SIBLING CORRELATIONS IN EARLY-CHILDHOOD MORTALITY: A COMPARISON ACROSS 77 LOW- AND MIDDLE-INCOME COUNTRIES

KIERON BARCLAY, MARTIN KOLK AND MIKKO MYRSKYLÄ

INTRODUCTION

Life begins in the family, and the childhood environment and our family relationships are enormously important for shaping the trajectories that each of us follow in life, for better or for worse. However, none of us choose which family we are born into. An understanding of the importance of the family for the health outcomes of children in both the short- and long-term is therefore a question of great significance: it is fundamentally a question about social justice. Behind a veil of ignorance, many would presumably express the preference that the family that one is born into should not determine ones life chances. However, we know that, even in a society with a reputation for socioeconomic equality such as contemporary Sweden, approximately 40-50% of the variation in educational outcomes is explained by the family that one is raised in (Bjorklund et al., 2003; Björklund and Jäntti, 2012). Some social stratification researchers have even concluded that relative social mobility is essentially the same across industrialized market economies with a predominantly nuclear family structure (Featherman et al., 1975; Erikson and Goldthorpe, 1987).

Despite the fact that it is well-known that the family is enormously important for life course trajectories, most research on health inequalities focuses on 'period' measures, such as the lifespan variation, or the standard deviation of life expectancy. These period measures are akin to using a Gini Coefficient to examine income inequality; they tell us how mortality is distributed across the population, but they do not tell us how important the family you start in will be for your health or the timing of your death. The potential importance of the family of origin for health and mortality is underappreciated, as are the implications for social inequality as a whole. The studies that do exist on this topic almost invariably examine a specific context at a specific period in time. However, the influence of the family on health outcomes is likely to have varied across time, and across societal contexts. There are few studies examining how the influence of the family on health has changed over time even within one specific society. Even fewer studies have attempted to examine changes over time across a wide range of different settings, with an accompanying attempt to identify the conditions that explain the differences across these settings . Needless to say, the potential implications of these research questions are

of great societal importance. To what extent is the influence of the family variant, or invariant, in spite of differences in the organization of public health conditions, health systems, welfare regimes, and the state of medical knowledge? Has the importance of the family for a child's health changed over time and with increasing levels of social and economic development, or are there simply trendless fluctuations?

In this study we will examine how sibling correlations in early childhood mortality vary across low- and middle-income countries, how this has changed between the 1930s and the 2010s, and how the importance of the family varies according to the level of development, using data from the Demographic and Health Surveys, and the World Bank Development Indicators. Although there is a large literature examining family clustering in early life mortality, very few studies have take a long-term perspective (Alter et al., 2001; Beise and Voland, 2002; van Dijk, 2018), and none have taken a comparative perspective (for an overview, see van Dijk, 2018). We will examine sibling correlations in infant mortality (first 12 month of life), and mortality at ages 1-5. We expect sibling ICCs in mortality to be higher at ages 1-5 than 0-1 because the importance of the family environment for child health is likely to grow progressively stronger from birth to age five. This is because mortality during the first month, and to a lesser extent the first year, of life has an important random component; if a child is born preterm, with low birth weight, or has a congenital abnormality, there is little that the parents can do about it. However, the child's health and wellbeing at ages 1-5 is subject to parental control to a much greater extent.

We also predict that sibling correlations in mortality should be low at the very lowest levels of development, because these are environments where communicable and vector-borne diseases play a far greater role in the health of populations, and families may be less able to guard against these exogenous risk factors. It is also likely that in the environments where levels of education are lowest, individuals may have a poor understanding of germ theory and be less likely to follow hygienic practices, which also increases the risk of communicable diseases. In general the poverty in these environments may also leave families more vulnerable to famines and other types of shortages, as well as to other random factors that they have little control over. We expect sibling correlations to increase as public health conditions improve and the random element of mortality decreases. There may also be a third stage; if public health conditions improve sufficiently, with ready and widespread access to high quality doctors and other medical facilities, sibling correlations in mortality may then decrease as the family becomes less important for child survival. We intend to test these hypotheses by examining the link between sibling correlations and several development indicators over time both between- and within-countries. The development indicators that we use in this study are infant mortality rates, life expectancy at birth, total fertility rates, and GNI per capita.

DATA

Demographic and Health Surveys. We used data from 228 Demographic and Health Surveys collected between 1985 and 2014 in 77 low-income countries. The DHS is a household survey, with a separate survey for women aged 15-49. The household response rates in the 228 surveys that we use range from 87.9-99.9% (mean 97.8%), while the response rates for the womans questionnaire ranges from 77.0-99.4% (mean 93.6%). Our analyses were based upon self-reported fertility histories, which we used to obtain information on maternal age at birth, birth order, sibling group size, birth intervals, birth year, and timing of child death. Since we used DHS surveys collected from 1985-2014, the fertility histories included births occurring between the 1930s and 2014.

World Bank Global Indicators. In order to understand how the sibling ICCs vary according to level of development we link various indicators of development by country-year. We sourced the data from the World Development Indicators database, which is the primary World Bank collection of development indicators, compiled from officially recognized international sources. It presents the most current and accurate global development data available, and includes national, regional and global estimates. In our analyses we particularly focus on national-level infant mortality rates, total fertility rates, GNI per capita, and e_0, estimated life expectancy at birth.

STUDY DESIGN

The approach that we use in this study is to estimate sibling intra-class correlations based upon multilevel survival models and group-level random effects. Since our data is drawn from a wide range of countries and and birth cohorts, we first compare sibling ICCs in mortality between countries. We then estimate these sibling ICCs for different five-year groupings of birth cohorts. By plotting our data and estimating OLS and country-level fixed effects regression analyses we then examine whether and how sibling ICCs in mortality have been changing over the period 1955 to 2014, and how sibling ICCs varying according to the level of development, as measured by infant mortality rates, total fertility rates, life expectancy at birth, and GNI per capita.

STATISTICAL ANALYSES

Accelerated Failure Time Models and Intra-class Correlations. We use accelerated failure time (AFT) models to estimate multilevel survival models and group-level random effects. Accelerated failure time models are fully parametric, and we fit our models using a Weibull distribution, which is appropriate for modeling early-life mortality (e.g. see Choe, 1981). In the

AFT model, the natural logarithm of the survival time, log t, is expressed as a linear function of the covariates. When we incorporate 2-level random-effects, this yields the model:

$$ln(T)_{ij} = \beta x_{ij} + u_j + \varepsilon_{ij} \tag{1}$$

$$ln(T)_{ij} = \beta x_{ij} + z_{ij}u_j + \upsilon_{ij} + \varepsilon_{ij} \tag{2}$$

for $j = 1,...,M$ clusters, with cluster j consisting of $i = 1,...,n_j$ observations, where i designates children nested within mothers j. The $1 \times p$ row vector x_{ij} contains the covariates for the fixed effects, with regression coefficients (fixed effects) β. Note, however, that we estimate 'empty models', and therefore omit fixed effects covariates. The $1 \times q$ vector z_{ij} contains the covariates corresponding to the random effects and can be used to represent both random intercepts and random coefficients. The random effects u_j are M realizations from a multivariate normal distribution with mean 0 and $q \times q$ variance matrix Σ. The random effects are not directly estimated as model parameters but are instead summarized according to the unique elements of Σ, known as variance components. Finally, υ_{ij} are the observation-level errors with density $\phi(\cdot)$. υ_{ij} are normally distributed with zero means and with variance σ_2^2. ε_{ij} is the error term; its variance, σ_1^2, is the variance component at the individual level. For AFT models in general, ε_{ij} are assumed to be independent and identically distributed, and their distribution depends on the model we are fitting. In the case of the Weibull distribution, ε_{ij} follow the Gumbell distribution, and their variance, γ, is:

$$\gamma = \pi^2/(6 \times \rho^2) \tag{3}$$

where ρ is the ancillary parameter of the Weibull distribution. We can compute the residual variance using the estimate of the log of ρ. A three-level formulation of the AFT multilevel model takes the following form:

$$ln(T)_{ijk} = \beta x_{ijk} + z_{ij}u_j + z_{ijk}u_{jk} + \upsilon_{ijk} + \varepsilon_{ijk} \tag{4}$$

where i designates children nested within mothers j, who are in turn nested within geographical areas k. For $k = 1,...,M$ clusters, each cluster k consists of $j = 1,...,n_k$ observations. For

$jk = 1, ..., M$ clusters, each cluster jk consists of $i = 1, ..., n_{jk}$ observations. The $1 \times p$ row vector x_{ijk} contains the covariates for the fixed effects, with regression coefficients (fixed effects) β. Note, however, that we also estimate 'empty models' for these three-level models, and again omit fixed effects covariates. Level-3 (geographical area) and level-2 (mother nested within area) random effects, u_i and u_{ij}, are normally distributed with zero means and with variances σ_3^2 and σ_2^2. ε_{ijk} are the error terms; their variance, γ, is the variance component at the individual level.

One of the advantages of the AFT metric, when fitting multilevel models, is that the variances of the random effects at different grouping levels can be interpreted as variance components of the log-time. This is possible because, in the AFT metric, we model the log-time as a linear combination of random effects. We can then use those variances and the variance of the residuals to compute intraclass correlations for the log-time. We estimate models using both a 2-level random intercept survival model where children are nested within a shared mother ID, as well as a 3-level random intercept survival model where children are nested within mothers, who are in turn nested within a shared geographical area. Although we can estimate the ICC for shared geographical areas, in this study we will focus on the mother-level ICC, estimated from both the two-level models, as well as the three-level models that estimate the mother-level ICC net of the shared geographical level ICC. The sibling ICCs from the 2-level ICC models are estimated as follows:

$$\rho = \frac{\sigma_2^2}{\sigma_1^2 + \sigma_2^2} \tag{5}$$

$$Y_{ij} = \beta_0 + \beta_1 X_1 + \upsilon_j + \varepsilon_{ij} \tag{6}$$

$$\rho^{(2)} = \frac{\sigma_2^2}{\gamma + \sigma_2^2} \tag{7}$$

While the sibling ICCs from the 3-level ICC models are estimated by:

$$\rho^{(2)} = \frac{\sigma_2^2 + \sigma_3^2}{\gamma + \sigma_2^2 + \sigma_3^2} \tag{8}$$

We then take these estimated sibling ICCs and examine how they have changed over time and by levels of social and economic development, detailed below.

Ordinary Least Squares and Fixed Effects Regression. To estimate how the sibling correlations in mortality vary over time and by levels of development we estimate linear regression models:

$$ICC_m = \beta_1 COHORT + \beta_2 IMR + \beta_3 e_0 + \beta_4 TFR + \beta_5 lnGNI + \varepsilon \tag{9}$$

where ICC_m is the sibling intra-class correlation in mortality during age window m, which refers to infant mortality and mortality at ages 1-5. On the right-hand side of the equation we include a covariate for $COHORT$, which corresponds to the 5-year birth cohort categories that we use in our estimation of the sibling ICCs. The other predictors, IMR (infant mortality rate), e_0 (life expectancy at birth), TFR (total fertility rate), and $lnGNI$ (natural log of GNI per capita) are the average values for each of these four variables during the 5-year period corresponding to each 5-year birth cohort. To reduce confounding from unobserved factors at the country-level, we also estimate linear fixed effects models, with the fixed effects specified at the country-level:

$$ICC_{m,ij} = \beta_1 COHORT_{ij} + \beta_2 IMR_{ij} + \beta_3 e_{0ij} + \beta_4 TFR_{ij} + \beta_5 lnGNI_{ij} + \alpha_j + \varepsilon_{ij} \tag{10}$$

where the subscript ij refers to cohort i nested within country j, and α_j refers to the country-level fixed effect. Note that cohort and period are essentially equivalent in our analyses of infant mortality, but our analyses of mortality at ages 1-5 links various development indicators during the 5-year birth cohort to mortality that may be measured in a period outside of that 5-year birth cohort. For example, an individual born in January 1994 whose mortality is followed over ages 1-5 will be followed over the period January 1995 to January 1999.

RESULTS

Descriptives. We begin by examining the distribution of the sibling ICC during different mortality age windows. As can be seen in Figure 1, most ICCs fall in the range 0.3 to 0.5 for mortality at ages 0-1 (left panel), though these ICCs are higher during the older age window (right panel), as anticipated. Some birth cohorts are characterised by extremely high sibling ICCs (see Figures S1 and S2 in the Appendices for scatter plots of sample n against mortality ICCs, as well as scatter plots for number of deaths in the analytical sample against mortality ICCs). We observe higher ICCs in settings where there are fewer infant deaths.

Sibling ICCs for Infant Mortality in Comparative Perspective. Figure S3 in the Appendices shows the ICCs for mortality at ages 0-1 and 1-5 across the 77 countries that we include in our analysis. Figure 2 shows that sibling ICCs in mortality at ages 0-1 were relatively flat between the 1950s and 1990s, though there is some evidence that ICCs were decreasing over time. However, since the 1990s there has been a clear upward trend in sibling ICCs. For ICCs in mortality at ages 1-5 there is also some suggestion that ICCs were decreasing in the early periods before increasing again. We have also plotted sibling ICCs for mortality at ages 0-1 and 1-5 against the infant mortality rate, life expectancy at birth, total fertility rate, and GNI per capita, which can be seen in Figures S4 through S7 in the Appendices. Although we will expand on this analysis in the time leading up to the workshop in January, these models suggest that sibling ICCs in mortality have increased with increasing levels of development. As our Appendices show, sibling ICCs are increasing at lower levels of infant mortality, higher levels of life expectancy at birth, lower TFRs, and higher levels of GNI per capita.

Naturally each of these development indicators are correlated with one another, and this method of visual inspection does not take account of unobserved factors within countries, nor

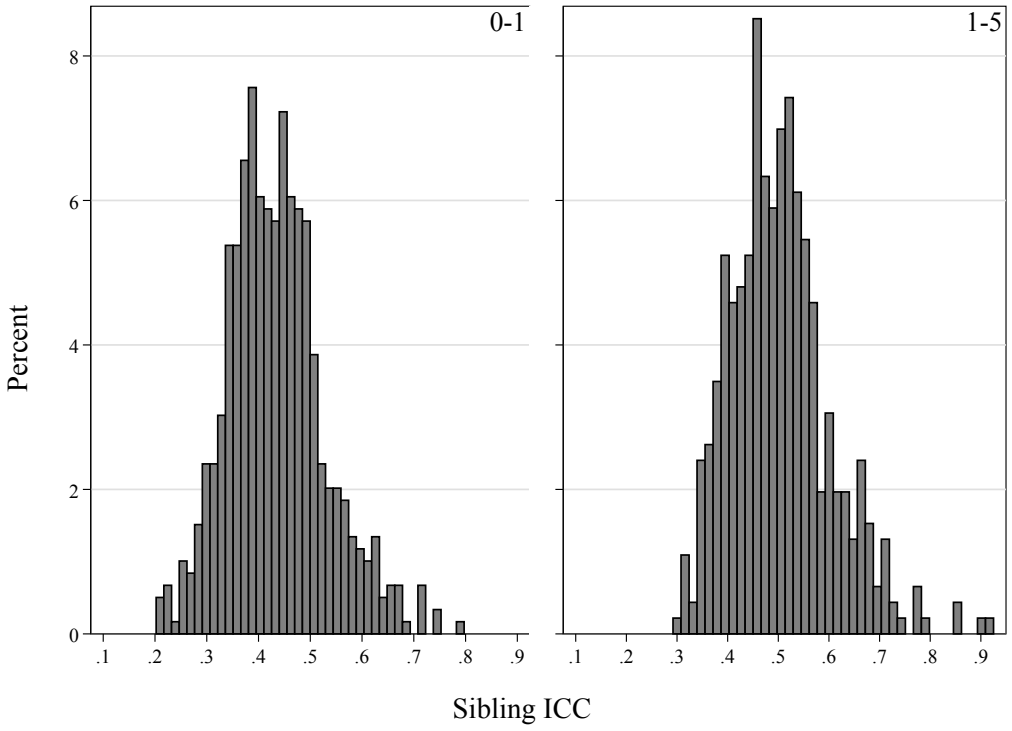


FIGURE 1. Distribution of Sibling Intra-class Correlations in 0-1 and 1-5 Mortality Across 77 Countries Included in the Demographic and Health Surveys, Based upon 5-year Cohort Analyses.

TABLE 1. Mortality at Ages 0-1 and Sibling ICCs in Mortality: Linear Regression Estimates, Including Country Fixed Effects

Development Indicator	Model 1	Model 2	Model 3	Model 4	Model 5	Model 6
IMR	-0.002(0.001)*					0.000(0.001)
IMR^2	0.000(0.000)*					0.000(0.000)
e_0		-0.021(0.008)*				0.002(0.013)
$e_0{}^2$		0.000(0.000)*				0.000(0.000)
TFR			-0.088(0.025)*			-0.035(0.033)
TFR^2			0.007(0.002)*			0.003(0.003)
*ln*GNI				-0.238(0.081)*		-0.143(0.095)
lnGNI^2				0.021(0.007)*		0.012(0.008)
Cohort					-0.041(0.010)*	-0.041(0.019)*
$Cohort^2$					0.003(0.001)*	0.002(0.001)*
N	531	534	534	464	534	463
Countries	76	76	76	72	76	72

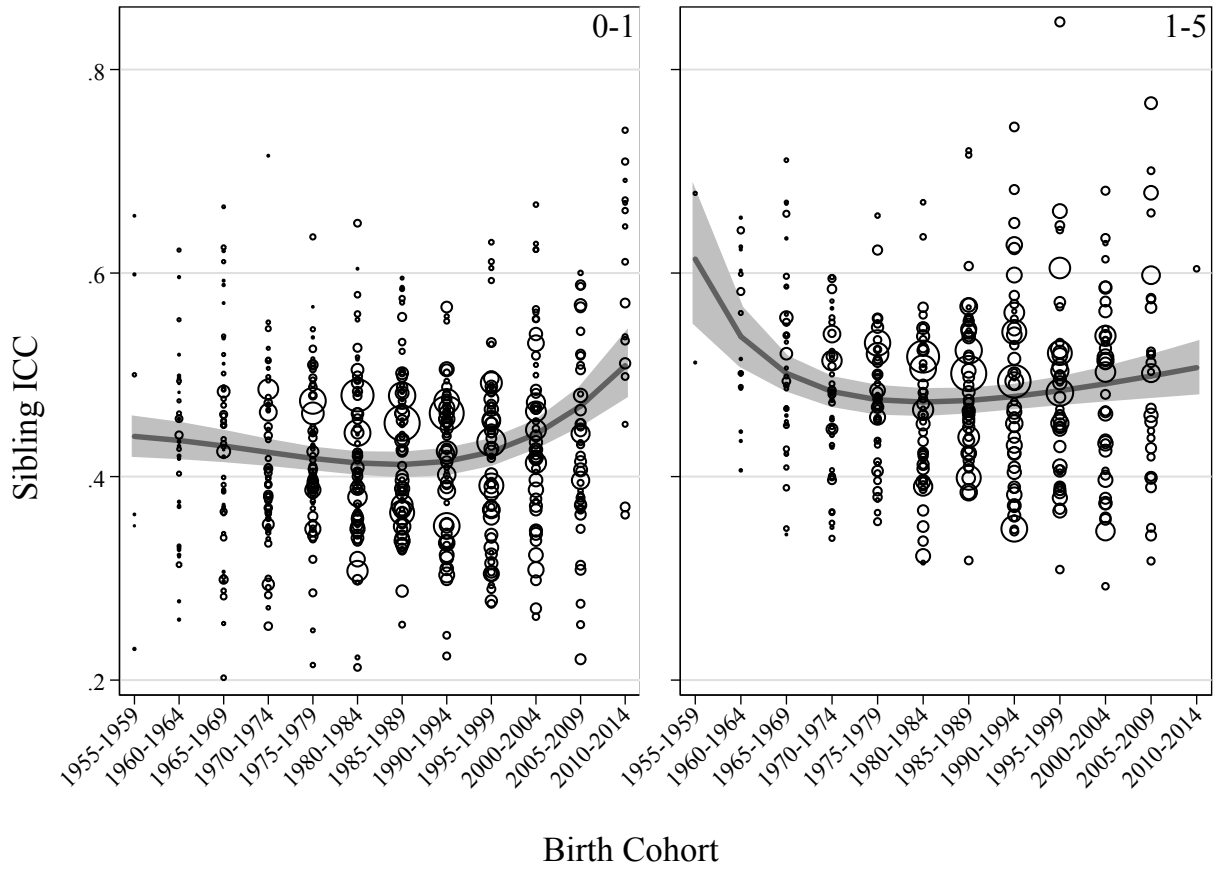


FIGURE 2. Cohort Sibling Intra-class Correlations in Mortality by Birth Cohort.

of the fact that some countries have higher levels of development and higher levels of sibling ICCs in mortality. To further examine the relationship between our indicators of development and the sibling ICCs in mortality we run linear regressions with fixed effects specified at the country-level. This allows us to examine how changes within each country over time lead are associated with changed in sibling ICCs in mortality at ages 0-1 and 1-5 (results for ages 1-5 still to be added). Given the curvilinear relationship between each of the development indicators that we employ, as well as birth cohort, and the sibling ICCs in mortality, we also include a continuous term as well as a quadratic term for each of these variables in our statistical model.

The results from these analyses are shown in Table 1 (estimates for OLS models with country-level fixed effects are shown in the Appendices in Table S1). As can be seen, there is a statistically significant relationship between each of the four development indicators that we use and sibling ICCs in mortality at ages 0-1, and these results support the idea that this relationship is curvilinear as was indicated in Figures S4 to S7. Table 1 also shows that there is a curvilinear relationship with birth cohort. In the final column of Table 1 we include all indicators simultaneously. In this case, however, we find that none of the development indicators are any longer associated with sibling ICCs in infant mortality, and the only variable that is still significantly associated is birth cohort. This suggests that calendar time is actually capturing changes in the ICC to a greater extent than the various development indicators that we are using.

Further work will follow in the run-up to the January workshop.

REFERENCES

Alter, G., Oris, M., and Broström, G. (2001). The family and mortality: a case study from rural Belgium. *Annales de Démographie Historique*, 1:11–31.

Beise, J. and Voland, E. (2002). A multilevel event history analysis of the effects of grandmothers on child mortality in a historical German population (Krummhörn, Ostfriesland, 1720-1874). *Demographic Research*, 7:469–498.

Björklund, A. and Jäntti, M. (2012). How important is family background for labor-economic outcomes? *Labour Economics*, 19(4):465–474.

Bjorklund, A., Lindahl, M., and Sund, K. (2003). Family background and school performance during a turbulent era of school reforms. *Swedish Economic Policy Review*, 10(2):111–136.

Choe, M. K. (1981). Fitting the age pattern of infant and child mortality with the Weibull survival distribution. *Asian and Pacific Census Forum*, 7(4):10–30.

Erikson, R. and Goldthorpe, J. H. (1987). Commonality and variation in social fluidity in industrial nations. part ii: the model of core social fluidity applied. *European Sociological Review*, 3(2):145–166.

Featherman, D. L., Jones, F. L., and Hauser, R. M. (1975). Assumptions of social mobility research in the US: The case of occupational status. *Social Science Research*, 4(4):329–360.

Rawls, J. (1971). *A Theory of Justice*. Harvard University Press, Cambridge, MA.

van Dijk, I. K. (2018). Early-life mortality clustering in families: a literature review. *Population Studies*, pages 1–21.

APPENDICES

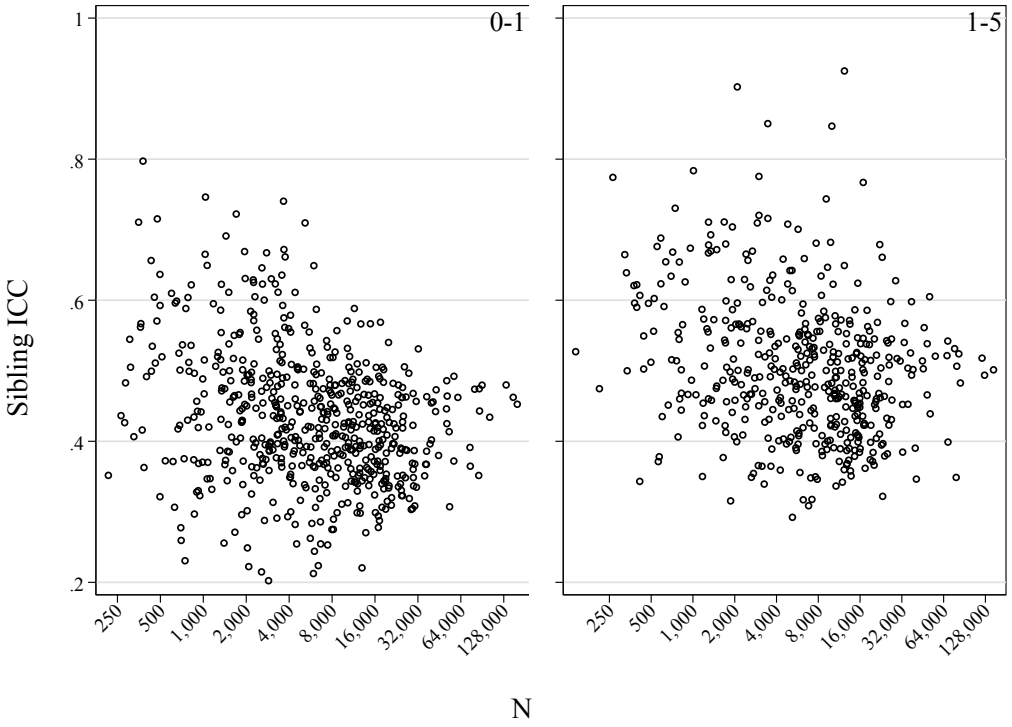


FIGURE S1. Sibling Intra-class Correlations in 0-1 and 1-5 Mortality by Analytical Sample Size Across 77 Countries Included in the Demographic and Health Surveys, Based upon 5-year Cohort Analyses.

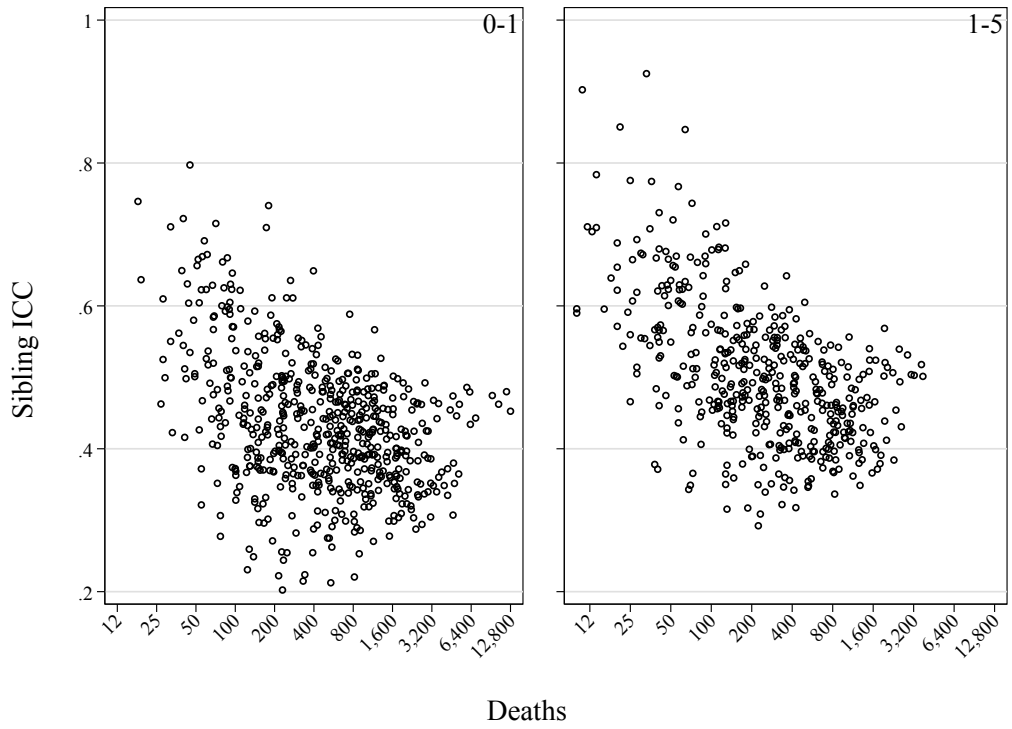
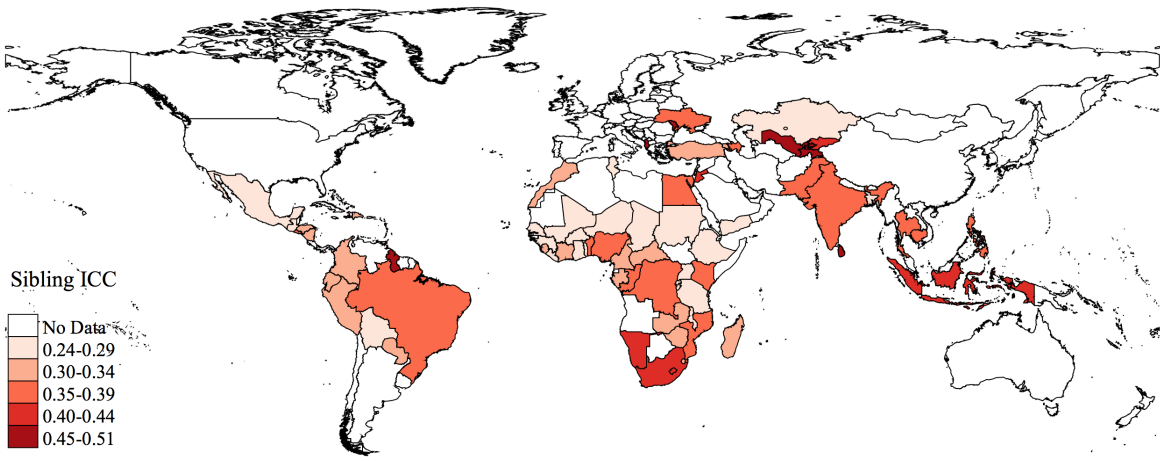
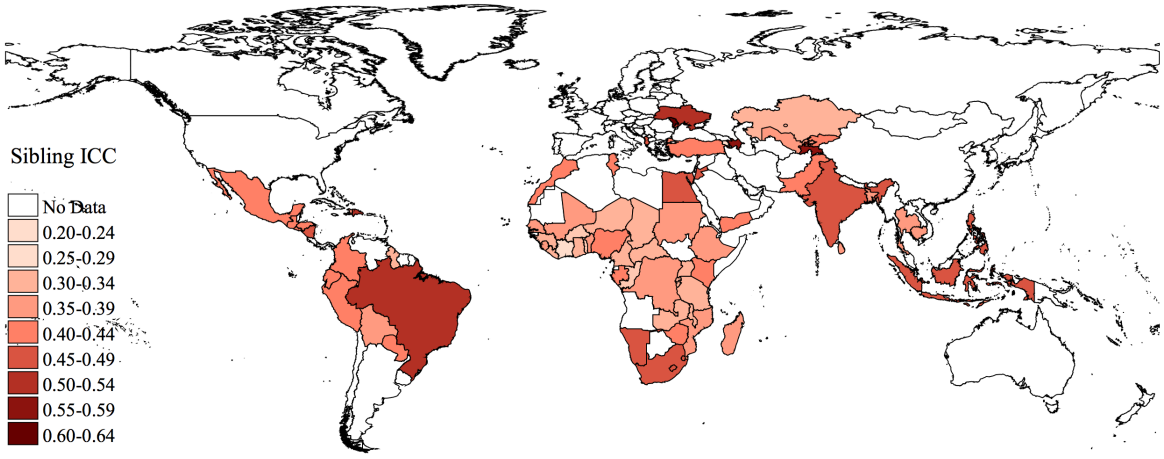


FIGURE S2. Sibling Intra-class Correlations in 0-1 and 1-5 Mortality by Number of Observed Deaths Across 77 Countries Included in the Demographic and Health Surveys, Based upon 5-year Cohort Analyses.



(a) 0-1 Mortality



(b) 1-5 Mortality

FIGURE S3. Mean Sibling Intra-class Correlations in Mortality Across 77 Countries Included in the Demographic and Health Surveys, 1935-2014.

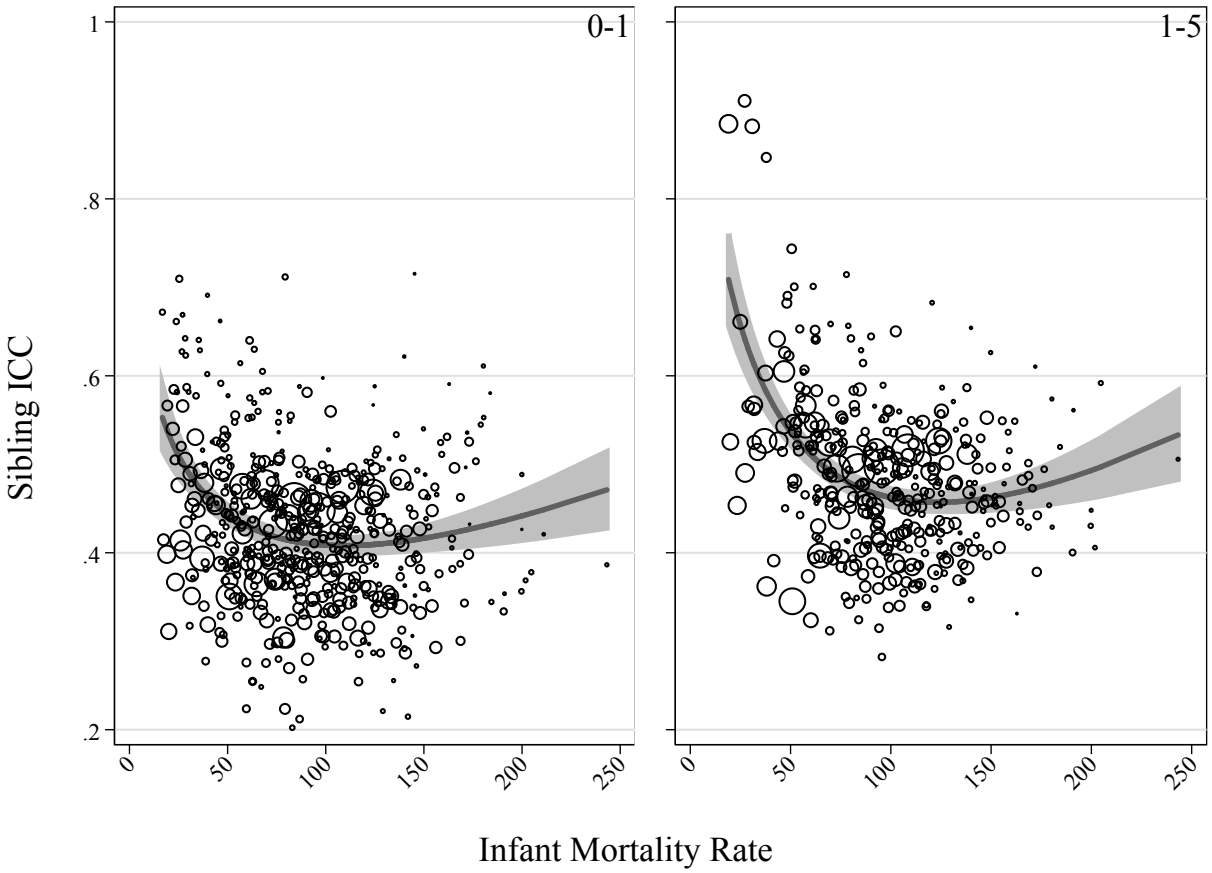


FIGURE S4. Cohort Sibling Intra-class Correlations in Mortality by Mean Country-Cohort Infant Mortality Rates.

FIGURE S5. Cohort Sibling Intra-class Correlations in Mortality by Mean Country-Cohort Life Expectancy at Birth.

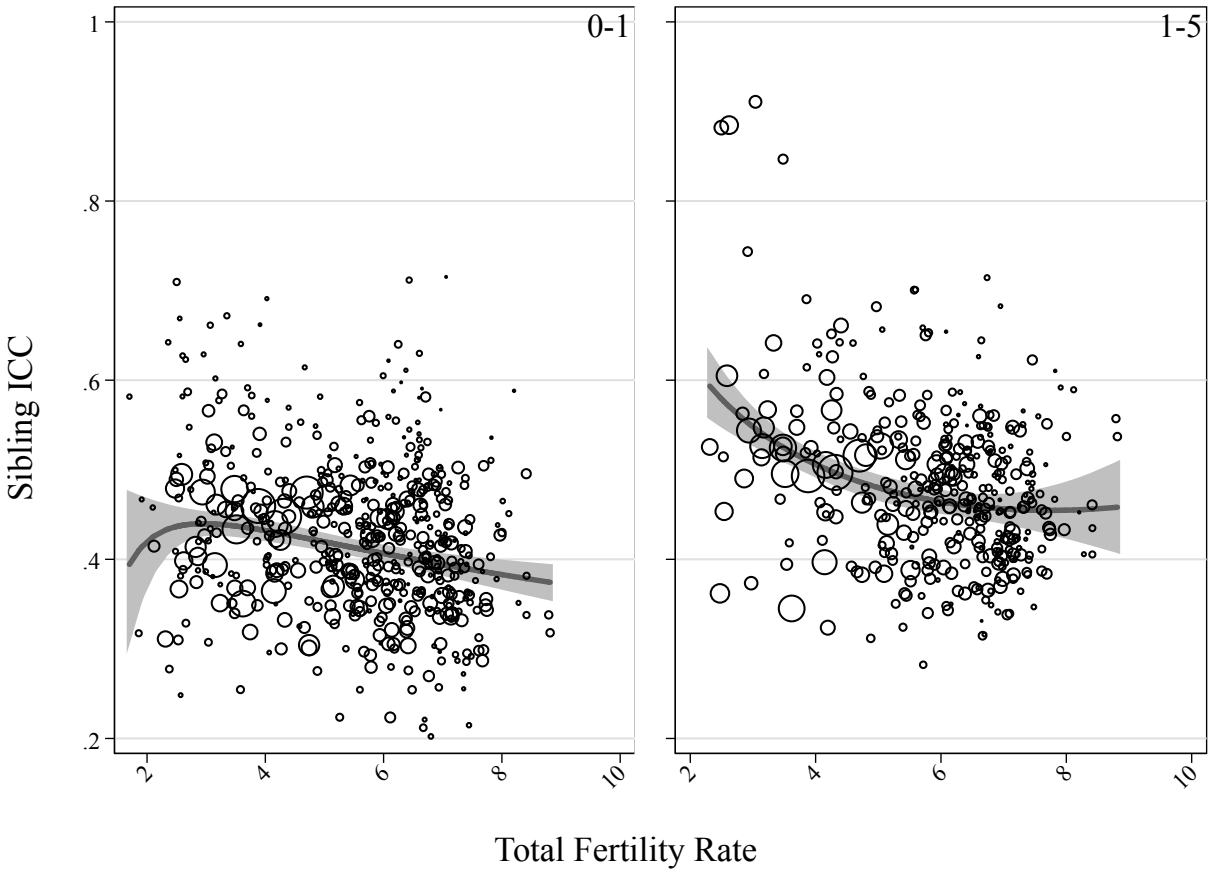


FIGURE S6. Cohort Sibling Intra-class Correlations in Mortality by Mean Country-Cohort Total Fertility Rate.

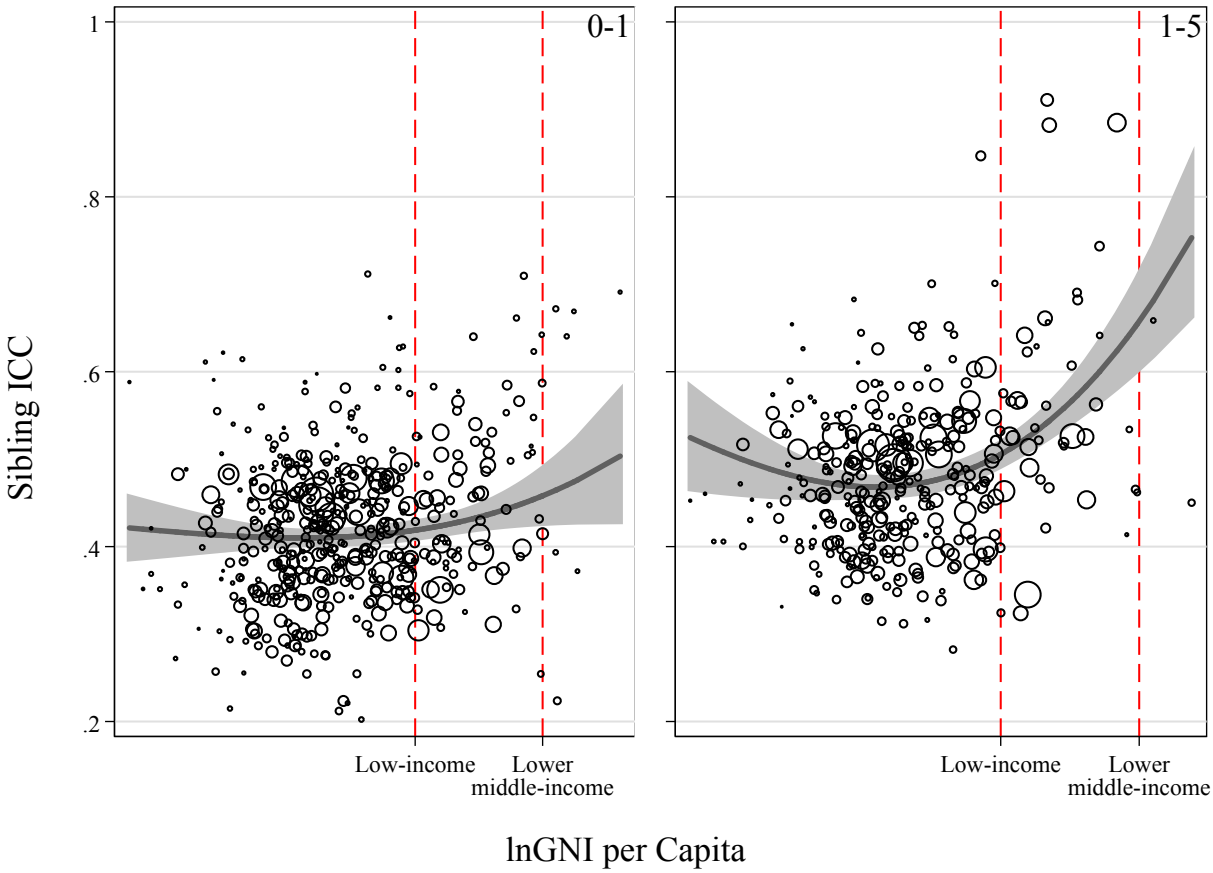


FIGURE S7. Cohort Sibling Intra-class Correlations in Mortality by Mean Country-Cohort *ln*GNI per capita.

TABLE S1. Linear Regression Estimates, No Country Fixed Effects

Development Indicator	Model 1	Model 2	Model 3	Model 4	Model 5	Model 6
IMR	-0.002(0.001)*					-0.001(0.001)
IMR^2	0.000(0.000)*					0.000(0.000)
e_0		-0.010(0.008)				0.010(0.011)
$e_0{}^2$		0.000(0.000)				0.000(0.000)
TFR			-0.008(0.024)			0.007(0.028)
TFR^2			-0.001(0.002)			-0.001(0.003)
*ln*GNI				-0.115(0.070)		-0.078(0.082)
*ln*GNI^2				0.011(0.005)		0.007(0.006)
Cohort					-0.034(0.011)*	-0.031(0.015)*
Cohort^2					0.002(0.001)*	0.002(0.001)*
N	531	534	534	464	534	463
Countries	76	76	76	72	76	72