California Center for Population Research
University of California - Los Angeles

Measuring Within and Between Group Inequality in Early-Life Mortality Over Time: A Bayesian Approach with Application to India

Antonio Pedro Ramos, and Robert Weiss

Measuring Within and Between Group Inequality in Early-Life Mortality Over Time: A Bayesian Approach with Application to India[*]

Antonio P. Ramos and Robert E. Weiss

April 19, 2018

Abstract

Most studies on inequality in early-life mortality (ELM) compare average mortality rates between large groups of births, for example, between births from different countries, income groups, ethnicities, or different times. These studies do not measure within-group disparities. The few studies that have measured within-group variability in ELM have used tools from the income inequality literature. We show that measures from the income inequality literature, such as Gini indices, are inappropriate for ELM. Instead we develop novel tools that are appropriate for analyzing ELM inequality. We illustrate our methodology using a large data set from India, where we estimate ELM risk for over 400,000 births using a Bayesian hierarchical model. We show that most of the variance in mortality risk exists within groups of births, not between them, and thus that within-group mortality needs to be taken into account when assessing inequality in ELM. Our approach has broad applicability to many health indicators.

Keywords: Complex Inference Targets; Compositional adjustments; Demographic methods; Health inequality; Income inequality.

1 Introduction

Inequality in early-life mortality (ELM) is a fundamental dimension of social inequality (Moser et al. 2005; Stuckler et al. 2010). It is usually defined in by differences in average mortality rates between groups of births characterized by a single demographic, such as race, gender, income or country. In high income countries, national averages of child mortality are less than 10 deaths per thousand births, while these rates can be over 200 deaths per thousand births in low income countries. Within-country disparities in early-life mortality can be just as large, particularly across groups defined by wealth and income (Victora et al. 2003; Sastry 2004; Wagstaff 2000) and race and ethnicity (Brockerhoff and Hewett 2000; Antai 2011; Jankowska and Weeks 2013).

If births within a group have similar mortality risk, quantification of between-group disparities based on group averages is sufficient to characterize disparities in ELM in human populations. If births within a single group have widely varying underlying mortality risk, then within-group variation in mortality risk can be higher than between-group variation. Demographic and epidemiological research suggests that births within a group can have very different mortality risks. Thus we should expect a group of births defined by only one or a few demographic characteristics to have a variety of mortality risks. The few studies that have measured inequality within groups of births find that the variation in mortality risk among individuals from the same group (e.g. births from the same country) is often larger than the difference in group averages (Gakidou and King 2002, 2003). If that is the case, it is important to quantify within-group inequality in mortality in a manner similar to the quantification of income inequality.

Previous work on the measurement of inequality in ELM has borrowed inequality measures from the income inequality literature (Gakidou and King 2002, 2003; Murray

2

2001) and these measures have been implemented by international organizations (World Health Organization 2000; NIMS et al. 2012). However, it is not clear that these measures are appropriate for studying mortality risk. Mortality is measured on the probability scale and thus is bounded inside the $[0, 1]$ interval. Income, however, is usually defined on the positive real line or, sometimes, on the whole real line. Properties that are appropriate for measures defined on the real line, such as scale invariance, are not appropriate for probabilities. Thus while it is clear that within-group variability in ELM should be quantified, it is unclear how to do so.

Standard demographic decomposition techniques, such as Kitagawa (1955) or Oaxaca (1973), have only been applied to explain *average* differences in ELM and other health indicators between populations. They do not quantify other aspects of distributional changes such as variances or quantiles, which are essential to the study of inequality. In fact, two populations with the same average rates of ELM can have very different distributions of mortality risk (Gakidou and King 2002). Thus, it is important to develop a framework that quantifies and explains differences in distributions of mortality risk beyond mean changes. Currently no methodology measures and explains differences in the distribution of mortality risk across populations in a manner similar to decompositional methods in labor economics (Fortin et al. 2011) or relative distribution methods (Handcock and Morris 1999).

Mortality risk is not an observable quantity, however. We only observe that births are still alive or not at a certain age. To measure and explain inequalities in ELM across births we propose to estimate birth risk using a statistical model. Only then can we proceed to the inequality analysis. To make proper inferences we will need to propagate uncertainty from the estimation stage to the analysis of inequality stage.

In this paper we develop new methods to analyze the distribution of mortality risk

3

with the objective of measuring and explaining health inequality within and across populations. We demonstrate our methodology using two waves of the Demographic and Health Surveys (DHS) with information on more than 400,000 births from India. We estimate the underlying infant mortality risk with associated uncertainty and complex posterior covariance structure using Bayesian hierarchical logistic regression fit via Markov chain Monte Carlo (MCMC). Bayesian estimation and MCMC methods can handle complex models and facilitate the propagation of model and estimation uncertainty into subsequent analysis of inequality. Hierarchical models are a natural choice in our context because cases in our data set are nested in larger sampling units: births are nested in mothers, mothers are nested within sampling clusters, which are nested in districts and districts in states. Our inference target is the empirical distribution of mortality risk both marginally and as a function of covariates.

We make a number of contributions with broader applications to the existing literature inequality in ELM. First, we show that the usual practice of measuring inequality in ELM between groups of births using measures from the income inequality is not appropriate. Thus we suggest that to investigate ELM one needs measures that are specifically developed for mortality.

Second, proper assessment of inequality requires the consideration of within-group variability. We introduce methods that are appropriate for quantifying inequality in ELM within and between groups of births. We introduce summary measures that quantify the overall difference between two distributions. We develop adjustments that parsimoniously summarize differences between distributions of mortality risk. We extend covariate adjustment methods from the demographic literature by expanding the types of comparisons. We propose comparisons between distributions, while most research in demography and public health only compares means or other summary

4

measures. Our methods allow us to separate the impacts of changes in population composition from changes in the covariate-outcome relationship for binary outcomes. This makes it possible to answer questions such as how the distribution of mortality risk in 1995 would have looked if there had been no changes in the distribution of maternal age since 1975. We also develop ANOVA methods applied after Bayesian model fitting to quantify within-and-between group variance in mortality risk. Gakidou and King (2002) have suggested that most of the variance in mortality risk occurs within groups of births, not between them. We develop methodology to formally demonstrate the truth of this hypothesis. Our methodology directly allows for complex inference targets while allowing for measurement of uncertainty. Inequality measures and ANOVA are examples of new complex inferential targets that benefit from Bayesian inference.

Third, our analysis uncovers several new patterns of inequality in ELM in India. Inequality in ELM in India has been widely studied (Bhattacharya and Chikwama 2012; Singh et al. 2011; De and Dhar 2013; Nidhi et al. 2013; Kumar and Sing 2014; NIMS et al. 2012). The level of aggregation varies in these studies (states, districts, etc.), but great inequality has been documented. However, inequality in ELM at the individual birth level has not been investigated previously. We show that looking at average mortality rates among socioeconomic groups provides an incomplete picture of ELM inequality in India. For example, demographic groups defined by caste, maternal age at the birth of the child and religion explain less than 4% of all variance in ELM risk. Groups defined by quintiles of wealth explain between 7% and 10% of all variance. District has the highest explanatory power but still explains less than 20% of all variance in ELM risk. These patterns are consistent over time and suggest that previous analyses have largely overlooked variability in ELM risk within groups of births. We show that differences over time in the distribution of ELM risk can be summarized by a simple multiplicative

shift. And, finally, we show that changes over time in covariate distributions only account for a small fraction of the changes over time in ELM risk.

The paper is organized as follows. In section 2 we show that income inequality measures are not appropriate for studying ELM. Section 3 suggests summary measures and adjustments that are appropriate for assessing inequalities in ELM. Section 6 introduces covariate adjustment methods. Section 7 discusses some of the implications of our findings. Details of the data and the complete model we use for estimating infant mortality risk are presented in the appendix.

2 Problems with applying measures from the income inequality literature to mortality data

Most measures of income inequality have a common form (Firebaugh 2002). For a population of $i = 1, \ldots, n$ births with mortality risk π_i, income inequality measures take the form

$$\text{Inequality} = \frac{1}{n} \sum_{i=1}^{n} f\left(\frac{\pi_i}{\mu}\right), \tag{1}$$

where $\mu = n^{-1} \sum_{i}^{n} \pi_i$ is the average risk μ in the population. Define $r_i = \pi_i/\mu$ which is the ratio of the i birth mortality risk to the population average risk; inequality measures (1) are functions of r_i. Three popular measures of inequality are the squared coefficient of

variance, Theil indexes, and variance of the logs

$$f_1(r_i) = (r_i - 1)^2 \tag{2}$$

$$f_2(r_i) = r_i \log(r_i) \tag{3}$$

$$f_3(r_i) = \left\{ \log(r_i) - \frac{1}{n} \sum_{i=1}^{n} \log(r_i) \right\}^2 \tag{4}$$

and the closely related Gini index

$$\text{Gini} = \frac{1}{2n^2} \sum_{i=1}^{n} \sum_{j=1}^{n} |r_i - r_j|. \tag{5}$$

There are at least three problems in applying (1) or (5) to mortality data. First, in (1) or (5) the average level of income μ of the population under study is in the denominator. This is not a problem for income, as average income is usually far from zero and income inequality measures are designed to be the same whatever currency or units income is measured in or if income is doubled or halved across the board. Applied to mortality data, these indices divide the expected mortality risk for each birth, π_i, by the average mortality risk, μ. However, as μ approaches zero, the ratio r_i becomes large for fixed π_i, which increases the value of the inequality index. Thus a numerical problem may give a false impression that ELM inequality is increasing even as ELM rates tend towards zero.

The second problem is that probabilities are bounded by the $(0, 1)$ interval, unlike income which is defined on the real line. For income inequality, for a set of incomes s_i, $i = 1, \ldots, n$ and constant c, for all measures (1) – (5) $\text{ineq}(s_1, \ldots, s_n) = \text{ineq}(cs_1, \ldots, cs_n)$, income inequality measures remain unchanged if we multiply every income by c. Scale-invariance, which is sensible for income data does not readily translate to mortality risk.

7

The probability scale imposes serious constraints on c as the distribution of $c\pi_i$ also needs to be in the $(0, 1)$ interval; at best c must be less than 1 over the maximum of the π_is; for probabilities that cover the full range $(0, 1)$, only choices of $0 < c < 1$ are acceptable.

The third problem with inequality measures from the income inequality literature is that the inequality indices are not consistent with basic intuition about justice and health. Inequality measures on probability distributions should lead to the same conclusion whether we are measuring probability of mortality, π_i, or probability of survival, $1 - \pi_i$. A country with high inequality in infant mortality has equally high inequality in infant survival. Thus we assert a symmetry property in the evaluation of inequality on probability distributions

- **Symmetry:** Suppose Z_1 and Z_2 are two probability distributions on $[0, 1]$. Then comparing inequality between Z_1 and Z_2 and comparing inequality between $1 - Z_1$ and $1 - Z_2$ should produce the same conclusion.

We use simulated data from a series of beta distributions with decreasing means that resemble mortality risk distributions to illustrate how inequality measures behave. Table 1 presents results for four beta(α, β) distributions, where β is fixed at 10 for all four distributions and α decreases from 1 to .5, .3 and .01. As α decreases, the mean $\alpha/(\alpha + \beta)$ and standard deviation $\alpha\beta/[(\alpha + \beta)^2(1 + \alpha + \beta)]$ of mortality risk decrease. The mean gives the same interpretation whether we look at survival or mortality: survival is increasing and mortality is decreasing top to bottom; we interpret this as inequality is also decreasing from top to bottom. The standard deviation is the same whether we look at mortality or survival and suggests that inequality is decreasing from top to bottom. The coefficient of variation gives opposite stories about inequality for mortality and survival as do the Gini and Theil inequality indices: increasing from top to bottom for mortality, but decreasing for survival.

8

Income inequality measures such as CV, Gini or Theil do not satisfy the symmetry property and therefore are not suited to assess inequality in probability distributions. Income distributions are fundamentally different from probability distributions. The idea of scale invariance is neither necessary nor appropriate for measuring inequality in mortality because probabilities do not scale. In contrast, the mean or variance appear to be possible measures of inequality for distributions of ELM risk.

We now look at other methodologies to measure and describe inequality in ELM risk.

3   Analyzing Inequality in Early Life Mortality

In this section we develop appropriate tools to study inequality in mortality data. We begin with definitions and notation. Consider two populations we wish to compare, a reference population $j = 0$, and a comparison population $j = 1$. Each population has a distribution Π_j of mortality risk. The distributions Π_j are induced by a distribution M_j on covariate X space and transformations G_j from X space to probability space. In practice both M_j and G_j will need to be estimated which we discuss in section 5, but we take them to be known for the current discussion. Let x_i be a random draw from M_j; then, for a logistic regression model, functions G_0 and G_1 would be inverse logit functions $\exp(\cdot)/(1 + \exp(\cdot))$ of $x_i'\beta_0$ or $x_i'\beta_1$ where β_0 and β_1 are regression coefficients for the reference and comparison populations. Populations 0 and 1 may differ in the distributions M_0 and M_1 of the covariates and in the values β_0 and β_1 of the regression covariates that multiply the covariates.

In comparing Π_0 to Π_1, we will make interpretable adjustments to Π_0 to make it be more similar to Π_1 so as to understand how and why Π_0 and Π_1 differ.

9

3.1 Location-Scale Adjusted Distributions

It is useful to summarize differences between distributions based on a few summary measures. If Π_0 and Π_1 were continuous distributions on the real line, we might recenter and rescale Π_0 so that the resulting distribution Π_A has the same mean and variance as Π_1. We create a new distribution, Π_A that represents a counterfactual or synthetic population. Then we could summarize the remaining differences between Π_A and Π_1 by a one number summary such as a Kullback-Leibler divergence or other divergence between distributions, or we could plot Π_A and Π_1 and inspect the differences graphically. General location-scale shifts do not work particularly well for distributions with bounded support, though with restrictions they can be useful. Let us assume that Π_0 and Π_1 enjoy full support on $(0, 1)$. Let $\Pi_A = a + b * \Pi_0$ be a location scale adjusted version of Π_0 with $a, b \geq 0$. While b could be negative, that flips the distribution around so that large probabilities become small and small probabilities become large, which violates the sense of manipulating one mortality distribution to be more like a second mortality distribution. Thus we need $0 \leq a < 1$, $0 < a + b \leq 1$, $0 < b \leq 1$ for Π_A to be a distribution contained in $(0, 1)$.

As mortality distributions are generally skewed with a mode close to zero and a long right tail, we prefer to restrict $a = 0$ and $0 < b < 1$. Assuming the mean μ_0 of Π_0 is greater than the mean μ_1 of Π_1, we can take $b = \mu_1/\mu_0$. This leaves Π_A with support in the range $(0, b) \subseteq (0, 1)$. If we think of Π_0 as the mortality risk in a country at an earlier time point and Π_1 as mortality risk at a later time, then choosing $b < 1$ is not a problem, as, in the current era, mean ELM risk is generally decreasing over time. If mortality risk increased, then we would manipulate Π_1 instead of Π_0. Medians could certainly be used in place of means as well and we have used both means and medians in our work.

Thus $b = \mu_1/\mu_0$ summarizes the differences between the distributions Π_0 and Π_A. To

10

fully understand the differences between Π_0 and Π_1, we would still need to summarize the differences between Π_A and Π_1.

3.2 Decomposition Methods

When comparing two distributions, decomposition methods are used to disentangle the effects of differences in coefficients β_0 versus β_1 from effects of differences in the distribution of the covariates (Kitagawa 1955; Oaxaca 1973). For example, mother's age at the birth of an infant is an important covariate for estimating mortality risk as births from younger and older mothers have higher mortality risk. For two populations the relative importance of maternal age may differ and the distribution of maternal age can differ for different populations.

A useful adjustment is to take the distribution of covariates for either population 0 or 1 and the regression coefficient β_j for the other population. Thus the distribution Π_A could be constructed as drawing x_i at random from covariate population M_1, then multiplying by coefficient vector β_0 from population 0 and taking the inverse logit to construct adjusted distribution Π_A. A conditional version for sampling x_i is also possible. Suppose that the density of $x_i = (x_{i1}, x_{i2})'$ in population j is $f_j(x_i) = f_j(x_{i1})f_j(x_{i2}|x_{i1})$. We can construct adjusted probability distribution Π_A by drawing from a combination covariate density $f_0(x_{i1})f_1(x_{i2}|x_{i1})$, then multiply by β_0 and taking the inverse logit. This constructs a population of probabilities based on population Π_0 except that the x_{i2} covariate distribution has been adjusted to look conditionally like covariates from population 1.

The distributions Π_A that we have described here and in the previous subsection are called adjusted or *counterfactual* distributions.

11

3.3 Comparing Distributions

We now consider methods to quantify the differences between Π_0 and Π_1 by comparing Π_0 to Π_A and also Π_A to Π_1. Let L be a comparison (divergence, distance) operator between distributions that satisfies the triangle inequality

$$L(\Pi_0, \Pi_1) \leq L(\Pi_0, \Pi_A) + L(\Pi_A, \Pi_1) \tag{6}$$

The Oaxaca decomposition (Oaxaca 1973), where $L(\Pi_0, \Pi_1) = |\mu_0 - \mu_1|/\mu_1$ is the absolute value of the relative difference in means, follows the triangle inequality when $\mu_0 \leq \mu_A \leq \mu_1$ where μ_A is the mean of the adjusted distribution. Similarly, $L(\Pi_0, \Pi_1) = |\mu_0 - \mu_1|$, the absolute difference in means also follows the triangle inequality.

3.4 Numerical Summaries for Comparing Distributions

It is often useful to quantify how much one distribution differs from another by a one number summary. Traditional approaches have summarized the distributions first with a one-number summary of the distribution and then compared the summaries. Measures such as the mean, variance, other moments, or quantiles can be used to compare and summarize distributions and we are not opposed to these methods, but we wish to consider other methods as well. We can measure how similar two distributions are by using divergence measures, provided the distributions share the same support. A commonly used measure is the Kullback-Leibler divergence,

$$K(\Pi_0, \Pi_1) = \int \log \frac{f_0(X)}{f_1(X)} f_0(X) \, dX, \tag{7}$$

12

where the densities $f_0(\cdot)$ and $f_1(\cdot)$ correspond to the distributions Π_0 and Π_1. The divergence $K(\Pi_0, \Pi_1)$ can be interpreted as the expected information for discriminating $f_0(X)$ from $f_1(X)$ based on a single observation from $f_0(X)$. Another useful measure is the L_1 norm

$$L_1(\Pi_0; \Pi_1) = \frac{1}{2} \int \left| f_0(X) - f_1(X)) \right| dX. \tag{8}$$

L_1 which quantifies how much probability mass needs to be moved so that one distribution becomes identical to the other one (Weiss 1995). The L_1 norm (8) is between 0 and 1, with 0 meaning the distributions are the same and 1 meaning the two distributions Π_0 and Π_1 do not share support.

4   Using Covariates to Explain Mortality Risk

Variation in mortality risk can often be explained by covariates. By analyzing subpopulations identified by levels of a categorical covariate, we split a large population into several subpopulations. Rather than being concerned with how one subpopulation compares to another subpopulation when we have a set of K subpopulations, we are more concerned with how distinct the subpopulations are and whether targeting one or anther subpopulation for intervention might be useful for reducing overall ELM risk. However, we have found that even covariates that explain significant amounts of risk are not necessarily very useful in identifying subgroups to target so as to alleviate high risk ELM, a situation we illustrate in section 6.

4.1 ANOVA Methods

Consider a categorical covariate X that identifies groups such as wealth quintiles or maternal age groups in our population. If most of the variability in risk occurs between groups, then knowing the mean risk for each group identified by X is highly informative about inequality in the population. In contrast, if most of the variability in risk occurs within groups, then comparing mean risk between groups does not provide much information on risk variation.

Variation in risk can be expressed as the between group variance plus the weighted sum of the within-group risk variances. The law of total variance produces the decomposition

$$\text{Var}(\Pi_0) = \text{E}_X \left(\text{Var}[\Pi_0|X] \right) + \text{Var}_X \left[\text{E}(\Pi_0|X) \right] \tag{9}$$

where $\text{E}_X(\text{Var}[\Pi|X])$ is the average within group variance and $\text{Var}_X[\text{E}(\Pi_0|X)]$ is the between-group variance of the group means. We fit Analysis Of Variance (ANOVA) models using mortality risk as outcome and group membership as predictor and use R^2 as a measure of how much of the total variance in ELM risk can be explained by membership in a particular group.

As an example, in section 6 we run separate ANOVAs for covariates and for births from a given year in India. We consider whether the R^2 might be increasing or decreasing over the years, to identify whether inequality across covariate categories might be increasing or decreasing with time.

It is not necessary that the ANOVA model be as complex as the Bayesian hierarchical model used to model the data. We consider the ANOVA as a summary measure of the posterior, not a model on its own right.

14

4.2 Time Trends

We are particularly interested in time trends in ELM and how group comparisons evolve over time in India. We use year to define subpopulations, and then use additional covariates such as income quintiles to further refine the population of births into sub-sub-populations. We use ANOVA to quantify the within-year inequality within and across income groups and then assess the trend in this summary over time. Among other additional comparisons, we use divergence measures to compare the distribution of mortality in the baseline year against the distribution of mortality risk in each of the following years. We can construct synthetic populations holding the distribution of some key covariate constant to evaluate its effect on mortality risk time trend. For example, we can estimate the distribution of mortality risk over time if the distribution of maternal age was fixed over time.

5   Accounting for Uncertainty When Evaluating Inequality

We estimate risk for subjects with a Bayesian hierarchical logistic random effects regression models fit using Markov chain Monte Carlo (MCMC). For each iteration $\ell = 1, \ldots, L$ of the MCMC , we have a point estimate $\pi_i^{(\ell)}$ of π_i for all births and we can now proceed with our program to evaluate ELM risk inequality as sketched in sections 3 and 4. For example, we can use the $\pi_i^{(\ell)}$s and each birth's wealth quintile x_i to calculate an $R^{2(\ell)}$. However, this calculation is for only a single iteration of the MCMC. Thus for all iterations $\ell = 1, \ldots, L$ we calculate an $R^{2(\ell)}$ producing a posterior distribution for R^2. This uncertainty can be used in comparing various R^2s for different covariates and thus determine which covariates do the best at distinguishing ELM risk.

As our sample π_i^{ℓ} is discrete, when we wish to calculate the divergence between two

subpopulations of π_i^ℓs, we use a kernel density smoother to first smooth out the distributions before calculating a divergence. Then, as with ANOVA R^2s, we can calculate distributions for divergences.

In our graphs and numerical summaries we present point estimates and 95% pointwise credible intervals calculated from the multiple MCMC samples.

6   Modeling of Early Life Mortality Risk in India

Our data on early life mortality in India comes from the Demographic and Health Surveys, (DHS) `http://www.measuredhs.com/`. We use data from two DHS surveys to construct a *restrospective panel* from 1975 to 1997. A third wave covering more recent years is available but does not include district level information which we use in our models, so we were unable to make use of the third wave data. We analyze births to mothers aged 15–35 from 1975 through 1998 to reduce truncation and censoring. We analyze a total of 408,706 births from 141,999 mothers in 3,806 sampling clusters taken from 443 districts and 26 states.

6.1   Model Specification and Estimation of Mortality Risk

We fit a hierarchical Bayesian logistic regression model to estimate each infant i's mortality risk π_i using covariates, time, and time varying covariate effects.

Let $i = 1, \ldots, N$ index the N infants, nested in $m = 1, \ldots, M$ mothers, nested in sampling clusters $c = 1, \ldots, C$, nested in districts $d = 1, \ldots, D$, nested in states $s = 1, \ldots, S$. Year of birth is indexed by $t = 1, \ldots, T$ with $t = 1$ for births from 1975 up to $t = 23 \equiv T$ for births from 1997. Let y_i be the binary response variable, whether infant i with covariate vector x_i born from mother $m(i)$, in sampling cluster $c(i)$, in district $d(i)$, in state $s(i)$, and in

16

year $t(i)$ was alive $y_i = 0$ at the age of one or not $y_i = 1$. We specify random intercepts for mother and cluster and bivariate random intercepts and time slopes at the district and state level. Let $\text{logit}(\pi) = \log(\pi/(1-\pi))$. Our Bayesian hierarchical logistic regression is

$$y_i|\pi_i \sim \text{Bern}(\pi_i) \tag{10}$$

$$\text{logit}(\pi_i) = x_i'\alpha + \delta_{m(i)} + \gamma_{c(i)} + \xi_{d(i),1} + \xi_{d(i),2} * t(i) + \tau_{s(i)1} + \tau_{s(i)2} * t(i), \tag{11}$$

where π_i is the unobserved probability of mortality for infant i, α is a vector of unknown coefficients corresponding to the covariates in x_i, δ_m is the mother random effect with variance σ_1^2, γ_c is the sampling cluster random effect with variance σ_2^2, $\xi_{d,1}$ and $\xi_{d,2}$ are district random intercepts and slopes with prior covariance matrix $\boldsymbol{\Sigma}$, and τ_{s1} and τ_{s2} are state random intercepts and slopes with prior covariance matrix $\boldsymbol{\Psi}$. Define $\xi_d = (\xi_{d1}, \xi_{d2})'$ and $\tau_s = (\tau_{s1}, \tau_{s2})'$. The random effects priors are $\delta_m|\sigma_1^2 \sim \text{N}(0, \sigma_1^2)$, $\gamma_c|\sigma_2^2 \sim \text{N}(0, \sigma_2^2)$, $\xi_d|\boldsymbol{\Sigma} \sim \text{N}_2(\mathbf{0}, \boldsymbol{\Sigma})$, and $\tau_s|\boldsymbol{\Psi} \sim \text{N}_2(\mathbf{0}, \boldsymbol{\Psi})$.

Covariates at the child level are birth order, birth year, gender, maternal age, with household level covariates of religion, caste, wealth quintile, residence (rural or urban), and maternal education. We use splines to capture non-linearities in the time trend and investigated whether covariates have time-varying effects, though these analyses are not shown in this paper. We include main effects and all $_9C_2 = 36$ two-way interactions between covariates. The effects for maternal age, birth order, and birth year are modeled using b-splines.

The intercept is given a $\text{N}(0, 9)$, prior, all main effects are given standard normal priors, and all two-way interactions are given $\text{N}(0, 0.5)$ priors. The web appendix gives the complete model specification.

While interest in individual probabilities π_i is standard in hierarchical logistic regres-

17

sions, we are not really interested in individual probabilities. Rather, we are interested in the entire collection of probabilities π_i, $i = 1, \ldots, N$ simultaneously as the key quantity of interest from the model. This set of probabilities are used as inputs to our inequality calculations. We generate samples from the posterior distribution of the ELM risks, $\pi_i^{(\ell)}$, $i = 1, \ldots, N$, and where $\ell = (1, \ldots, L)$ indexes the L MCMC samples.

6.2 Analysis of Inequality

Disparities in mortality risk are not well-captured by looking at national averages of mortality rates. For example, the infant mortality rate (IMR) in India was 12% in 1975 and 6% in 1995, both calculated as unadjusted means from our data. While this is a remarkable decline, these numbers do not quantify important aspects of the distribution of mortality risk as estimated by our model. To illustrate, figure 1 displays three kernel density estimates: one representing the distribution of mortality risk in India for 1975, which is the less peaked, longer right tailed density, and one for 1995, which has the higher mode and smaller right tail and a third density mostly hidden next to the 1995 density which we discuss in a moment. The shaded areas display 95% pointwise credible regions; we generated L kernel density estimates from L MCMC samples $\pi_i^{(\ell)}$ and calculate the .025 and .975 percentiles at each point along the x axis. The modal mortality risk was 2% in 1975 and 1% in 1997, which means that the actual mortality risk most infants experienced in both years were quite different from the IMR. In 1975 26% of infants have mortality risk higher than the 12% 1975 IMR while in 1995 only 11% of infants have mortality risk higher than the 1975 IMR of 12%. The uncertainty in the kernel density estimates shows that the distribution of ELM risk from our model is well determined in both years, and thus we are comfortable in asserting that the mortality risk distribution is stochastically much smaller in 1995 than in 1975. The third density

18

plotted in Figure 1 is almost indistinguishable from that of 1997 and was constructed by taking the probabilities in 1975 and multiplying by the ratio of the 1997 median to the median for 1975. The resulting adjusted 1975 density is virtually indistinguishable from the 1997 distribution. Summary statistics from figure 1 are given in table 2.

Looking at the median of π_i by wealth quintile we find that the mortality rate by quintile is: lowest quintile .08; second .07; middle .05; fourth .04; highest .03, which shows that the mortality rate is almost 3 times higher than the poorest group compared to the richest. However, these calculations ignore within-group variability. Figure 2 shows box plots of posterior means $E[\pi_i|Y]$ of mortality risk for every infant in our data, by wealth quintile, where Y denotes the entire data set. By contrast with the means by quintile, figure 2 shows that the distribution of mortality risk is highly variable within each wealth quintile. For example, 10% of births from the two lowest quintiles have mortality risk higher than 12% and 30% of births from other wealth quintiles have higher mortality risk that the .08 median mortality risk among the poor. This suggests that summaries such as means hide useful information and that quantifying individual mortality risk is necessary.

We quantify how much of the changes in ELM over time can be explained by the multiplicative adjustment. We use 1975 as the comparison year to compare against each subsequent year $1976, \ldots, 1995$. In figure 3 we use KL (left) and L_1 (right) divergence measures to compare differences in the distributions year t versus 1975, with (lower dotted curve) and without (upper solid curve) the multiplicative median adjustment. Bands are pointwise 95% posterior intervals. The differences between the baseline year and the other years are close to zero after the adjustment and do not grow larger with time. The overall distribution of mortality risk in India grows increasingly different from the baseline distribution in 1975, yet these changes can mostly be explained by a

19

parsimonious multiplicative shift.

Figure 4 analyses the effect of the change in maternal age distribution over time on the distribution of mortality risk. The left panel plots histograms of maternal age in 1975 (pink) and 1995 (grey). While there is a great deal of overlap in the age densities, maternal age in 1995 has fewer younger mothers as compared to 1975. The right-hand panel displays the distribution of ELM risk for 1975 and 1995 and an adjusted version of 1975 where we have adjusted the maternal age distribution to look like that in 1995. Adjusting for maternal age in 1975 changes the adjusted 1975 distribution slightly towards smaller values, but not far enough. Table 3 presents Kullback-Leibler and L_1 divergences quantifying the differences between the adjusted 1975 distributions of infant mortality and the 1995 (unadjusted) distribution after adjustments for the various covariates in our sample with 95% posterior intervals in parentheses. Single covariate adjustments alone do not explain the changes in distribution of mortality risk from 1975 to 1995.

Figure 5 shows the extent to which variability in mortality occurs within covariate categories and doesn't change over time. We look at the R^2 for caste, religion, states, maternal education, wealth, residence, and districts over time. For all covariates and all times, the vast majority of variation in inequality in infant mortality is within categories and not between categories. For religious group and for caste, the covariate contains almost no information on the variability of inequality: for these groups, almost 100% of the variation is within category. These patterns are consistent over time. This suggests that socioeconomic groups described by a single covariate are much more heterogeneous than previously thought and that infants within a single sub-group defined by levels of wealth or other covariates have very different mortality risks. This also confirms the notion that comparisons between the means of groups defined by levels of a categorical

20

demographic variables ignores substantial within-group variability in mortality risk due to the high levels of heterogeneity. This finding highlights the importance of measuring inequality within as well as between groups of births. This is also an important contribution to the literature on India and it has been overlooked by previous research.

7   Discussion and Conclusions

While in this paper we have looked at ELM, our results extend to mortality data in general, such as adult mortality. Because mortality risk is a probability, it is fundamentally different from income and our results suggest that measures from income inequality literature are inappropriate for mortality.

Our findings have implications for measures of inequality in mortality and shed light on how to design summary measures to quantify inequality in mortality. Currently inequality in early-life mortality uses measures from the income inequality literature or calculates average differences between large groups of births, such as between countries or between income groups within-countries. Our results show that these analyses are incomplete, and possibly misleading. For example, several studies have suggested that some Sustainable Development Goals (SDG) were not achieved due to high levels of inequality (NIMS et al. 2012; Houweling and Kunst 2010; Gwatkin 2005). UN General Assembly Resolution 68/261, which highlights Sustainable Development Indicators as a key measurement for measuring progress in reducing early-life mortality, recognizes this fact by recommending that health indicators should be disaggregated, where relevant, by subpopulations (Economic and Social Council: Statistical Commision 2016). Our approaches identify within-group variability and thus we can identify inequalities that are missed when one only looks at average levels.

Our methods have broad applicability to other health outcomes, including those that are not defined in the probability scale, such as life expectancy or malnutrition. Our methods are particularly useful when scientists suspect that within-group variability can be substantial or when researchers are interested in other aspects of the distributional differences between two populations besides mean differences. For example, our methods can be used to determine if the changes over time in height among children is driven by relatively slow growth of certain high-risk children versus faltering of the entire population.

Although we use a Bayesian approach in this paper, our methods are potentially compatible with a frequentist approach. If a researcher can fit a frequentist model and simulate predictions for mortality risk, for example using bootstrap methods, our methods can be applied to the bootstrap predictions. One notable advantage of the Bayesian approach is that is makes inference easier. For example, the popular LASSO Tibshirani (1996) can do variable shrinkage and selection, but does not naturally provide standard errors for coefficient estimates and so also won't provide uncertainty measurements for probabilities which are functions of the coefficient estimates. By contrast, in a Bayesian model, we can easily calculate the probability that one sub-population is more unequal than another sub-population by counting the proportion of MCMC samples in which one sub-population's measure of inequality is higher than another. The frequentist approach can depend on being able to fit a complex model to bootstrap samples, something that can fail, for example in attempting to bootstrap a logistic regression with a small fraction of cases or where a variance component gets set to zero when using maximum likelihood.

References

Antai, D. (2011), 'Regional Inequalities in Under-5 Mortality in Nigeria: A Population-based Analysis of Individual and Community Level Determinants', *Population Health Metrics* **9**(1), 6.

Bhattacharya, P. and Chikwama, C. (2012), 'Inequalities in Child Mortality in India', *Asian Population Studies* **7:3**, 243–261.

Brockerhoff, M. and Hewett, P. (2000), 'Inequality of Child Mortality among Ethnic Groups in Sub-Saharan Africa', *Bulletin of The World Health Organization* **78**(1), 30–41.

De, P. and Dhar, A. (2013), 'Inequality in Child Mortality Across Different States of India: A Comparative Study.', *Journal of Child Health Care* **17**(4), 397–409.

Economic and Social Council: Statistical Commision (2016), 'Report of the Inter-Agency and Expert Group on Sustainable Development Goal Indicators'. Available at `https://unstats.un.org/unsd/statcom/47th-session/documents/Decisions_final_unedited.pdf`.

Firebaugh, G. (2002), *The New Global Geography of Income Inequality*, Harvard University Press.

Fortin, N., Lemieux, T. and Firpo, S. (2011), Decomposition Methods in Economics, *in* O. Ashenfelter and D. Card, eds, 'Handbook of Labor Economics', Vol. 4, Elsevier, chapter 1, pp. 1–102.

Gakidou, E. and King, G. (2002), 'Measuring Total Health Inequality: Adding Individual Variation to Group-Level Differences', *International Journal for Equity in Health* **1**(3).

Gakidou, E. and King, G. (2003), Determinants of Inequality of Child Survival: Results from 39 countries, *in* C. J. L. Murray, ed., 'Health Systems Performance Assessment: Debates, Methods and Empiricism', World Health Organization, chapter 36.

Gwatkin, D. (2005), 'How Much Would the Poor Gain from Faster Progress towards the Millenium Development Goals for Health?', *Lancet* **365**(9), 813–817.

Handcock, M. S. and Morris, M. (1999), *Relative Distribution Methods in the Social Sciences*, Springer.

Houweling, T. A. and Kunst, A. E. (2010), 'Socio-Economic Inequalities in Childhood Mortality in Low and Middle Income Countries: A Review of the Evidence', *British Medical Bulletin* **93**, 7–26.

Jankowska, Marta, M. B. and Weeks, J. R. (2013), 'Estimating Spatial Inequalities of Urban Child Mortality', *Demographic Research* **28**(2), 33–62.

Kitagawa, E. M. (1955), 'Components Between the Differences Between two Rates', *Journal of American Statistical Association* **50**, 1168–1194.

Kumar, A. and Sing, A. (2014), 'Is Economic Inequality in Infant Mortality Higher in Urban Than in Rural India?', *Journal Maternal Child Health* **18**, 2061–70.

Moser, K. A., Leon, D. A. and Gwatkin, D. R. (2005), 'How does Progress towards the Child Mortality Millennium Development Goal affect Inequalities between the Poorest and Least Poor? Analysis of Demographic and Health Survey data', *British Medical Journal* **331**(7526), 1180–1182.

Murray, C. (2001), 'Commentary: Comprehensive Approaches are needed for full understanding', *British Medical Journal* **323**(7314), 680–681.

Nidhi, J., Singh, A. and Pathak, P. (2013), 'Infant and Child Mortality in India: Trends in Inequalities across Economic Groups', *Journal Population Research* **30**(4), 347–365.

NIMS, ICMR and UNICEF (2012), 'Infant and Child Mortality in India: Levels, Trends and Determinants', National Institute of Medical Statistics, Indian Council of Medical Research (ICMR), and UNICEF India Country Office, New Delhi, India. Available at http://unicef.in/PressReleases/374/The-Infant-and-Child-Mortality-India-Report.

Oaxaca, R. (1973), 'Male-Female Wage Differentials in Urban Labor Markets', *International Economic Review* **14**(3), 693–709.

Sastry, N. (2004), 'Trends in Socioeconomic Inequalities in Mortality in Developing Countries: The Case of Child Survival in Sao Paulo, Brazil', *Demography* **41**(3), 443–464.

Singh, A., Pathak, P. K., Chauhan, R. K. and Pan, W. (2011), 'Infant and Child Mortality in India in the Last Two Decades: A Geospatial Analysis', *PLoS ONE* **6**(11), e0026856.

Stuckler, D., Basu, S. and Mckee, M. (2010), 'Drivers of Inequality in Millennium Development Goal Progress: A Statistical Analysis', *PLoS Medicine* **7**(3), e1000241.

Tibshirani, R. (1996), 'Regression Shrinkage and Selection via the LASSO', *Journal of the Royal Statistical Society, Series B* **58**(1), 267–288.

Victora, C., Wagstaff, A., Schellenberg, J., Gwatkin, D., Claeson, M. and Habicht, J. (2003), 'Applying an Equity Lens to Child Health and Mortality: More of the Same is not Enough', *The Lancet* **362**(9379), 233–241.

Wagstaff, A. (2000), 'Socioeconomic Inequalities in Child Mortality: Comparisons Across Nine Developing Countries', *Bulletin of The World Health Organization* **78**(1), 19–29.

Weiss, R. E. (1995), 'An Approach to Bayesian Sensitivity Analysis', *Journal of the Royal Statistical Society, Series B* **58**(4), 739–750.

World Health Organization (2000), *The World Health Report 2000: Health Systems: Improving Performance*, World Health Organization. Available at http://www.who.int/whr/2000/en/.

Beta	Mean		SD		CV		Gini		Theil	
(α, β)	Mort	Surv	Mort	Surv	Mort	Surv	Mort	Surv	Mort	Surv
(1, 10)	0.0909	0.9091	0.0829	0.0829	0.9126	0.0912	0.4761	0.0476	0.3778	0.0044
(0.5, 10)	0.0476	0.9524	0.0628	0.0628	1.3193	0.0660	0.6210	0.0311	0.6831	0.0023
(0.3, 10)	0.0291	0.9709	0.0500	0.0500	1.7160	0.0515	0.7207	0.0216	0.9864	0.0014
(0.1, 10)	0.0099	0.9901	0.0297	0.0297	3.0022	0.0300	0.8787	0.0088	1.8303	0.0005

Table 1: Income inequality measures applied to simulated beta beta(α, β) ELM distributions. The first column (α, β) gives the parameters of the beta distribution. After the first column, each pair of columns gives a particular measure applied to the probability of Mortality (Mort) distribution and the probability of Survival (Surv) distribution. Pairs of columns are the mean, the standard deviation (SD), coefficient of variation (SD/mean), Gini and Theil inequality measures.

	1975	1995
IMR	12.0%	6.0%
Mode	1.8% (1.6%, 2.3%)	1.0% (0.9%, 1.3%)
Mean	9.6% (9.2%, 10.1%)	5.2% (5.0%, 5.4%)
Median	5.4% (5.1%, 6.0%)	2.8% (2.6%, 3.1%)
Percent infants with		
$\pi_i >$ 1975 IMR	25.7% (24.2%, 27.3%)	10.8% (9.6%, 11.7%)

Table 2: Summary statistics of the densities in Figure 1. The infant mortality rate (IMR) is the empirical proportion of births that experienced mortality and is calculated from the raw data.

Covariate	Entropy	L1 Norm
Overall	0.182 (0.144, 0.230)	0.278 (0.225, 0.359)
Maternal Education	0.130 (0.100, 0.168)	0.229 (0.181, 0.294)
Maternal Age	0.169 (0.133, 0.214)	0.267 (0.215, 0.342)
Birth Interval	0.177 (0.141, 0.225)	0.275 (0.222, 0.353)
State	0.177 (0.141, 0.223)	0.275 (0.222, 0.355)
Religion	0.178 (0.142, 0.226)	0.276 (0.223, 0.355)
Gender	0.182 (0.144, 0.230)	0.278 (0.225, 0.359)
Caste	0.185 (0.148, 0.233)	0.280 (0.226, 0.362)
Birth Order	0.185 (0.148, 0.234)	0.280 (0.227, 0.362)
Residence	0.196 (0.157, 0.245)	0.289 (0.234, 0.370)
Wealth	0.234 (0.191, 0.289)	0.318 (0.259, 0.407)

Table 3: The effect of various single covariate adjustments to 1975 to see how close the distribution of ELM then gets to the distribution in 1995, as measured by the L_1 norm and entropy. Numbers in parenthesis are 95% posterior intervals.
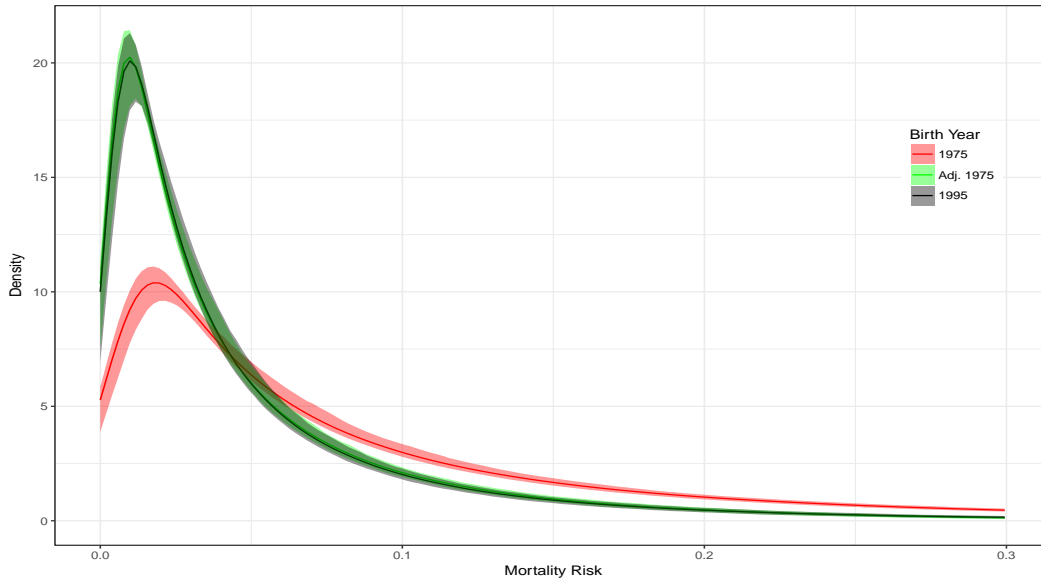
Figure 1: Kernel densities of mortality risk for 1975 and 1995 and a counterfactual adjusted 1975 population. Intervals around the main curve are 95% pointwise posterior bounds for the density height. The counterfactual 1975_A population takes the values from 1975 and multiplies by the ratio of the median risks from 1995 to 1975. The differences between 1975 and 1995 appear to be fully summarized by the median adjustment because it is almost impossible to distingush the gray live from the green line.
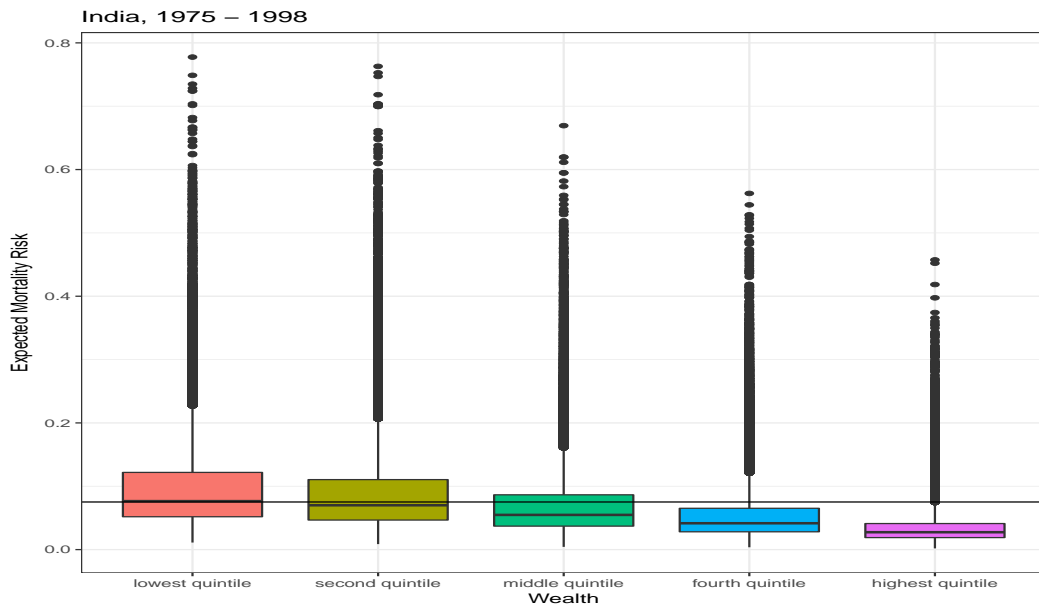
Figure 2: Distribution of mortality risk by wealth quintile in India. Each box plot summarizes the distribution of individual infant average mortality risk $E[\pi_i|Y]$ in each wealth quintile.
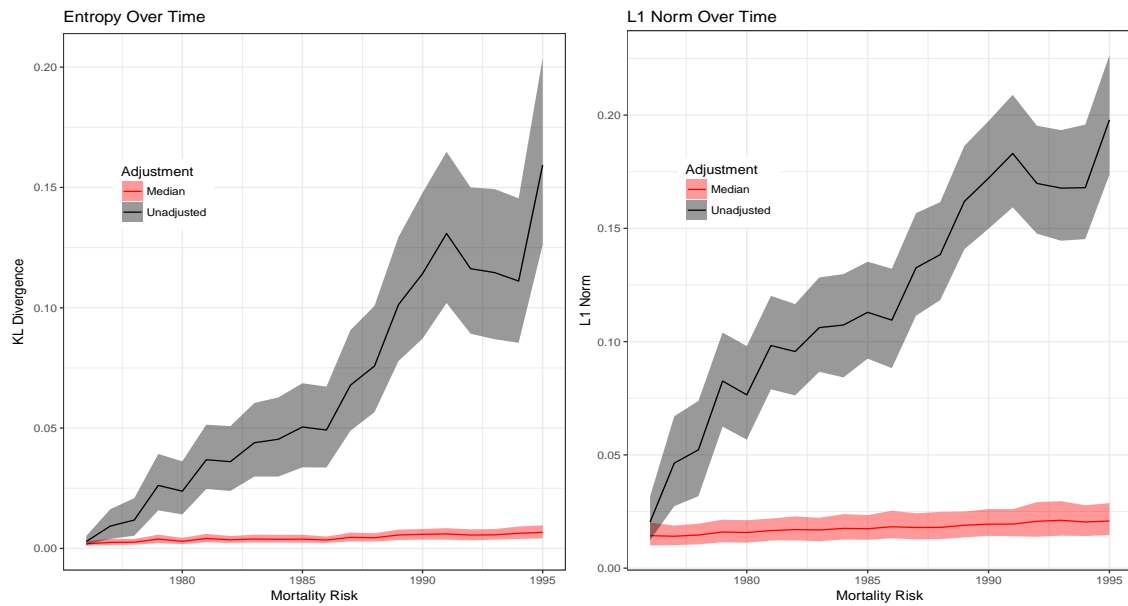
Figure 3: Inequality trends in infant mortality in India, 1975–1995, with a multiplicative median ratio adjustment in the relative distribution. The baseline year is 1975 and all subsequent years are compared against the baseline year.
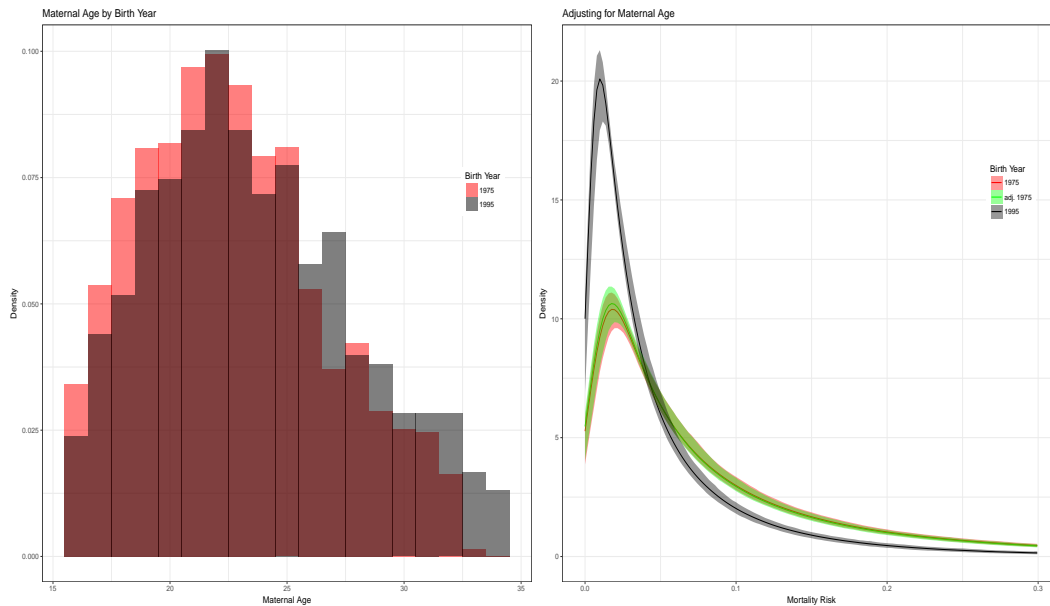
Figure 4: Distribution of maternal age and its effects on the distribution of mortality risk for 1975 and 1995. Graph on the left is a histogram with the distributions of maternal age for 1975 (pink) and 1995 (grey). The graph on the right represent three densities of mortality risk, for 1975 (green), for 1995 (pink), and 1975 adjusted to have the same distribution of maternal age that 1995 (green). Most of the differences in the distribution of mortality risk between these two years cannot be explained by the change in the distribution of maternal age.
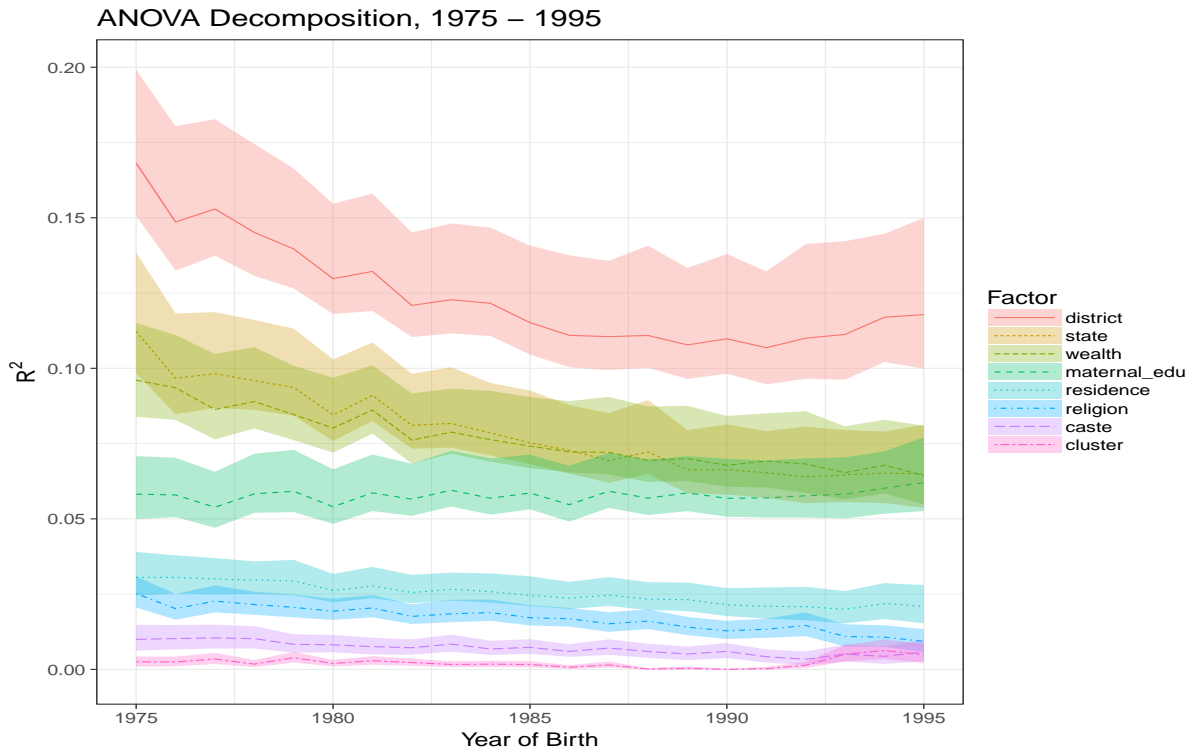
Figure 5: Fraction of mortality risk explained by individual covariates for India over time, 1975–1997. Each line represents the trend for a particular covariate, showing how much of the variability in mortality risk is due to between-group sources. The shaded areas represent 95% pointwise credible intervals.

Apendix: Priors

We use gamma priors on the precision parameters. For the mother random effects,

$$\text{mother}_j \sim N(0, \sigma^2_{\text{mother}})$$

$$\sigma^2_{\text{mother}} \sim \text{Inverse Gamma}(3, 2)$$

Similarly, for clusters,

$$\text{cluster}_l \sim N(0, \sigma^2_{\text{cluster}})$$

$$\tau^2_{\text{cluster}} \sim \text{Inverse Gamma}(3, 2),$$

For districts,

$$\begin{pmatrix} \text{district}_{k,1} \\ \text{district}_{k,2} \end{pmatrix} \sim N_2\left(\mathbf{0}_{2\times 1}, \Sigma_{\text{district}}\right)$$

$$\Sigma_{\text{district}} \sim \text{Inverse Wishart}\left(\begin{pmatrix} 1 & 0 \\ 0 & 0.1 \end{pmatrix}, 4\right)$$

For states,

$$\begin{pmatrix} \text{state}_{m,1} \\ \text{state}_{m,2} \end{pmatrix} \sim N_2\left(\mathbf{0}_{2\times 1}, \Sigma_{\text{state}}\right)$$

$$\Sigma_{\text{state}} \sim \text{Inverse Wishart}\left(\begin{pmatrix} 1 & 0 \\ 0 & 0.1 \end{pmatrix}, 4\right)$$

All fixed effects are given mean zero normal priors. The intercept term has variance

9, all other main effects have variance 1, and two-way interactions have variance 0.5. In total, there are

$$\sum_{h=1}^{2} \binom{9}{h} + 1 = 46$$

Different combinations of covariates. The categorical covariates can have differing numbers of levels, so this gives a total of $6,505$ regression parameters to be estimated.