Measuring Inequality in Early Mortality Across All Births?
A Bayesian Approach with application to the India

Antonio Pedro Ramos and Robert E. Weiss
Abstract

To date, most studies on early life survival analyze between-group comparisons across large groups of births (e.g. across countries, income groups, ethnicities, or time). However, while average between-group differences are informative, they mask important disparities in the mortality risk within groupings. The few studies that have looked at mortality across the entire population of births, however, have used inappropriate tools. We propose new methodology to measure and explain inequalities that overcome problems in previous research. First, we show that measures based on the income inequality literature, such as Gini indices, are not appropriate for measure inequality in mortality risk, because the latter is defined on the probability scale. We develop alternative approaches. We present and extend relative distribution methods and their decompositions. We also discuss which numerical summaries and statistical inference are useful for our measures. We illustrate our approach with a large data set from India, where we calculate underlying death risk for over 400,000 birth using a Bayesian Hierarchical Model. We show that our approach uncovers several unknown but important patterns in the dynamics of inequality in infant mortality. We discuss extensions and policy implications of this methodology.
1 Introduction

One of the most fundamental dimensions of social inequality is inequality in early life survival (Moser and Gwatkin 2005, Stuckler, Basu, and Mckee 2010). Recent scholarship has documented the highly varied distribution of early life survival in the world (CITE). For example, while in the richest countries, national averages of child mortality are less than 10 deaths per thousand births, these rates can be higher than 200 deaths per thousand births in poor countries. Yet, within-country disparities in child mortality can be even larger than disparities across countries. For example, large inequalities exist across income groups, where the poor in a country fare worse than the country’s rich (Victora et al. 2003, Sastry 2004, Wagstaff 2000). Studies have also documented significant child mortality inequities across other groups such as race and ethnicity (Brockerhoff and Hewett 2000, Antai 2011, Jankowska and Weeks 2013).

However, within any given group, some births may have a much higher mortality risk than others. The few studies that have measured inequality across all births find that the variation in the mortality risk among individuals from the same group (e.g. births from the same country) is often larger than the averages across groups of individuals (e.g. national averages of child mortality) (Gakidou and King 2002, Gakidou and King 2003). If that is the case, it is important to document inequalities across all births and not simply the average risk within given groups. We want to document within-groups heterogeneity. And we also want to use this information to design public policy that will not only improve group averages but also reach those at highest risk within a country.

Previous work on the measurement of inequality across all births borrowed inequality measures from the income inequality literature (Gakidou and King 2002, ?, ?). These approaches have been implemented by international organizations, for example, World Health Organization (2000) and UNICEF (2012). However, it is far from obvious that these inequality measures can be applied to death risk. The latter are measured on the probability scale and thus are bounded between in the (0, 1) interval. Income, however, is usually defined on the positive real line or, sometimes, on the whole real line. Properties appropriated for measures on the real line may not be appropriated on the probability scale. For example, the scale invariant property of inequality measures says that for any set of values $X$ and constant $C$ \( \text{ineq}(X) = \text{ineq}(X + C) \). This means that if we multiply everybody’s income by the same amount $C$ inequality remains unchanged. However, the probability scale imposes serious con-
strains to $C$ as distribution of $C \ast X$ cannot have values outside the unit interval $(0, 1)$. This forces $0 < C < 1$. Secondly, income inequality measures also tend to include the average level of income of the population under study in their denominator. This is not a problem for income, as the average income is almost always far way from zero. However, as average infant or child mortality rates approach zero, these measures may not behave well on the probability scale. To see this, consider a simple example, the coefficient of variation, a widely used measure inequality. It is defined as the standard deviation, $\sigma$, divided by the mean, $\mu$, of a set of values, such that $cv = \sigma/\mu$. For a fixed $\sigma_j$, as the mean goes to zero in a series of populations, the coefficient of variation goes to infinity $\lim_{\mu \to 0} cv \to \infty$. In substantive terms, this means that the inequality indices may increase just because mean group level of child mortality is approaching zero.

However, while measures based on the income inequality literature seems to be problematic there are many other measures that have not been used by previous literature on the topic. These indices include several divergence and similar measures from the statistical literature that can help us to quantify how much two distributions diverge. Thus we also investigate these. Finally, one of the original motivations for looking at mortality risk across all births instead of group level averages was the promise to compare the full distribution of mortality risk across populations (Gakidou and King 2002). However, this goal was never fully accomplished because previous research basically replaced one measure — the group level averages — by other indices. In line with this original motivation, we introduce and develop relative distribution methods in the context of comparing probabilities of death across. Thus while we agree that it is important to quantify inequality in health at the individual level, it remains an open question how to do it.

In this paper, we propose a methodology to measure and explain changes over time in inequality in mortality risk across all births from any given population. We present several alternative approaches to those based on the income inequality literature. We extend the relative distribution framework and their decompositions to work on the probability scale (Handcock and Morris 1999). We also introduce several inequality measures based on the statistical literature, including alternative summary measures and methods that look at the full distribution of death risk. We also show how to explain changes in inequality by creating counterfactual scenarios based on covariate adjustment techniques. Since the underlying death risk is estimated from the data, we also discuss how to propagate uncertainty from the estimation step of the analysis to the inequality step. Finally we also discuss inferential issues
such as how to calculate that the probability of inequality in one year is higher than another year.

We use a large data set from India to illustrate our methodology. India is a large developing democracy, where inequality in early mortality is still a problem (UNICEF 2012). In India large disparities exist: Bhattacharya and Chikwama (2012) look at inequality in child mortality in India, at the district level, and found an increase concentration of child mortality in the less developed districts; however, they also documented that inequality in female child mortality has declined. Singh (2011) investigate spatial disparities in child mortality and found that while economic status in becoming less important at explaining child malnutrition and child survival, the effect of mother’s education has actually become stronger over time. De and Dhar (2013) focus on disparities across states and found that inequality in child mortality is more concentrated in the comparatively developed states than the poorer states in India. (Nidhi Jain 2013) focused on trends in economic inequalities with respect to infant and child mortality suggesting a decreasing trend in economic inequality in infant mortality but an upward trend in economic inequality in child mortality in India. Kumar and Sing (2014) finds that infant mortality in rural India is substantially higher than that in urban India. A recent study blame India inequalities in early mortality for its inability to reach the Millennium Development Goals target to early death mortality (UNICEF 2012). All of these studies, however, look at the between-group inequality; none of them look at mortality within-groups, largely assuming within-group homogeneity in the death risk.

Using data from India, we calculate the underlying mortality rate of each birth in our data and associated measures of uncertainty using Bayesian hierarchical models for binary outcomes fit via Markov Chain Monte Carlo (MCMC). We use Bayesian estimation and MCMC methods because (1) they can handle complex models and (2) they facilitate the propagation of model uncertainty into subsequent analysis of inequality. We apply hierarchical models because most demographic surveys have grouping level information available (e.g. mothers, sampling clusters, districts, states, etc). Using in-sample predictions from the model, we describe, characterize and decompose changes in inequality over time.

We investigate a largely untested assumption in the literature on inequality in early mortality: that most of the inequality in early mortality exists between groups of birth. We group births into several traditionally defined groups: religion, caste, maternal education, income, and states. Contrary to the conventional wisdom, we find that most of the variability in death risk comes from within-groups, not between-groups. This suggests that, by focusing on
between-group comparisons only, we may miss individuals within these groups that are not otherwise considered in policies solely based on group average levels. Finally, we show that these patterns are not explained by changes in the demographic composition of the population. None of these patterns can be seen if only traditional methods are used.

This paper makes several important contributions to the existing literature on inequality in early mortality. First, we show that existing approaches based on income inequality literature are not well suited to measure inequality in mortality risk. This is particularly important because the World Health Organization currently implements an approach that we show to be problematic. Second, we present several alternatives from the statistical literature. Within relative distribution framework, we refine the existing decomposition approaches to handle distributions on the probability scale. Our treatment in the context of Bayesian inference is also novel. Third, our case study of India does show that measuring inequality across all births are important because in fact most of the variance in mortality risk exist within groups of births, not between them. This finding challenges the conventional wisdom that most groups (e.g. socioeconomic) are somewhat homogeneous. It also has public policy implications as it suggest that better results might be achieved by targeting high risk births (e.g. regardless of their socioeconomic status) as opposed to predefined groups. Finally, we discuss some extensions of our methodology.

First we explain why most methods from the income inequality literature do not work on mortality risk. We illustrate the problem with a constructed example. Then we introduce several alternatives based on the relative distributions methods and the statistical literature. All of these methods are designed to describe inequality across populations. Thus we introduce covariate adjustment methods as a step toward explanatory modeling. We also discuss how Bayesian inference makes possible probability statements about inequality such as what is the probability that inequality in one population is higher than in another. To illustrate our methodology, we provide an example from a large data set from India. We discuss limitations of this work and further research at the end.
2 Measuring and Explaining Inequality in Early Life Across All Births

In this section we discuss how to quantify and explain inequality in early mortality. First we show that ratio based measures from the income inequality literature are not appropriated for mortality risk. Next we introduce several alternative approaches. We start with relative distributions methods that allow for comparing the full distribution of mortality risk across entire populations and thus are closely related to the motivation of the early literature on inequality (Gakidou and King 2002). We then discuss how to compare decompose distributions of mortality risk with the objective of parsimoniously characterize differences across distributions. Following, I discuss how to use ANOVA to decompose the variance in within and between-group components for several groups. Then we discuss several numerical summaries that are appropriated for probabilities, some of which are based on the relative distribution methods and some that are based on the statistical literature. Finally, we discuss how to use covariate adjustment techniques to generate counterfactual scenarios with the object of explaining different of inequality across population. In the following section, we illustrate all these methods with our case study of India.

2.1 Inequality Indices based on Ratios

We now discuss measures of inequality that are drawn from the literature on income inequality, some of which have been applied to the health literature. These measures can be derived from a common formula but different measures capture different aspects of the distributional change. These is useful, because by using a common formula, we can see the common limitations of these measures to the probability scale. As shown by Firebaughah (2002), these indexes differ only because they employ different functions for the income ratios. The general formula for a population of 1, ..., n births with death risk given by \( \pi_1, ..., \pi_n \) is

\[
\text{Inequality} = \frac{1}{N} \sum_i f(r_i), \quad \text{where} \\
r_i = \frac{\pi_i}{\frac{1}{n} \sum \pi_i}, \quad \text{and} \\
S = \frac{1}{n} \sum \pi_i 
\]

where \( r_i \) is the ratio of the death risk for the \( i \) child to the average birth death risk in the population in that sample, \( S \) Based on this general approach, we can consider four popular
measures of inequality for the $f(\cdot)$: squared coefficient of variance, variance of the logs, Theil, and Gini index respectively,

\begin{align*}
    f_1(r_i) &\equiv v_i = (r_i - 1)^2 \\
    f_2(r_i) &\equiv t_i = r_j \log(r_i) \\
    f_3(r_i) &\equiv l_i = \log\{r_i - E[\log(r_i)]\}^2 \\
    f_4(r_i) &\equiv g_i = r_i(q_i - Q_i)
\end{align*}

where $E$ is the expected value, $\log$ is the natural logarithm, $q_i$ is the proportion of total population in units with higher death risk than unit $i$ and $Q_i$ is the proportion of total population in units with lower risk than unit $i$ so that $q_i + Q_i = 1$. A key point to note across all these indices is that in $r_i$, $\pi_i$ is being divided by the population average mortality risk, $\frac{1}{n} \sum_i \pi_i$, so that as average probability of early death approaches zero, $r_i$ might increase, thus giving an impression that inequality is increasing.

### 2.1.1 World Health Organization Measure

This measure has been proposed to specifically measure inequality in early death across all births. It is also similar to the income-based measures as it also divides by population mean death risk. The formulae from Gakidou and King (2002) is given by

\[ \left( \frac{1}{2n^2} \right)^3 \left( \frac{1}{n} \sum_i \pi_i \right)^3 \]

where $r_i$ and $r_j$ are the mortality risk for individuals $i$ and $j$. As we can see, this measure also divides by the mean death risk in the population, which may also be problematic as the average probability approaches zero.

### 2.1.2 A problem with ratio based measures

We design a constructed example to investigate how ratio based inequality measures work on the probability scale. Consider the following property

- **Symmetry**: If $Z$ has a distribution on $[0, 1]$ then $1-Z$ should have the same measure of inequality.

This sheds light on the fact that any inequality measure should produce the same results whether we are measuring probability of mortality ($\pi_i$ probability of death) or probability of
survival \(1 - \pi_i\). For example, for a country with high inequality in infant mortality it should also have high inequality in infant survival. We are measuring the same quantity but using alternative definitions so that the results should not be dependent on the choice of indicator. We use simulated data from a series of beta distributions with decreasing means that resemble mortality risk distributions to test which measures satisfy the Symmetry property as average mortality rates approaches zero and what happens to the measures as average mortality risk decreases toward zero.

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>(1,10)</td>
<td>0.089</td>
<td>0.911</td>
<td>0.084</td>
<td>0.084</td>
<td>0.4789</td>
<td>0.0480</td>
</tr>
<tr>
<td>(0.5,10)</td>
<td>0.048</td>
<td>0.952</td>
<td>0.063</td>
<td>0.063</td>
<td>1.315</td>
<td>0.066</td>
</tr>
<tr>
<td>(0.3,10)</td>
<td>0.029</td>
<td>0.98</td>
<td>0.05</td>
<td>0.05</td>
<td>1.696</td>
<td>0.05</td>
</tr>
<tr>
<td>(0.01,10)</td>
<td>0.009</td>
<td>0.999</td>
<td>0.009</td>
<td>0.009</td>
<td>3.038</td>
<td>0.029</td>
</tr>
</tbody>
</table>

Table 1 – Simulated data to test the ratio based measures.

Table 1 presents results for four populations of probabilities as described by four beta distributions, Beta\((a, b)\), where \(b\) is fixed at 10 for all four distributions and \(a\) decreases from 1 to .5, .3 and .01. As we move across the beta distributions, the mean levels declines, which represents the over time trends we observe. The variance is not affected whether we look at mortality or survival. However, the coefficient of variation displays very different levels of inequality for mortality and survival and the size of the difference increases as the mean levels declines. This pattern is similar to what we see for the other inequality measures. The conclusion is that those indices do not satisfy the symmetry property and therefore are not suited to asses inequality in probability distributions. We now look at other methodologies to measures and explain inequality in child mortality.

### 2.2 Comparing Distributions of Mortality Risk

One of the original motivations for looking at mortality risk across all births was the promise to compare the full distribution of mortality risk across populations (Gakidou and King 2002). However, this goal was never fully accomplished because previous research replaced the group average mortality risk by other numerical summaries. As discussed by Gakidou and King (2002), however, populations with similar numerical summaries may have very different distributions of mortality risk. For example, populations with the same mean...
or variance in mortality risk but different levels of polarization. In particular, numerical summaries tell us very little about where the differences in two distribution come from. For example, does the average mortality risk in one year is higher than in the other because most births have higher death risk of because of a fewer number of very high risk births. Is one year more polarized than another? What is the difference among the top 10% high risk births? To answers these questions it is ideal to use some methodology that actually contrast two distributions across every possible quantile.

Relative distribution methods (Handcock and Morris 1999) allow us to compare full distribution of mortality risk so that we can precisely quantify at where and how two distributions diverge. Thus it is closely related to the original motivation of earlier works. Although relative distribution methods do not numerical summaries but it an useful complement to numerical summaries. It is framework that combines graphical tools, summary measures and provides statistical inferences. They can be used to compare inequality in infant mortality across different groups. The framework is thus an important complement for more traditional numerical summaries that contrast distributions based on a single number. However, there are also measures based on the relative distribution — the polarization indices — and measures that can be used in conjunction with relative distributions.

Essentially, relative distributions compare two distributions, a reference distribution, $F_0$, and a comparison distribution, $F$ by creating a new random variable. The new random variable is a ratio of two cumulative density functions (CDF), the reference CDF and the comparison CDF. Let $Y_0$ be a random variable representing the distribution of child mortality risk across the reference group and let $Y$ be a similar random variable for a comparison group with cumulative density functions $F_0$ and $F$ with probability densities are $f_0$ and $f$, respectively. Following Handcock and Morris (1999), the relative distribution $R$ of the comparison group to the reference group, $\Pi_0$ to $\Pi$, is the random variable created by the transformation.

$$R = F_0(\Pi)$$

The random variable $R$ has values $r$ with $r \in (0, 1)$. $R$ tells us about the percentile rank in which a value $r$ for a death risk from the comparison group has in the reference group. It tells us how more or less frequent any range of values from the comparison distribution would
have in the original distribution. $G(r)$ is a monotone transformation and thus invertible.

\[
G(r) = P_r(R \leq r) = P_r(F_0(Y) \leq r) = P_r(Y \leq F_0^{-1}(r)) = G(F_0^{-1}(r)) = G(Q_0(r))
\]

where $F_0^{-1}$ is the quantile function of $Y_0$. The PDF of $R$ can also be easily derived by using the definition of pdf and the chain rule and the inverse function theorem from calculus:

\[
g(r) = \frac{d}{dr} G(Q_0(r)) = \frac{d}{dr} G(F_0^{-1}(r)) = G'(F_0^{-1}(r)) \times (F_0^{-1})'(r) = f(F_0^{-1}(r)) \times \frac{1}{f_0(F_0^{-1}(r))} = \frac{f(G_0(r))}{f_0(G_0(r))}
\]

This PDF, the relative density, is a ratio of densities, evaluated at a given quantile of the reference distribution. The relative density can be interpreted as a density ratio, the % rank that the original comparison group would have been in the reference group.

\[
g(r) = \frac{f(y_r)}{f_0(y_r)},
\]

where $y_r \geq 0$ is the $r^{th}$ quantile of $R$. A few remarks are in order. Relative distributions are invariant to monotone transformations (Handcock and Morris 1999). Suppose $M(\cdot)$ represent any monotone function such that $Y_0^T = M(Y_0)$ and $Y^T = M(Y)$, then

\[
R = F_0(Y) = F_0(Y^T)
\]

Equation 21 is useful because we can work on either the logit, probit or probability scale or any monotone transformation of those and get the same answers.

To detect changes in the distribution of mortality risk over time, we can compare distributions of mortality risk from one year to another and ask questions such as, “at which quantile
on the reference year’s distribution would someone from the comparison year fall?”. If the distributions are the same, we expect this relative data, $r$ to be uniformly distributed. Deviations from uniformity provide insights into group differences because we can quantify differences between two populations at any point of the distributions, thus offering an extremely flexible approach for studying health inequalities. Relative distribution works in any scale.

### 2.2.1 Decomposing Distributions of Mortality Risk

In many scientific applications it is useful to summarize differences across distributions based on a few parameters (Handcock and Morris 1999). For example, suppose we are comparing over time changes in the entire distribution of mortality risk in a population. Does the changes in the median of the distribution represents the experience of the most of the population? Or, instead, do we have a scenario of increasing polarization of mortality risk where reduction on the mortality risk across years varies by quantile? Alternatively, we can also investigate the remaining differences between two distributions once a simple shift is taken into account. By doing so, we could, for example, identify subsets of births that did not experience the same over time change of most births, which would lead to further analysis and investigations.

We can use relative distribution methods to think about these decompositions. A comparison distribution $F(X)$ is a location-shifted version of the reference distribution $F_0(X)$ when the differences between them can be summarized by an additive or, $F(X) ≡ F(X + C)$, or a multiplicative shift $F(X) ≡ F(X \times C)$, for some constant $C$. In this case the difference between two distributions can be parsimoniously summarized by the shift, $C$. Essentially, we are creating a new distribution, $F_A(X)$ that represent a counterfactual or synthetic population. Throughout this paper we will adjusted the reference distribution $F_0(X)$ but the adjustment can be done in both direction. In most cases simple shifts won’t be able to characterize all of the differences across two distributions. Thus in general it is not the case that $F_A(X) ≡ F_0(X)$. However, we can quantify the remaining differences between $F_0(X)$ and $F_A(X)$ after the adjustment. In the probability scale, additive shifts are problematic for $F_A(X)$ having full support on $(0,1)$ as they produce distribution with support outside the bounds. Multiplicative shifts with $C < 1$ does produce $F_A(X)$ with support within the bounds. We discuss decompositions that are useful in the probability scale.

More formally we are now working with 3 random variables, the reference distribution $Y_0$, the comparison distribution, $Y$, and adjusted distribution $Y_A$, which describes a hypothetical
population after the adjustment. Handcock and Morris (1999) that the overall relative distribution can be decomposed into two effects where $y_r$ is the $r^{th}$ quantile of $R$ defined earlier. We have 3 comparisons

$$\frac{f(y_r)}{f_0(y_r)} = \frac{f(y_r)}{f_A(y_r)} \times \frac{f_A(y_r)}{f_0(y_r)}$$ \hspace{1cm} (23)

The left hand side is the overall relative distribution without any adjustments. The second ratio compares the adjusted reference distribution, $Y_A$ with the comparison distribution, $Y$. If the differences between $Y_0$ and $Y$ are fully summarized the the adjustment, $C$, then the ratio will be 1. The third ratio contrasts the adjusted density $f_A$ with the unadjusted density $f_0$.

### 2.2.2 Multiplicative Adjustment

A natural adjustment on the probability scale is the multiplicative adjustment. Define $Y_A = Y \times \rho$ where $\rho = \text{location}(Y_0) / \text{location}(Y)$. We shall say that $Y_A$ is the location adjusted reference distribution with CDF $F_A$ is $F_A(y) = F(y \times \rho)$. However, because we are working in the probability scale we have some restriction on the decompositions. In particular, we now that the adjusted distribution $Y_A$ need to be on the range $(0, 1)$ and thus it cannot be multiplied by any number. In particular, $0 \leq Y_A = Y_0 \ast \rho \geq 1$. The adjustment factor is always positive, $\rho \geq 0$ thus we only need to impose the additional restriction that location($Y_0$) $\geq$ location($Y$). When making over time comparison for the same population, this is not a major restriction because the decline in early mortality over time. However, in other situations this restriction might be a severe limitation for the analysis.

### 2.2.3 Exponential Adjustment

Another approach is to do the adjustment is another scale and then transform the distribution back to the probability scale. An additive decomposition in a complementary log scale is equivalent to a power transformation in the probabilities. Define the original scale $Y \in (0, 1)$, and transformations $Z = \exp(Y) \in (0, \infty)$ and $X = -\exp(Z) \in (-\infty, \infty)$ so that $X = \exp(-\exp(Y))$, the complementary log transformation of the original distribution which are now defined in for the real number line. In general, it is not possible to perform an additive adjustment on the probability scale but it is possible to do so in the co-log scale because it is
defined in the real number line. Thus $X_A = X_0 + \rho$ where $\rho = \text{location}(X_0) - \text{location}(X)$

$$Y_A = \exp(-\exp(X_0 + \rho)) = \exp(-(\exp(X_0))) \times \exp(\rho))$$

(24)

$$= \exp(-\exp(X_0))) \times \exp(\rho))$$

(25)

This approach has the advantage of not having the restriction and thus it has broader applicability. However, which one will work best will depend on a particular application.

### 2.3 Anova to decompose within and between-group inequality

An important motivation for quantifying inequality in earlier mortality across all births was the suspicion that within-group variation is significantly higher than the between-group variation. If the within-group differences are higher than between-group then previous analysis between-group comparisons missed where most of the variability exist. However, previous studies have not quantified the variation between and within groups. Many inequality indices from the income inequality literature decomposable into between and within group components. We suggest to focus on the variance, which is a proper measure for probabilities. We implement an analysis of variance (ANOVA) to decompose within and between-group in early mortality. Do to so, we fit linear regression models using mortality risk as an outcome and group membership as a predictor. These approach allow us to quantify how much of the variance is coming from any group membership or combination of group membership.

### 2.4 Numerical Summaries

Numerical summaries are standard procedures to compare distributions in inequality analysis. They can be used in conjunction with relative distributions methods and they make it possible to quantify differences according to certain metrics. We discuss summary measures based on relative distribution, income inequality literate, and the statistical literature. In the following section we investigate which measures are appropiated for the probability scale where mortality risk is defined.

#### 2.4.1 Measuring Polarizational Distribution

Polarization indices are based on the adjusted relative distribution, $Y_A$. They measure deviations of the adjusted relative distribution from the uniform distribution. They are similar to measures such as the median absolute deviation. By doing so, these indices quantifies
changes in the tails of the distribution instead of focus on movements in the middle of the distribution. We define the median polarization indices (MPR) of $Y$ relative to $Y_0$ as

$$M_{RP}(F; F_0) = 4 \int_0^1 |r - \frac{1}{2}|g_0^A(r)dr - 1,$$

(26)

Where $g_0^A(r)$ is the pdf of the adjusted relative distribution.

This measure weighs the mass in the upper and lower tails more heavily than the mass in the center. It is the mean absolute deviation of the location-matched relative distribution, scaled to produce an index that varies between -1 and 1. Positive values represent more polarization, i.e., increases in the tails of the distribution, and negative values represent less polarization, i.e., convergence towards the center of the distribution. If the location-matched relative distribution is uniform, then the index will be zero.

$$L_{RP}(F; F_0) = 8 \int_0^{\frac{1}{2}} |r - \frac{1}{2}|g_0^A(r)dr - 1,$$

(27)

$$U_{RP}(F; F_0) = 8 \int_{\frac{1}{2}}^1 |r - \frac{1}{2}|g_0^A(r)dr - 1,$$

(28)

Thus, $M_{RP}(F; F_0) = .5U_{RP}(F; F_0) + .5L_{RP}(F; F_0)$

(29)

### 2.4.2 Measuring Distributional Divergence

Suppose we have two densities that we wish to compare, $F$ and $F_0$. It is often useful to quantify how much one distribution differs from another with a single quantity rather than a graph. For example, we want to measure the divergency between two distribution before and after the relative distribution adjustment to learn about how much of two distributions can be parsimoniously explained by a single parameter. A commonly used measure is the Kullback-Laibler divergence

$$K(F; F_0) = \int_{-\infty}^{\infty} \log \frac{f(x)}{f_0(x)}dF(x) = \int_0^1 \log(g(r))(g_0(r))dr,$$

(30)

can be interpreted as the expected information for discriminating $g(r)$ from the uniform distribution based on a single observation from $Y_0$. Details of these measure and its relation with relative distribution methods can be found (CITE). Another useful measure is $L_1$ norm

$$L_1(F; F_0) = \frac{1}{2} \int_0^1 |F(\pi) - F_0(\pi)|d\pi,$$

(31)

$L_1$ Norm quantifies how much probability mass needs to be moved so that one distribution become identical to the other one (CITE).
2.5 Explaining Inequality

So far, we have focused on describing inequality in mortality risk. However, we also want to explain it. We want to understand the key drivers of these observed changes. For example, what are the key factors behind changes in inequality over time? Are inequality changes driven by shifts in the demographic composition of the population, or by changing in the covariate-outcome relationship? To understand these changes the key is to construct counterfactual scenarios with hypothetical or synthetic populations so that we can create some new Y’s to separate out changes in the composition of the population from changes in the covariate-response relationship between Y and Y. For example, the age of the mother at the birth of the child is an important predictor of early life survival, as births from young mothers have higher mortality risk. Thus as the distribution of the age of the mother changes in the population, we should expect changes in mortality levels and possibly in inequality levels. For example, higher dispersion in the age of the mother at birth should be associated with higher levels of inequality. However, the relative importance of the age of the mother for early life survival may also change. For example, as medical technology and sanitation disseminates, the age of the mother at the birth of the child should become a less important factor in infant survival. However, both changes often happen simultaneously and thus we need a statistical methodology to distinguish each one of these effects. We describe the basic approach looking at a single variable at a time.

The basic ideas come from probability rules relating conditional and marginal effects. Now we are working with random vectors, were Y is the mortality risk and Z is the covariate we want to adjust for:

\[ P(Y = y) = \int P(Y = y | Z = z) P(Z = z) dz \]  

(32)

Define the random vectors for the reference and comparison populations, \((Z_0, Y_0)\) and \((Z, Y)\) respectively. \(Y_0\) and \(Y\) are the distribution of mortality risk for two different populations and \(Z_0\) and \(Z\) are the covariates we want to make an adjustment for. For simplicity of the exposition, let’s consider the discrete case. In the discrete case, assume that the support for both \(Z\) and \(Z_0\) is \(\{1, 2, \ldots, K\}\) and that their probability mass functions are given by \(\{\pi_k\}_{k=1}^{K}\) and \(\{\pi^0_k\}_{k=1}^{K}\). These mass functions represent populations compositions with respect to \(Z\) and \(Z_0\). Define densities \(f_{y|z}(y|K)\) and \(f_{y_0|z_0}(y|K)\), which are the conditional effects defining the covariate-response relationships for \((Y, Z)\) and \((Y_0, Z_0)\). We can write the marginal densities
of $Y$ and $Y_0$ as

$$f_0(y) = \sum_{k=1}^{K} \pi_k f_{Y_0|Z_0}(y|k), \text{ and}$$

$$f(y) = \sum_{k=1}^{K} \pi_k f_{Y|Z}(y|k) \quad (33)$$

Now suppose we want to create a counterfactual population $f_{0C}$ with the same conditional distribution of reference population $(Y_0, Z_0)$ but with the population composition of the comparison group $(Y, Z)$. We define such a population as having probability mass function $\pi_k^0$ but density $f_{Y|Z}(y|K)$

$$f_{0C}(y) = \sum_{k=1}^{K} \pi_k^0 f_{Y_0|Z_0}(y|k) \quad (35)$$

This can be extended to the case of continuous variable. Assume the covariate $Z$ is continuous with density $f_Z(z), z \in \mathbb{R}$

$$f_{0C} = \int f_Z(z) f_{Y|Z}(y|k) dz \quad (37)$$

Now we can re-express the adjusted distributions, $f_{0C}$, in terms of the relative distributions

$$f_{0C}(y) = \int_0^1 g_Z(r) f_{Y_0|Z_0} Q_{Z_0}(r) dr \quad (38)$$

where $g_Z(r)$ is pdf of the relative distribution of $Z \rightarrow Z_0$ and $Q_{Z_0}(r) = F_{Z_0}^{-1}(r)$ is the quantile function for $Z_0$.

### 2.6 Statistical Inference for Inequality Measures

Consider that the mortality risk can be estimated from Bayesian models solved by MCMC numerical methods. Thus distributions of mortality risks can be seen as non-standard inferences targets from MCMC samples, being non-linear complex functions of the parameters and data. Thus we use posterior and predictions from the Bayesian hierarchical model to calculate posterior distributions of inequality indices and relative distributions. We can calculate the posterior probability that a given index is higher in one population than another by calculating the proportion of MCMC samples where the one population is higher than in another population. Doing so, we also propagate the uncertainty in estimating parameters into inequality analysis.
3 Application: A Case study of India, 1975-1998

3.1 Data

Our data set for India come from the Demographic and Health Surveys, (DHS) (http://www.measuredhs.com/), a nationally representative surveys. We use data from 2 DHS surveys to construct a retrospective panel from 1975 to 1997. A third wave of the survey, covering more recent years, was available. We did not use it because it did not included district level information, making it impossible to estimate district random effects. We analyze infant mortality from births from mothers aged 15-35 between 1975 and 1998 to avoid censoring effects.

We have a total of 408,706 births from 141,999 mothers in 3,806 sampling clusters taken from 443 districts and 26 states. In our model, we use covariates at the child level (birth order, gender, maternal age, religion, caste) and household level (wealth index, maternal education). We have 4 clustering covariates, which we assign to random effects: mothers, sampling clusters, districts and states, all nested one inside the next. Sampling cluster are small geographic units used for sampling purposes.

<table>
<thead>
<tr>
<th>cluster level</th>
<th>sample size</th>
</tr>
</thead>
<tbody>
<tr>
<td>births</td>
<td>408,706</td>
</tr>
<tr>
<td>mothers</td>
<td>141,999</td>
</tr>
<tr>
<td>sampling clusters</td>
<td>3,806</td>
</tr>
<tr>
<td>districts</td>
<td>443</td>
</tr>
<tr>
<td>states</td>
<td>26</td>
</tr>
</tbody>
</table>

Table 2 – Sample size for each clustering level.

3.2 Estimation of Mortality Risk

We use Bayesian logistic random effects models to predict children’s survival, using covariates from a variety of levels, including birth level (gender, age of the mother at birth of the child, birth order, birth interval), and mother level (wealth, and education of the mother). The effect of classical predictors of child mortality can be estimated directly from the data and models can have random effects at several levels of the data simultaneously. Time trends, time shocks, and time varying effects of key covariates can be accommodated in these models.
3.2.1 Model Specification

Let \( i \) index children, \( m \) index mothers, and \( c \) index (sampling) clusters, \( d \) index districts, and \( s \) index states, with child nested within mother, nested in cluster sampling, nested within districts, nested with states, such that we have \( i = 1, ..., n_{mcds} \) children for each mother-cluster-district combination and similarly, \( m = 1, ..., M_{cds} \) mothers, and \( c = 1, ..., C_{ds} \) sampling clusters, \( d = 1, ..., D_s \) districts, \( s = 1, ..., S \) states, and \( t \) indexes years. Then the model is

\[
Y_{imcds}|\pi_{imcds} \sim \text{Bern}(\pi_{imcds}) \quad (40)
\]

\[
\text{logit}(\pi_{imcds}) = \beta_0 + x_{imcds}'\alpha + \delta_{mcds} + \gamma_{cds} + \xi_{ds} + \tau_s \quad (41)
\]

\[
\delta_{mcds} \sim N(0, \sigma^2_1) \quad (42)
\]

\[
\gamma_{cds} \sim N(0, \sigma^2_2) \quad (43)
\]

\[
\xi_{ds} \sim N(0, \sigma^2_3) \quad (44)
\]

\[
\tau_s \sim N(0, \sigma^2_4), \quad (45)
\]

where

- \( Y_{imcds} \) is the response variable, whether child \( i \) born in year \( t \) from mother \( m \), in sampling cluster \( c \), from district \( d \), and state \( s \) was alive at 1 year old \( Y_{imcds} = 0 \) or not \( Y_{imcds} = 1 \).

- \( \pi_{imcds} \) is the unobserved mortality probability of the \( imcds \) child.

- \( x_{imcds}' \) is a vector of covariates.

- \( \alpha \) is a vector of fixed effects corresponding to the covariates in \( x_{imcds} \).

- \( \delta_{mcds} \) is mother random effect with variance \( \sigma^2_1 \).

- \( \gamma_{cds} \) is the sampling cluster random effect with variance \( \sigma^2_2 \).

- \( \xi_{ds} \) is the district random effect with variance \( \sigma^2_3 \).

- \( \tau_s \) is the state random effect with variance \( \sigma^2_4 \).

Probability \( \pi_{imcds} \) is the key quantity of interest from the model. It will be used later as input to inequality calculations. Covariates are by each level: child: sex, birth order, maternal age at birth of child, and gender of the child; mother and household levels: wealth and maternal education. Mothers random effects \( \delta_{mcds} \) accounts for correlation between children with the same mother and unmeasured variables at the level of the mother and household. Similarly, \( \gamma_{cds}, \xi_{ds} \), and
\(\tau_s\) account for correlations between observations in the same geographic location and unmeasured variables at the geographic level. \(\xi_{ds}\) accounts for correlations between observations in the same district or state and unmeasured variables at the district or state level with variances is \(\sigma^2_3\) and \(\sigma^2_4\) and \(\tau_s\).

### 3.2.2 Specification of Time Trends and Time-Varying Covariate Effects

As our scientific goal is to longitudinally model changes in probability of death across all births, we need to allow for time-varying effects. We sketch our general approach to model specification for time trends. We use spline to capture non-linearities in the time trend. Then we investigate whether covariates have time-varying effects. We experiment with (usually linear) time by covariate interactions and, as necessary, Generalize Additive Models (GAM) for the covariate effects.

### 3.2.3 Bayesian Estimation, Prediction and Priors

We fit the model in a Bayesian framework with MCMC methods. It is easy to get estimates, predictions and measures of uncertainty (Gelman and Hill 2006) from these methods. Random effects and sampling are easily accommodated and we can generate predictions and non-standard inferential quantities of scientific interest. Predictions from the MCMC output are used to make inferences about mortality risk across the entire population of births. We use priors that provide relatively little information about the location of fixed effects parameters. Thus for scientific purposes, our results are numerically comparable, though not identical, to maximum likelihood estimates. We make in-sample predictions. Thus for each child in our dataset we generate a posterior distribution of the mortality risk, \(\tau_{imeds}\).

### 3.3 Results from the statistical model

Our model indicates that the main time trends is non-linear: our exploration suggest a model for the main time trend with a cubic spline model with the knots at the following years (1978, 1981, 1985, 1986, 1990, 1992). A few covariates had linear time-varying effects: wealth, religion, and caste. We also find that the age of the mother at the birth of the child has a quadratic effect on mortality.

After deciding on the functional form of the fixed effects predictors, we fit increasingly complex Bayesian hierarchical models using MCMC methods. We progressively add random
effects for mothers, sampling clusters, districts, and states. The final random effects model includes random intercepts for mothers, sampling clusters, district and states. We also had non-linear random slopes for time by state. We use standard methods to check for convergency of the chain. We have use DIC and Deviance to help decide among models. The results are reported in table ???. From model 1 to 5 we include additional random intercepts, one for each level of clustering. Model 6 include random effects bent lines by state. The inclusion of them all reduce both DIC and Deviance statistics, indication that these clustering levels do provide additional information for the model.

As we can see in figure 1, the results for the regression coefficients are quite consist across models. We briefly discuss the fixed effect results from model 6. The intercept represents an average infant mortality rate for a female, from rural area, lowest quintile of income, hindu religion, from 'other backward caste’ caste, with no education, who was born from a mother aged 18 years old in 1975. A birth with these characteristics has it a posterior probability of death of 0.07\[0.05, 0.11\]. Income does has a large effect, even after account for all other factors. From the poorest to the richest, the second quintile is not significantly different from the baseline case. However, the middle quintile has a mortality risk of .063\[0.058, 0.07\]; the fourth quintile has a mortality risk of .055, \[0.05, 0.06\]; and the highest quintile, a mortality risk of .03\[0.03, 0.04\]. Maternal education also have a very strong independent effect on mortality risk. Primary education is associated with a change in the mortality risk of .064\[0.061, 0.066\]; to

<table>
<thead>
<tr>
<th></th>
<th>Model 1</th>
<th>Model 2</th>
<th>Model 3</th>
<th>Model 4</th>
<th>Model 5</th>
<th>Model 6</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mothers</td>
<td>NA</td>
<td>1.28</td>
<td>0.17</td>
<td>0.09</td>
<td>0.09</td>
<td>0.09</td>
</tr>
<tr>
<td></td>
<td>( NA, NA)</td>
<td>( 1.22, 1.34)</td>
<td>( 0.14, 0.19)</td>
<td>( 0.08, 0.11)</td>
<td>( 0.08, 0.11)</td>
<td>( 0.08, 0.11)</td>
</tr>
<tr>
<td>Cluster</td>
<td>NA</td>
<td>NA</td>
<td>1.19</td>
<td>1.15</td>
<td>1.16</td>
<td>1.16</td>
</tr>
<tr>
<td></td>
<td>( NA, NA)</td>
<td>( NA, NA)</td>
<td>( 1.14, 1.25)</td>
<td>( 1.10, 1.21)</td>
<td>( 1.10, 1.21)</td>
<td>( 1.10, 1.21)</td>
</tr>
<tr>
<td>District</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
<td>0.16</td>
<td>0.09</td>
<td>0.10</td>
</tr>
<tr>
<td></td>
<td>( NA, NA)</td>
<td>( NA, NA)</td>
<td>( NA, NA)</td>
<td>( 0.13, 0.19)</td>
<td>( 0.08, 0.11)</td>
<td>( 0.08, 0.11)</td>
</tr>
<tr>
<td>States</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
<td>0.40</td>
<td>1.69</td>
</tr>
<tr>
<td></td>
<td>( NA, NA)</td>
<td>( NA, NA)</td>
<td>( NA, NA)</td>
<td>( NA, NA)</td>
<td>( 0.22, 0.62)</td>
<td>( 0.90, 2.68)</td>
</tr>
<tr>
<td>Deviance</td>
<td>204,829</td>
<td>178,955</td>
<td>178,708</td>
<td>178,491</td>
<td>178,389</td>
<td>178,245</td>
</tr>
<tr>
<td>DIC</td>
<td>229,622</td>
<td>217,542</td>
<td>217,062</td>
<td>216,328</td>
<td>216,226</td>
<td>216,184</td>
</tr>
</tbody>
</table>

Table 3 – Random Effects and Godness of Fit Statistics for Models 1-6. Model 1-5 includes random intercepts only. Model 6 also includes random bent lines for time trends by states.
secondary education of \(0.053 [.051, 0.056]\); and to higher education of \(0.037 [.033, 0.041]\). All other religious groups have substantially lower mortality than hindus: holding everything else constant, a birth from a muslim mother has a mortality risk of \(-0.062 [.056, 0.067]\); from a Christian mother, the mortality risk is \(-0.055 [.046, 0.066]\); birth from Sikh mothers not significantly different hindu mothers; the residual category for religion, ‘others’, is also not significantly different from the baseline. Being a member of a scheduled tribe increase mortality from the baseline to \(0.08 [.07, .12]\) and being a member of the schedule tribe reduce it to \(0.66\). Of course, in the actual data set these effects add up. So often, a rich person that is also a Christian, have a mother with higher education and belong to the scheduled tribe, will have a mortality risk of only \(0.011\) in 1975. The age of the mother has a significant effect as well, even though the quadratic effect term is mostly not substantively significant. Being born in an urban place does seems to be associate with lower mortality but it is also not statistically significant by conventional levels. Similarly, being a male infant seems to be associate with higher mortality than being a female but, again, not quite statistically significant. Bent lines show a statistically significant effect over time: as we move on time, the average risk of death declines. However, In all cases, the (linear) time-varying interactions are substantively very small and often not statistically significant.

### 3.4 Inequality Analysis

We now present results for the inequality analysis from model 6. We use in-sample predictions for the mortality risk for every birth \(\pi_{imcds}\) for 1000 MCMC samples thus propagating uncertainty from to the model fit to the inequality analysis. We start by contrasting mortality risk between 1976 and 1996. Figure 2 plots the mortality risk for 1976, the reference year \(F_0\) and 1996, the comparison year \(F\), multiplicative \(F_{mult}\) and exponential \(F_{exp}\) adjustments on \(F_0\), and several relative distribution comparisons. For each one of these plots, we show the mean levels and 95 % confidence interval from the mcmcm samples. The graph on the left plots kernel densities for \(F\) and \(F_0\) and their adjustments. Unsurprisingly, we confirmed that mean levels of infant mortality were declining in India between 1976 and 1996. However, in our approach we can also see details about the distribution of mortality risk that were not possible by just calculating the mean levels. The distribution of mortality risk is much more skewed in 1996 than in 1976, as in 1996 there are much more births with very low mortality risk. In left plot we can also see the two adjusted distributions, \(F_{mult}\) and \(F_{exp}\). These adjust-
Figure 1 – Results for the fixed effects from models 1-6 on the logit scale. Each dot is a point estimate. The lines are 95 posterior % confidence intervals. The coefficients that cross the vertical dotted line are not significantly different from zero. Knots are the change in slopes for the years (1978, 1981, 1985, 1986, 1990, 1992, 1995). Results for the predictors are similar across models.
Figure 2 – Distribution of mortality risk for 1976 and 1996. The plot on the left displays mortality risk for the reference (1976) and comparison (1996) years as well as two adjustments on the reference years, a multiplicative and an exponential adjustment. The plot on the center shows the overall relative distribution as well as two residual adjustments. The graph on the left shows the size of the adjustment effects. The adjustments make the reference distribution very similar to the comparison distribution. We can also look at the relative distributions to have a more precise sense about the differences between 1976 and 1996 and how much of it can be accounted for by the adjustments. The graphs on the center of the panel display the overall relative distributions for $F_0$ to $F$ and also the residual effects, $F_0$ to $F_{\text{adj}}$, for each adjustment. The adjustments overlap to a great extent, which confirms that once the adjustment are taking into account, the reference and the comparison distributions are very similar. These adjustments can also highlight the differences between 1976 and 1996 very clearly: Births which mortality risk are in the lower quintiles of the distribution in 1976 are up to 3 times or more common in 1996. Similarly, the frequency of high risk births, as defined by the distribution of mortality risk in 1976, is 2 times or more less common in 1996. Finally, the graph on the right of the panel displays two relative distributions, comparing the exponential and multiplicative adjustment effects $F_{\text{adj}}$ to $F_0$ more directly making it more clear that they are not identical. The median adjustment produces a relative distribution that is closer to a uniform distribution, meaning that there are very few differences between $F_{\text{adj}}$ and $F_0$ after the adjustment. However, the exponential adjustment produce results but these differ at the upper and lower limit of the distributions. It indicates that for this data, the multiplicative adjustment seems to produce a better sum-
Figure 3 – Trends in mortality risk for India, 1975-1998. These graphs display mean and variance. The bottom graphs show ANOVA decomposition for every year. COLOR DO NOT MATCH VERY WELL: I WILL KEEP WORKING ON IT.

...mary of the differences between these two distributions. None of these patterns cannot be seen by traditional between-group comparisons, which only describes the mean levels for each year with no sense of the dispersion around the means.

Since it is difficult to look at relative distributions for every year, we calculate summaries measures to examined trends for the entire period. Next plot we focus on mean, variance, and its decomposition. Mean and variance are both declining over time. This is consistent with what we know from looking at the previous plot. The decline in national averages of infant mortality is also consistent with it. However, by decomposing the variance we can address an important question in the literature: to what extend the variability in mortality occurs within and between large groups births. If most of the variance occur between large groups of births than knowing between-group differences is highly informative of the variability in mortality...
risk in the population. However, if most of the variance occur within-group, then focus on differences among large groups is misleading from a demography and policy perspective. On the graph on the bottom right, we decompose the variance by the main socioeconomic groups: caste, religion, states, maternal education, wealth, and districts. As in any analysis of variance, $R^2$ indicate how much of the variance is explained by the group membership. Our results are surprising. For all these grouping, almost all of the variation in inequality in infant mortality come from within-group variation, not from between group variation. In fact, for some groups, such as religious groups and caste, the grouping contains almost no information on the variability of inequality: for these cases, almost 100% variation come from within-groups. These patterns are constant over time. This suggest that socioeconomic groups are much more heterogeneous than previously thought and that births from the same group does have very different mortality risks. It also confirms the notion that comparison between large groups of births tend to mask variability in mortality risk due to high levels of heterogeneity. Because districts is the grouping level that explain most of the variance, we explore interactions between it and other levels. The objective is a preliminary exploration to see with smaller grouping cells have higher exploratory power. We thus interact district with caste, maternal education, religion, and wealth. Maternal education and wealth interacting with districts do explain more than districts alone sometime two times more.

We look at polarization indices to investigate over time changes in the tails of the distribution. The left panel shows the lower polarization indices, the middle panel the overall polarization indices and the right panel the upper polarization indices. We are comparing
every other year with the baseline year. Zero mean not change in polarization and the shaded areas are the 95% confident intervals based on the mcmc samples. All polarization indices display a similar pattern over time. Because of the large confidence intervals it seems that the polarization was fairly constant over the years. These indices is showing us that despite the decline, disparities in the risk of death remain quite large in India.

Now we also look at divergency measures across two distributions. The objective is quantify how different are the distributions and how much of it is taken into account by the multiplicative adjustment. As in the case of the polarization indices, we use 1975 and the comparison year and all subsequent years are compared against it. As we can see, the diverge between the baseline year and the others is basically gone once we adjust for the median differences via multiplicative adjustment. It means that over time changes in the distribution of mortality risk in India can be explained by a very simple multiplicative shift.

The broader from the analysis so far is this: group averages of early life survival produce a very partial picture of the mortality risk across all births. In particular, it does mask the heterogeneity that exist across large groups of births, where some births have a much higher mortality risk than others. The results presented here support the need for a more accurate picture, in order to document less favorable trends of greater disparity within certain social groups. Finally, we also show that over time changes in inequality in infant mortality can be parsimoniously described in term of a multiplicative adjustment. These findings also suggest that our methodology is useful to identify changes in inequality that are otherwise hidden by
standard approaches.

3.5 Demographic Decompositions

We now look at the demographic sources of this changes.

4 Concluding Remarks

We have presented an integrated approach to measure and explain over time changes in inequality in early life survival across all births from any given population. By looking at the full distribution of mortality risk, we add to an existing body of research that tried to document inequality in early mortality across entire populations. However, we made several contributions to these early efforts. First, we show that inequality measures from the income inequality literature do not work well on the probability scale and thus that these should not be used to describe inequality in early life mortality. This is an important point that have been previously unnoticed. We present several alternative approaches based on the relative distribution methods and on the statistical literature. We also discuss how to use covariate adjustments to explain changes across populations. Finally, we discuss statistical inferences for our inequality measures. To illustrate our approach we use a large data set from India. Using hierarchical models, we have estimated the latent death rate for each child in our data along with their associated uncertainty measures. Using in-predictions from these models we have characterized over time changes in the inequality across all births, accounting for both sampling and estimation uncertainty. We found that most of the death risk variation come from within-group variability, not from between-group. Finally, we provide evidence that these changes are not driven by over time changes in the demographic composition of the population. These patterns are statistically and substantive significant and consistent with a wide range of inequality indices. None of them has been previously documented.

The methodology developed here is not without its own limitations. The underlying death rate is a latent variable that needs to be estimated from a statistical model. The ability of our approach to describe and explain inequality is thus dependent on the estimation of the mortality risk. We are particularly concerned about including all relevant time-varying effects and interactions. Without doing so, we will be forcing the covariates to have the similar effects over time and across demographic groups which will limit our ability to apply our decomposition methods. Another limitation is that our methodology works well for well-
behaved changes in mortality risk. In that case, simple decompositions can characterize differences across distributions, as in the case of India. We believe that most developing countries will exhibit similar patterns. However, other scenarios are possible. For example, consider high mortality episodes, such as wars and natural disasters. These will likely produce very different distributions of mortality risks, for example, by creating discontinuities due to higher risk subpopulations. In these cases, other decomposition techniques may be useful, for instance working with mixture of distributions.

Our treatment of relative distribution and inequality measures can be extended in several ways. Here we show the usefulness of the decomposition the reference and the comparison distribution by a simple parameter. However, one can decompose the distribution by multiple parameters. Similarly, we introduced covariate adjustment for one variable at the time but the same logit can be applied for several covariates as well. Here the only additively decomposable inequality measured was the variance. While it is not possible to use additively decomposable measures from the income inequality literature on mortality risk, one could further explore other measures from the statistical literature. Another area with have not discussed was regarding the estimation of the mortality risk itself. More flexible models, such as non-parametric Bayesian models and machine learning models can be used to estimate the latent death risk. These models are very flexible and can deal with a vast number of interactions among covariates. It would be particularly interesting in investigating whether predictions from different models produce different results for our inequality analysis.
References


Kumar, Abhishek, and Abhishek Sing. 2014. “Is Economic Inequality in Infant Mortality Higher in Urban Than in Rural India?” Journal Maternal Child Health 18.


