. Overall, two patterns emerge. First, there is a geographical pattern whereby the large effects are in the North-Central states (with the exception of Tamil Nadu), and the small effects are in the South and East of the country (with the exception of Punjab). The second evident pattern is that scarring is high in states that are both socially and economically backward. A number of exceptions show that economic development alone does not consistently generate lower scarring. Further research into the correlates of scarring using more disaggregate data (e.g. district) is merited.

Percentage of raw clustering explained by scarring

Comparison of the estimated odds in Table 2 with the raw odds presented in Table 1 shows that estimated scarring accounts for a substantial fraction of the clustering seen in the raw data. A striking case here is that of Kerala. This state has a large raw odds ratio but, the odds that describe scarring, after controlling for inter-family heterogeneity, are considerably smaller. This shows that family characteristics account for most of the exceptionally low death risk in Kerala. Since the contribution of family unobservables to the model is relatively small in this state (Table 2, column [5]), the observable family characteristics that we condition upon seem to hold the key to understanding the success of this state in achieving low mortality. This is consistent with the outstandingly high levels of female (and male) education in this state.

In order to facilitate further interpretation, *the marginal effect* associated with γ is provided in Column [2] of Table 2. This 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.[17] The Table also shows, in Column [4], the percentage of raw persistence (or clustering) that is explained by scarring. The estimates show that this percentage is very large. The within-family process of scarring accounts for between 15.3% (in West Bengal) and 61.8% (in Haryana) of the clustering seen in the data, with the rest being explained by differences between families in observed and unobserved traits. Indeed, Column [5] confirms that the percentage of the error variance that is attributable to unobserved heterogeneity is smallest in Haryana, and is large in West Bengal.

How much would mortality fall if scarring were eliminated?

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 are in Column [6]. They suggest that, in the absence of scarring, mortality rates would fall by between 2.2% (in Punjab) and 7% (in Madhya Pradesh and Uttar Pradesh). Note that these estimates include the probability of death attached to first-borns. Since the weight attached to first-borns is greater in smaller families, estimates of the impact on infant mortality of eliminating scarring will tend to be smaller in states that have lower fertility levels [see the state-level data on fertility in Appendix Table A1].

The results have strong implications for policy, as discussed in Section 1. Scarring, unlike inter-family heterogeneity, involves responsive behaviour, which may

[17] 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.

be amenable to policy, while the latter involves largely untreatable factors like genes or fixed behavioural traits.[18]

6.2. Unobserved heterogeneity

As discussed in Section 1, 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). Table 2 Column [5] presents estimates for each of the 15 Indian regions of the intra-family correlation coefficient that obtains after controlling for observable family and child-specific characteristics including the survival status of the previous sibling. This is defined as the proportion of variance that is attributed to unobserved heterogeneity within the total error variance. It ranges between 0.02 in Haryana and 0.18 in the North-East region. As we may expect, the proportion of clustering attributable to scarring is inversely correlated with the intra-family correlation coefficient. The estimates reject the null of no family-level unobservables in ten states, at the 5% significance level. However, it is quite striking that in five states, unobservables have limited power to explain death clustering. These are the relatively developed states of Punjab, Haryana, Maharashtra, Kerala and Tamil Nadu. Previous research in demography has tended to

[18] 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.

over-estimate the contribution of fixed family traits on account of neglecting to allow for scarring.

6.3 Diagnostics

It is only in 2 of the 15 states (Kerala and Gujarat) that none of the interaction terms that allow for distinct effects of the covariates on death risks for first-born children is significant. A test of the null hypothesis that $\theta=0$ in equation (2) is reported in Table 3 Column [1]. The null is rejected in 7 of the 15 states at 10% or less. Rejection of the null 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 (see Section 3). 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. The model then collapses to a simple random effects model. The Table also presents a test of the hypothesis that $\theta=1$. This cannot be rejected in any of the states, which indicates that the unobserved heterogeneity terms in the equations for the first child and for subsequent children are perfectly correlated.[19]

Estimates of the parameter p_0 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 no infant deaths in the family. In line with the discussion above, very few families in Haryana belong to this group of families. In contrast, in Uttar Pradesh, West Bengal and Gujarat, almost 90% of families belong to this group.

6.4. Other covariates

In considering the effects of the other covariates (**x**), reported in Appendix Table A2, it is important to remember that these are conditional effects, obtained from a model that

[19] It is clear from the above that there are some states for which we cannot reject $\theta=1$ but, nevertheless, we cannot reject $\theta=0$. This simply reflects the imprecision with which this parameter is estimated.

includes the previous child's mortality status (y_{ij-1}) and a family-level random effect (α_i). This renders them incomparable with previous results in the literature that were 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! The odds ratio suggests that the risk of death increases by 64% if a girl is born in Punjab, holding all other covariates constant. Girls are significantly less likely to suffer infant death in Kerala, West Bengal and in the North Eastern region, consistent with popular belief that the status of women is relatively high in these parts. [20]

Multiple births suffer significantly higher risks in all states, the odds ratio being in excess of 10 in most. Most states show higher risks amongst children of higher birth order The education of fathers shows significant effects in 7 states and that of mothers in 8 states; it is only in Bihar and Rajasthan that neither parents' education has any effect. The odds ratio associated with parental education ranges between about 0.4 and 0.8 except in the case of Andhra Pradesh where both effects are remarkably large (see Appendix Table A2). The odds ratio associated with the mother's age at birth of the index child (which was entered as a quadratic) reported for two values, 15 and 25. It has a significant effect in every state other than Bihar.

Ethnicity effects are only significant in the states of Punjab and Uttar Pradesh, where they indicate greater death risks amongst scheduled castes and tribes and amongst

[20] 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.

other backward castes, respectively.[21]. Children in Muslim households enjoy lower death risks in four states: Andhra, Maharashtra, Uttar Pradesh and Rajasthan, the odds ratio indicating a huge gain in Andhra. In Rajasthan, being of a religion other than Hindu or Muslim also confers an advantage. The cohort dummy coefficients indicate, overall, that the survival chances of infants were smaller for mothers born in the 1950s and 1960s than in the subsequent decade and a half. In six states, both dummies are significant while, in another six, only the 1950s dummy is significant. In the three states of Bihar, Punjab and Tamil Nadu, both dummies are insignificant.

7. Conclusions

There is a remarkable degree of death clustering amongst siblings across India. 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. This could simply reflect the play of genetic or environmental factors shared by siblings of the same family (inter-family heterogeneity). Alternatively, observed death clustering might result from the event of death of a child influencing the survival chances of the next sibling (scarring). Using data on 223702 children spread across the 15 major states of India, this paper attempts to disentangle these two sources of death clustering. The results are striking. Sizeable scarring effects are identified and, in 14 of the 15 states, this effect is significant at the 5% level. In a model that controls comprehensively for inter-family heterogeneity, the 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,

[21] 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.

ranging between 15% in West Bengal and 61.8% 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 by between 2.2% (Punjab) and 7.1% (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.

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. This said, it is the case that the one state in which the scarring effect is weakest, and only significant at the 7% level, is Punjab, which is economically the most prosperous. 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. Further research into the processes underlying scarring is merited, and one step in this direction is to estimate birth spacing jointly with mortality.

Evidence of scarring implies a greater payoff to policy interventions that reduce child deaths, since it implies that preventing the death of a child immediately reduces the risk of death of siblings of that child. Understanding the scarring process also offers relevant insights into the relations of family behaviour, fertility and mortality (see Section 2). The emphasis in this paper on separating causal processes from correlations amongst siblings that arise from shared family traits has potential application in a number of other areas, for example, in analysis of other child outcomes.

Appendix: The relative position of the 15 regions

Some relevant socio-demographic and economic indicators are presented in Appendix Table A1. These data refer to recent years whereas the analysis in this paper pertains to deaths over a longer retrospective period, in the course of which state rankings for different indicators have changed. It is useful therefore to identify the states that have been consistently rich or poor (in terms of per capita gross state domestic product), and those that have been consistently better performers in terms of social and demographic indicators. Punjab and Maharashtra are the rich states, followed by Haryana and Gujarat. Uttar Pradesh, Bihar and Orissa are the poorest states, and Madhya Pradesh, Rajasthan and Andhra Pradesh have been changing positions just above them. Kerala, West Bengal, Karnataka and Tamil Nadu have quite consistently been middle-ranking in per capita GDP. In terms of social and demographic indicators, Kerala is an outlier and it is interesting that, despite its fairly sluggish economic development, it has had by far the lowest mortality rates. Across a range of human development indicators, Kerala stands well ahead of the rich states of Punjab and Maharashtra. On the other hand, economic and social backwardness do overlap amongst the northern states of Uttar Pradesh, Bihar, Orissa, Madhya Pradesh and Rajasthan. This set (minus Orissa, which lies to the East while the rest are Central-West) is popularly known as the BIMARU region (e.g. Dreze and Sen 1997). This is an acronym with an edge because *bimar* means *sick* in Hindi.

The rest of this section presents brief remarks on the summary statistics in Appendix Table A1. More than 68 per cent of Indian children who die before the age of 5 are infants. Improvements in technology and health-care practices tend to have a larger impact on deaths that occur after the age of one than on infant deaths (i.e. up to age one). Consistent with this, the proportion of under-5 deaths that occur in infancy is, at 80 per cent, highest in Punjab and Kerala, the states that lead in terms of economic and social progress respectively. About 13-21 per cent of birth intervals are less than 18 months. Since short preceding birth intervals are known to increase mortality risk for the index child, it is surprising at first glance to find that this proportion is highest in Punjab, the wealthiest of the states. Further comment on this is in the Results section. The age of mother at first birth ranges between 15 (in Madhya Pradesh and Andhra Pradesh) and 19 (in Punjab and Kerala). The proportion of mothers without any formal education ranges between 40 and 81 per cent (with Bihar and Rajasthan at the upper end of the range), if we exclude Kerala, in which only 11 per cent of mothers are uneducated. In Kerala and Punjab, about a quarter of mothers have achieved at least a secondary level of education, the corresponding proportion in most other states being less than a tenth. The proportion of fathers without any formal education is between about 20 and 40 per cent, except in Kerala, where it is 8 per cent. There is also considerable inter-state variation in electricity supply, with only 4% of households having no supply in Punjab as against 83% in Bihar.

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. NBER Working Paper 9209, National Bureau of Economic Research, Inc.

Altonji, J. G. and Dunn, T. A. (1996a),The effects of family characteristics on the return to education, *The Review of Economics and Statistics*, Vol. 78, No. 4, 692-704.

Altonji, J. G. and Dunn, T. A. (1996b), 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. (2004), Sibling death clustering in India: genuine scarring *vs* unobserved heterogeneity. Revised February 2004. Available as Discussion Paper 03/552, Department of Economics, University of Bristol.

Atella, V. and Rosati, F. (2000), Uncertainty about children's survival and fertility: A test using Indian microdata, *Journal of Population Economics*, Vol. 13, Issue 2, 263-278.

Baulch, R. and Hoddinott, J. (2000) (eds), *Economic Mobility And Poverty Dynamics In Developing Countries,* Frank Cass.

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.

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.

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 (2001), *Indian Planning Experience: A Statistical Profile*, January. New Delhi: Planning Commission.

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. J. and Singer, B. (1984), A Method for Minimizing the Impact of Distributional Assumption in Econometric Models for Duration Data, *Econometrica,* 271-320.

Hyslop, D. R. (1999), State dependence, serial correlation and heterogeneity in intertemporal labour force participation of married women, *Econometrica*, 67(6), 1255-94.

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

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

Kaestner, R. (1997), Are brothers really better? Sibling sex composition and educational achievement revisited, *The Journal of Human Resources*, Vol. 32, No. 2. Spring, 250-284.

Kirigia, J. M., Sambo, L. G., Aldis, W. and Mwabu, G. M. (2004), Impact of disaster-related mortality on gross domestic product in the WHO African Region, *BMC Emergency Medicine*, 4(1):1.

Koenig, M.A., Phillips, J.F., Campbell, O.M. and D'Souza, S. (1990), Birth interval and childhood mortality in rural Bangladesh, *Demography,* 27, 251-265.

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

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

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.

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.

Rosenzweig, M. R. and Wolpin, K. Kenneth I. L. (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.

Rosenzweig, M. R. and Wolpin, K. I. (1995), Sisters, siblings, and mothers: the effect of teen-age childbearing on birth outcomes in a dynamic family context, *Econometrica*, Vol. 63, No. 2, March, 303-326.

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 Corporation. Stata Statistical Software: Release 7.0. College Station, TX; 2001.

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.

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

Williamson, J. (1988), Migration and Urbanization, in H. Chenery and T. N. Srinivasan (eds), Handbook of Development Economics, Volume 1.

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

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 – Raw Data Probabilities Of Infant Death

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]	Odds Ratio * [5]
Central					
Madhya Pradesh	0.113	0.223	0.085	0.138	3.09
Uttar Pradesh	0.116	0.241	0.092	0.150	3.15
East					
Orissa	0.105	0.226	0.083	0.143	3.22
Bihar	0.080	0.240	0.061	0.178	4.83
West Bengal	0.076	0.194	0.060	0.134	3.79
North					
Rajasthan	0.100	0.211	0.080	0.131	3.06
Haryana	0.066	0.202	0.053	0.149	4.56
Punjab	0.060	0.143	0.055	0.088	2.85
West					
Gujarat	0.085	0.187	0.070	0.117	3.07
Maharashtra	0.059	0.138	0.048	0.090	3.15
South					
Andhra Pradesh	0.092	0.190	0.075	0.115	2.89
Karnataka	0.076	0.190	0.062	0.128	3.57
Tamil Nadu	0.071	0.160	0.060	0.099	2.96
Kerala	0.036	0.125	0.029	0.096	4.78
North-East	0.061	0.166	0.052	0.114	3.64

Notes: All probabilities are for death in *infancy*. * This is calculated as the ratio of column [2]/(1-column[2)] to column [3]/(1-column[3]). It is the raw data counterpart of the odds ratio associated with the scarring coefficient (γ) that is reported in column [1] of Table 2.

Table 2: Random Effects Logit Regression Results

STATE	Odds-Ratio associated with the death of previous sibling ($exp\ \hat{\gamma}$) [Pvalue][1] [1]	Estimated Marginal Effects[2] [2]	Raw data death clustering[3] [3]	Raw Clustering Explained by Scarring [2]/[3] % [4]	Estimated intra-family correlation coefficient, ρ [P-value][4] [5]	Reduction in infant mortality if no scarring %[5] [6]
Central						
Madhya Pradesh	2.07 [0.00]	0.0680	0.138	49.2	0.10 [0.00]	7.05
Uttar Pradesh	1.94 [0.00]	0.0635	0.150	42.5	0.12 [0.00]	6.96
East						
Orissa	1.90 [0.00]	0.0671	0.143	47.0	0.08 [0.00]	5.44
Bihar	2.35 [0.00]	0.0634	0.178	35.5	0.16 [0.00]	6.61
West Bengal	1.50 [0.00]	0.0206	0.134	15.3	0.22 [0.00]	3.09
North						
Rajasthan	1.93 [0.00]	0.0674	0.131	51.6	0.07 [0.00]	5.71
Haryana	2.50 [0.00]	0.0923	0.149	61.8	0.02 [0.00]	6.07
Punjab	1.42 [0.07]	0.0203	0.088	23.1	0.12 [0.00]	2.21
West						
Gujarat	1.98 [0.00]	0.0510	0.117	43.4	0.11 [0.00]	5.51
Maharashtra	1.70 [0.00]	0.0291	0.090	32.5	0.09 [0.00]	3.20
South						
Andhra Pradesh	1.45 [0.00]	0.0266	0.115	23.2	0.13 [0.00]	3.08
Karnataka	1.58 [0.00]	0.0328	0.128	25.6	0.15 [0.00]	2.79
Tamil Nadu	2.09 [0.00]	0.0558	0.099	56.3	0.05 [0.06]	5.15
Kerala	1.99 [0.04]	0.0277	0.096	28.9	0.07 [0.03]	2.92
North-East	1.70 [0.00]	0.0268	0.114	23.5	0.18 [0.00]	3.10

Notes: 1. The P-values refer to a one-sided test of odds-ratio=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. 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 Column [4] Table 1.

4. These p-values refer to a test of the hypothesis that the correlation coefficient is zero; see section 4 for further details.

5. 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 γ=0 is imposed after estimation.

Table 3: Model Diagnostics

STATE	θ [P-value for $\theta=0$] [P-value for $\theta=1$] [1]	Estimated proportion of families without any deaths $\phi/(1+\phi)$ [2]
Central		
Madhya Pradesh	0.56 [0.025] [0.078]	0.79
Uttar Pradesh	0.76 [0.000] [0.158]	0.91
East		
Orissa	0.74 [0.123] [0.588]	0.62
Bihar	1.03 [0.000] [0.896]	0.69
West Bengal	0.89[0.000] [0.647]	0.92
North		
Rajasthan	0.38 [0.354] [0.131]	0.64
Haryana	0.34 [0.942] [0.888]	0.42
Punjab	1.16 [0.093] [0.817]	0.68
West		
Gujarat	0.94 [0.033] [0.892]	0.88
Maharashtra	0.82 [0.166] [0.760]	0.71
South		
Andhra Pradesh	1.31 [0.006] [0.518]	0.74
Karnataka	0.44 [0.143] [0.062]	0.56
Tamil Nadu	0.19 [0.881] [0.524]	0.73
Kerala	1.79 [0.541] [0.788]	0.63
North-East	0.77 [0.000] [0.176]	0.76

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.

Appendix – Table A2 – Odds ratios from the model estimates

	MadhP	UttP	Orissa	Bihar	WBeng	Rajast	Haryan	Punjab	Gujar	Mahar	AndhP	Karnat	TamilN	Kerala	NE
All children															
Female	0.929	1.039	0.939	0.971	**0.811**	1.036	1.155	**1.639**	0.861	0.991	0.734	0.952	0.955	**0.685**	**0.865**
Ma's year of birth – after 1970 (base)															
pre 1959	**1.511**	**1.946**	**1.478**	1.782	**2.480**	**1.669**	**1.670**	1.391	**1.678**	**1.639**	**49.399**	**1.666**	1.357	**4.565**	**1.412**
1960-1969	**1.212**	**1.286**	1.046	1.377	1.265	**1.328**	1.277	1.190	1.180	1.254	**8.947**	**1.448**	1.026	**3.030**	1.162
Religion – Hindu (base)															
Muslim	0.841	**0.721**	0.832	0.545	0.957	**0.754**	0.821	0.628	0.855	**0.662**	**0.084**	0.795	0.773	1.053	1.162
Other	0.973	0.746	1.033	0.720	1.034	**0.439**	0.933	0.883	**0.416**	0.981	0.244	0.908	0.815	1.520	1.039
Ethnicity- other (base)															
Scheduled caste/tribe	1.099	1.134	0.878	1.170	1.177	1.115	0.937	**1.397**	0.975	1.047	2.666	0.936	0.557	1.680	0.864
Other backward caste	1.170	**1.208**	0.903	1.089	1.418	0.901	0.955	1.341	1.117	1.085	1.836	0.809	0.605	1.170	1.049
Ma's educ: none(base)															
Incomplete primary	0.884	0.844	0.916	0.766	0.929	1.108	1.335	1.360	0.863	0.970	0.276	0.963	0.957	**0.520**	0.858
Complete primary	0.809	**0.820**	0.981	1.015	**0.496**	0.914	0.711	**0.675**	0.786	1.126	1.073	0.732	0.916	0.773	0.845
Incomplete secondary	0.912	0.825	0.814	0.575	**0.400**	0.764	0.712	1.337	**0.659**	0.742	**0.122**	0.684	0.969	0.568	**0.797**
Comp sec or higher	**0.392**	**0.471**	**0.553**	0.972	0.581	0.653	0.910	0.782	**0.523**	0.995	0.918	0.855	0.896	**0.368**	0.953
Pa's educ: none (base)															
Incomplete primary	1.026	1.186	0.934	0.867	0.907	1.127	0.668	1.199	0.857	1.152	0.442	0.779	0.943	1.135	0.928
Complete primary	0.968	**0.831**	0.952	1.032	1.135	0.955	0.620	0.673	0.716	1.096	1.223	**0.547**	0.861	0.952	0.805
Incomplete secondary	0.933	**0.835**	**0.734**	0.777	0.944	0.908	0.722	0.939	0.929	0.922	0.241	**0.587**	0.840	0.759	**0.787**
Complete secondary	0.943	**0.838**	0.754	0.498	0.736	0.831	**0.667**	0.834	0.732	**0.501**	**0.071**	**0.610**	**0.501**	0.600	0.741
Higher	1.057	0.849	**0.450**	0.602	**0.538**	0.855	0.645	0.784	0.872	**0.451**	0.173	0.726	**0.553**	0.794	**0.571**
Mother's age = 15 yrs	**0.868**	**0.867**	**0.906**	0.824	**0.854**	**0.900**	**0.833**	**0.831**	**0.873**	**0.833**	**0.824**	**0.832**	**0.842**	**0.811**	**0.860**
Mother's age = 25 yrs	**0.929**	**0.928**	**0.958**	0.943	**0.931**	**0.940**	**0.928**	**0.907**	**0.944**	**0.921**	**0.943**	**0.934**	**0.944**	**0.972**	**0.938**
Multiple birth	**13.833**	**8.050**	**11.105**	11.509	**11.148**	**7.815**	**9.882**	**6.267**	**12.550**	**12.366**	**34173**	**19.193**	**8.637**	**9.510**	**8.162**
Birth order 3	1.045	1.038	0.971	1.151	**1.333**	0.978	0.886	1.183	0.882	0.900	3.191	**1.429**	1.064	**0.566**	1.006
Birth order 4	1.178	**1.245**	0.915	1.350	1.351	0.946	1.288	**1.846**	0.901	1.186	**7.173**	**1.421**	1.484	0.799	**1.360**
Birth order 5	**1.332**	**1.373**	0.936	1.489	1.379	1.149	1.032	1.701	1.168	**1.613**	**8.049**	1.297	1.402	0.830	**1.388**
Birth order >5	**1.693**	**1.761**	1.125	1.283	1.316	**1.401**	**1.766**	**3.578**	1.088	1.380	3.156	**1.796**	1.662	1.901	**1.693**
Previous sibling died	**2.070**	**1.941**	**1.897**	**2.363**	**1.502**	**1.929**	**2.499**	1.420	**1.976**	**1.702**	**1.448**	**1.580**	**2.089**	**1.988**	**1.702**
Intercept	**2.500**	**2.663**	1.743	11.201	2.162	1.176	**12.670**	4.510	2.913	5.883	**15.793**	**9.250**	**13.295**	9.877	2.501

Notes: (i) All variables except mother's age are binary indicator variables. (ii) The figures in bold indicate coefficients significant at 5% or less. (iii) The model for the North Eastern (NE) region included binary indicators for the individual states in this region. (iv) The model also included interactions of all regressors with a dummy for the first born child (not shown).