TECHNICAL NOTE N° **IDB-TN-03057**

Sibling Correlations in Schooling around the World: A New Database

Ercio Muñoz
Gabriela Jaque

Inter-American Development Bank
Gender and Diversity Division

November 2024

Sibling Correlations in Schooling around the World: A New Database

Ercio Muñoz
Gabriela Jaque

Inter-American Development Bank
Gender and Diversity Division

November 2024

Cataloging-in-Publication data provided by the
Inter-American Development Bank
Felipe Herrera Library
Muñoz, Ercio.
Sibling correlations in schooling around the world: a new database / Ercio Muñoz, Gabriela Jaque.
p. cm. — (IDB Technical Note ; 3057)
Includes bibliographic references.
1. Educational mobility-Latin America. 2. Educational mobility-Caribbean Area. 3. Educational equalization-Latin America. 4. Educational equalization-Caribbean Area. I. Jaque, Gabriela. II. Inter-American Development Bank. Gender and Diversity Division. III. Title. IV. Series.
IDB-TN-3057

http://www.iadb.org

Sibling Correlations in Schooling around the World: A New Database*

Ercio Muñoz†
Inter-American Development Bank

Gabriela Jaque‡
Universidad de Chile

November 22, 2024

Some Preliminary Results

Abstract

We estimate sibling correlations in schooling (i.e., the fraction of inequality in educational outcomes that can be attributed to factors shared by siblings) for 94 countries, accounting for 85% of the world population. With this new database, we document several findings. On average, at least 51% of the inequality in schooling can be explained by shared family background. There are large regional differences, with North America and South Asia displaying the lowest and highest sibling correlations, respectively. There is important heterogeneity within some regions. The average sibling correlation has been decreasing across cohorts, mainly driven by North America and Latin America and the Caribbean region. Across countries, educational mobility appears to be strongly correlated with several social and economic variables.

JEL-Codes: D63, I24, J62.
Keywords: *Sibling correlations, Intergenerational mobility, Education, Family background.*

*This preliminary draft is posted for conference presentations. Any views expressed in this paper are those of the authors and should not be attributed to the Inter-American Development Bank.

†Email: erciom@iadb.org.
‡Email: gjaqueh@fen.uchile.cl.

I Introduction

How much of the variation in schooling is due to family background around the globe? There is a long interest in social science in understanding how family background matters for individual socioeconomic outcomes, including education. In particular, economists have made important progress documenting the intergenerational relationship between parent's and offspring's educational outcomes. However, the evidence using broader measures of the impact of the family remain scarce. This paper aims to fill this gap by estimating the level of sibling similarity in educational outcomes (i.e., the fraction of total inequality in educational attainment that can be attributed to factors shared by siblings, such as family background) across a large group of developed and developing countries.

We use 292 census samples spanning 94 countries to construct a new database of educational mobility across generations with global coverage, accounting for 85% of the world population. We use household rosters to identify cohabiting siblings aged 21 to 30 years and estimate sibling correlations in educational outcomes. Using this database, we document patterns across countries and over time and explore the association between this measure of mobility and a rich set of social and economic variables.

We document several empirical findings. First, we find that, on average, at least 51% of the inequality in schooling can be explained by shared family background. Second, we report mobility by region, showing that, on average, countries from North America have the lowest level of sibling correlations, while the opposite is true for countries in South Asia. Third, there is important heterogeneity within some regions of the developing world. For instance, several regions have countries among the ten most mobile and the ten least mobile globally. Fourth, we show that mobility in education (as measured by sibling correlations) has been increasing globally but that this trend is driven by North America and Latin America and the Caribbean. Fifth, we show that educational mobility is strongly correlated to several variables related to the economy, education, health, labor market, demography, infraestructure, and governance.

This paper contributes to the literature in several ways. First, it complements recent studies by estimating a new indicator of educational mobility across generations that goes beyond the parent-offspring association or the role of a given set of circumstances. Second, it significantly improves the coverage of previous studies on sibling correlations by accounting for 85% of the world population and approximately 91% in the case of developing countries. Third, it uses census data, which contain large samples and provide high cross-country comparability to study educational mobility around the world. Fourth, the paper provides novel evidence on changes in mobility between cohorts and regions. Lastly, it explores how sibling correlations are associated with a rich set of social and economic variables at the country level.

The rest of the paper is organized as follows. Section II details the data and methods used in the paper. Section III describes the main patterns derived from the estimates of sibling correlations, and section IV explores the statistical association between educational mobility and a set of variables previously used in the literature. Finally, section V offers some concluding remarks.

II Data and Methods

Data. This paper uses population and housing census data obtained from IPUMS International (Ruggles et al., 2024), hosted at the University of Minnesota Population Center, which reports harmonized representative samples (usually 10%) of full count census micro data sets for a large number of countries.[1] These censuses are conducted with the goal of obtaining updated social and demographic information about the entire population and their housing conditions.

Educational attainment. There are two questions about educational attainment in the harmonized data set. The first one reports the total years of schooling completed by

[1]Census data has been used to study educational mobility across generations in, for example, Alesina, Hohmann, Michalopoulos, and Papaioannou (2021, 2023); Card, Domnisoru, and Taylor (2022); Derenoncourt (2022); Munoz (2024).

each individual (formal schooling regardless of the track or kind of study), and the second one is re-coded by IPUMS to capture educational attainment in terms of the level of schooling completed[2] and contains four categories: (1) less than primary completed, (2) primary completed, (3) secondary completed, and (4) university completed. In the main analysis of the paper, we use the latter variable, which has better country coverage, with years of schooling imputed based on the average years of schooling computed for each level of education in those samples where both variables are available. As we will show later, we obtain estimates of sibling correlations from these imputed variables that are closely aligned with those obtained with years of schooling as originally reported.

Country coverage. The data set includes 94 countries with information about levels of schooling and 64 countries with information on years of schooling, which respectively account for 85% and 64% of the World's population (see Table 1). For most of the regions of the World, these countries represent more than 58% of their population. Latin America and the Caribbean and North America are the regions with the largest coverage in terms of population, both considering information about years of schooling and levels of education. By income level, the coverage is 91% of the population for developing countries (57% with years of schooling) and 61% for high-income countries (26% with years of schooling). To the best of our knowledge, this database provides the largest coverage regarding the number of countries, as well as the world population accounted for in the literature of sibling correlations.[3]

Identification of siblings. The data collection is organized at the household level, so it is possible to link those individuals who live in the same household at the time of the interview through a variable that, using 62 different values, details the relationship between each individual and the head of the household. Based on this variable, we identify

[2]Note that it does not necessarily reflect any particular country's definition of the various levels of schooling in terms of terminology or number of years of schooling. This variable applies, to the extent possible, the United Nations standard of six years of primary schooling, three years of lower secondary schooling, and three years of higher secondary schooling.

[3]Concurrent work in Ahsan, Emran, Jiang, Han, and Shilpi (2023) offers estimates for 53 developing countries using Demographic and Health Surveys.

Table 1: Country coverage

	Levels of education		Years of schooling	
Region	Countries	Population	Countries	Population
East Asia & Pacific	12	87.87	10	27.22
Europe & Central Asia	17	58.21	1	1.12
Latin America & Caribbean	25	99.54	24	99.45
Middle East & North Africa	7	61.48	2	9.11
North America	2	99.98	1	90.08
South Asia	4	96.99	3	85.28
Sub-Saharan Africa	27	78.65	23	69.26
World	94	85.17	64	51.20

Notes: The table shows the number of countries covered by the database (with information about levels of education as well as years of schooling) and the population share that they account for. Table A2 shows the country coverage by income level.

individuals that are likely to be siblings. For example, individuals classified as children of the head of the household are listed as siblings (the Appendix provides the identification details). We restrict the sample of siblings to individuals aged between 21 and 30 years, with the idea of maximizing the likelihood of them completing their education while having enough probability of cohabitation. A potential concern that may emerge from this sample restriction is the possibility of getting biased estimates due to coresidence bias, however, Ahsan, Emran, Jiang, and Shilpi (2023) provide evidence that this type of bias is small for measures such as sibling correlations.[4]

Measurement. We estimate a sibling correlation in educational attainment following previous literature (see for example, Grätz et al., 2021), which is derived from:

$$y_{ij} = \beta' X_{ij} + \epsilon_{ij} \tag{1}$$

where y_{ij} denotes the measure of educational attainment (e.g., years of schooling) for the jth sibling in family i, X_{ij} is a vector of exogenous variables that account for life-cycle stage

[4]Munoz and Siravegna (2023) show that the bias due to cohabitation restriction is also small in the context of upward mobility and that the rankings derived from other indicators of mobility such as the intergenerational regression coefficient using only coresident samples can be reliable.

(e.g., a cubic in age) with β as the associated vector of coefficients,[5] and ϵ_{ij} is a disturbance term that represents educational attainment net of life-cycle effects.

The disturbance term has two components:

$$\epsilon_{ij} = \alpha_i + \mu_{ij} \tag{2}$$

where α_i is a permanent component common to all siblings of family i and μ_{ij} is an idiosyncratic individual component. We assume that these two components are orthogonal, which implies that the variance of the disturbance term can be expressed as:

$$\sigma^2_\epsilon = \sigma^2_\alpha + \sigma^2_\mu \tag{3}$$

Hence, the correlation of educational attainment among siblings is

$$\rho = \sigma^2_\alpha / (\sigma^2_\alpha + \sigma^2_\mu) \tag{4}$$

which is the proportion of the population variance in educational attainment that is due to factors shared by siblings.[6]

III Sibling Correlations around the World

In this section, we document observed patterns based on the estimated sibling correlations. First, we analyze patterns across geographical regions of the World. Second, we explore the variation in educational mobility across countries. Finally, we inspect patterns across cohorts by estimating our mobility indicators by different birth cohorts based on the year in which the census data was collected.

Evidence across regions. As described in our data section, the country coverage by

[5] Our baseline estimates do not include age. Our estimates barely change when it is included.

[6] We estimate this correlation in Stata using the command "mixed" for estimation of linear mixed-effects models using restricted maximum likelihood.

region varies from nearly 58% to almost 100% (see Table 1). In particular, the coverage in terms of population accounted for in North America, Latin America and the Caribbean, and South Asia is much better than in other regions of the World. The lowest coverage is in Europe and Central Asia and the Middle East and North Africa (58.21% and 61.48% of their population, respectively). Nonetheless, we aggregate the available information to compare the evidence of the importance of family background across these regions.

Table 2 summarizes our main results. Given the heterogeneity in the number of samples available for each country and the year in which data collection happened, we aggregate the information in three different ways, keeping all cohorts, just the first cohort of each country, and just the last cohort of each country. We compute simple averages of country cohorts (i.e., census samples) within each region, which reflects the level of mobility in the average country from a given region. We find that South Asia shows the highest level of sibling correlations (lowest educational mobility). Conversely, the lowest level of sibling correlations (highest educational mobility) can be found in North America. Europe and Central Asia and Middle East and North Africa also show higher mobility than other regions, while Sub-Saharan Africa and East Asia and Pacific show lower levels. Interestingly, by the level of mobility, Latin America and the Caribbean appear in the middle of the seven regions when taking all cohorts or the last cohort but with very low mobility when taking only the first cohort, which suggests that levels of mobility have been changing across cohorts in this region. Disaggregating by income level (see Table A3), sibling correlations are, on average, lower in high-income countries than in developing countries (0.43 vs. 0.54 when the last cohort is considered).

Evidence across countries. Figure 1 presents an overview of the level of sibling correlations in schooling across the world using the most recent cohort for each country. The figure ranks countries in ascending order by level of sibling correlations, dividing them into two groups (bottom half and upper half). In line with the results from the previous section (see Table 2), countries from North America are in the bottom half (i.e., highest mobility),

Table 2: Sibling correlation by region

Region	Average	First cohort	Last cohort	Countries
North America	0.44	0.45	0.37	2
Europe & Central Asia	0.45	0.45	0.43	17
Middle East & North Africa	0.47	0.47	0.49	7
Latin America & Caribbean	0.57	0.60	0.52	25
Sub-Saharan Africa	0.54	0.55	0.55	27
East Asia & Pacific	0.53	0.54	0.56	12
South Asia	0.62	0.62	0.61	4
World	0.53	0.54	0.51	94

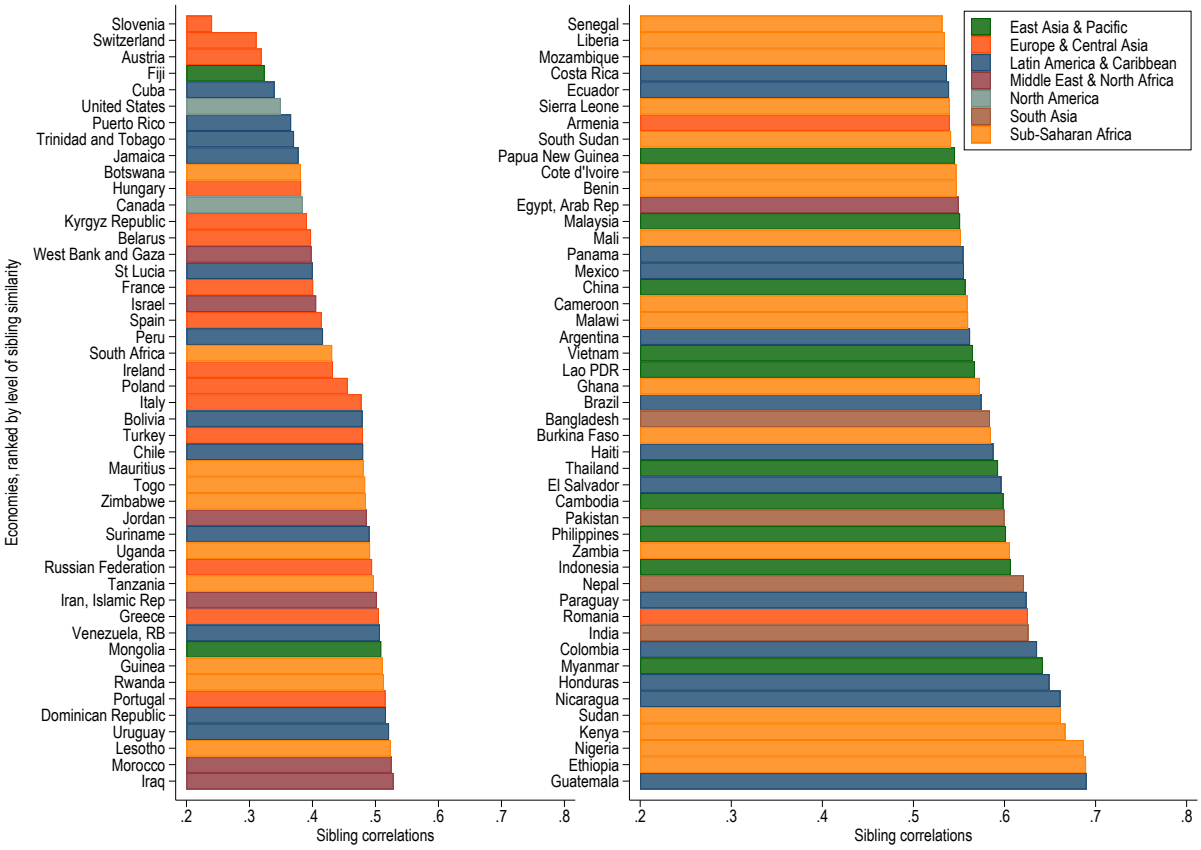
Notes: The table reports simple averages of our estimates from each region. Columns "Average", "First cohort", and "Last cohort", respectively, use all the birth cohorts available for each country, only the first cohort (or census sample) available from each country, and the latest birth cohort (or census sample) available from each country. Column "Countries" reports the number of countries with data for each region. Table A3 in the Appendix provides these statistics by income level.

while all the countries from South Asia are in the upper half (i.e., lowest mobility). However, for other regions of the World, we can find countries in both groups, and they are very diverse in levels of mobility. For example, Latin America and the Caribbean have Cuba, Puerto Rico, Trinidad and Tobago, and Jamaica among the ten countries with the highest mobility and at the same time Guatemala, Nicaragua, Honduras, and Colombia among the ten countries with the lowest mobility. In a similar fashion, East Asia and Pacific have Fiji with high mobility and Myanmar with a relatively low level of mobility, while Sub-Saharan Africa has Bostwana with high mobility and Ethiopia, Nigeria, Kenya, and Sudan with low mobility.

Sibling correlations over time. Given the time coverage of our data set, we document how the level of mobility has changed across birth cohorts at the global level. Figure 2 shows that the average level of mobility has been increasing across cohorts (i.e., the average level of sibling correlations has decreased across cohorts).[7] This pattern can be seen in high-income economies as well as in developing economies (see Figure A6 of the Appendix). However, this

[7]Figure A4 shows a similar pattern when using the subset of economies with information on years of schooling and Figure A5 shows that this pattern holds when only using economies with at least 4 censuses.

Figure 1: Sibling correlations around the World (from most mobile to least mobile)

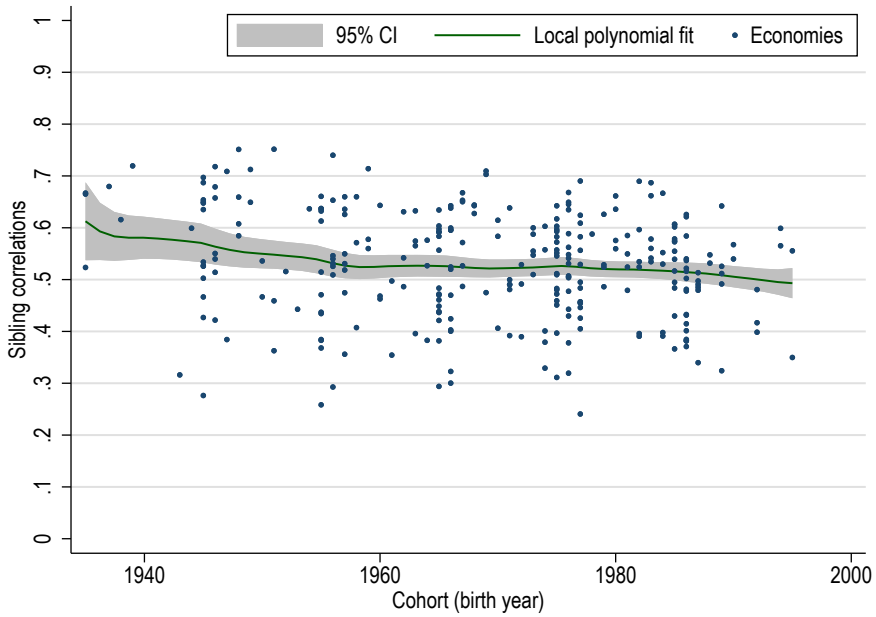


pattern does not hold for all regions of the World, as shown in Figure A7 of the Appendix. Regions such as Latin America and the Caribbean, and North America have a clear downward trend in their average level of sibling correlations, while other regions, such as Sub-Saharan Africa, do not show clear trends across cohorts.

IV Correlates of Sibling Similarity

In this section, we use our newly created database to explore the association between our measure of sibling similarity in schooling and a set of correlates explored in the literature on intergenerational mobility. In particular, we replicate the selection of correlates from

Figure 2: Sibling correlation in schooling over time



Van der Weide, Lakner, Mahler, Narayan, and Gupta (2024).[8] Figure 3 reports the coefficients obtained from regressing our estimates of sibling similarity on a wide set of correlates related to the economy, education, health, labor market, demography, infrastructure, and governance.

We find that the size of the economy, the size of the government (in terms of revenue and also expenditure), and the level of income equality are negatively associated with the level of sibling similarity.

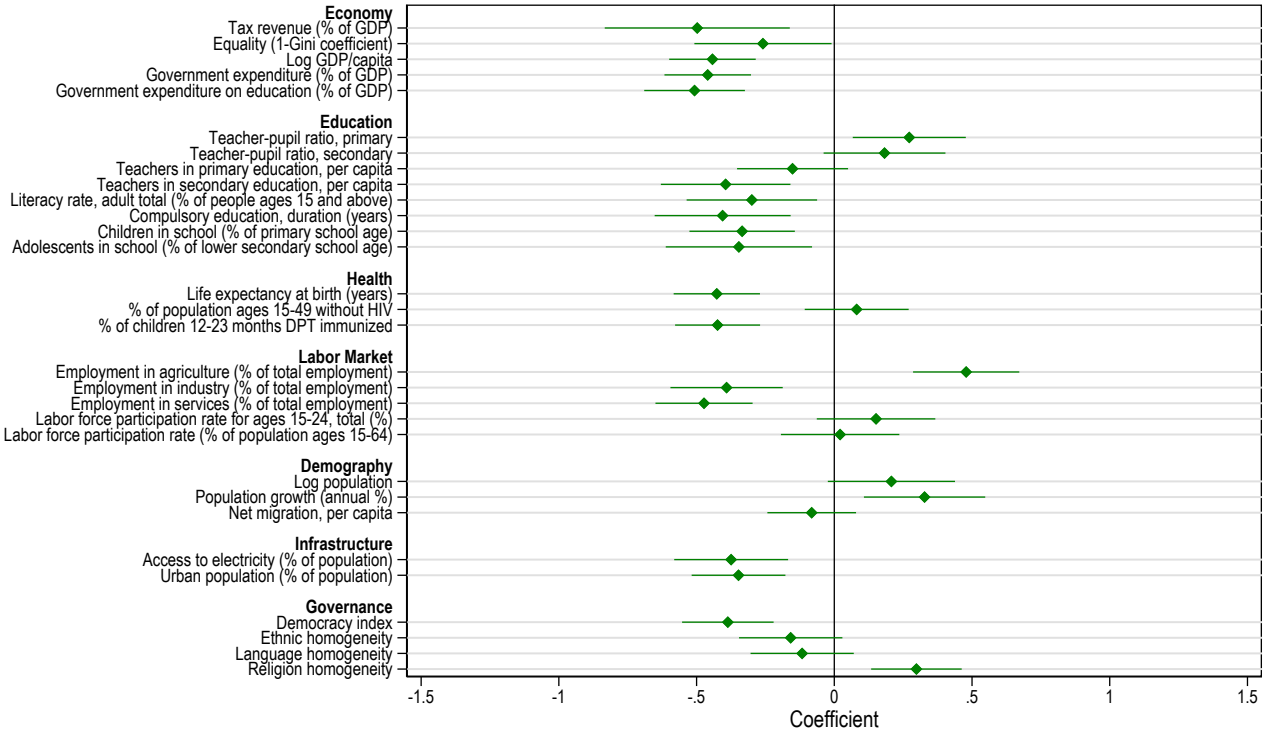
Most of the variables related to education and health are significantly correlated with the level of sibling similarity. A higher teacher-pupil ratio is associated with higher sibling correlation, while the opposite is true for other indicators associated with enrollment, literacy, or the availability of teachers. Interestingly, we find a statistically significant association between our measure of mobility and years of compulsory schooling, which is in contrast to Van der Weide et al. (2024).

In the remaining variables, we find that the employment composition among agriculture,

[8]This selection is guided by data availability and previous literature (see Alesina et al., 2021; Asher, Novosad, & Rafkin, 2021; Chetty, Hendren, Kline, & Saez, 2014).

industry, and services is related to sibling similarity. Population growth is positively correlated with sibling similarity, while access to electricity or the share of the urban population is negatively correlated. Finally, we find that the democracy index is negatively associated with sibling similarity, but religious homogeneity is positively associated.

Figure 3: Correlates of sibling similarity



Notes: The figure plots the coefficients from a regression of our measure of sibling correlations (standardized to have mean 0 and variance 1) on each covariate (standardized to have mean 0 and variance 1), including 95% confidence intervals with standard errors clustered by country. The uncertainty of the estimate of sibling correlation is not taken into account by these confidence intervals. Each sample is matched with a correlate 10 years earlier such that a given cohort aged 21-30 years is associated with the value of the correlates when they were around primary or secondary school.

V Concluding Remarks

In this paper, we create and examine a new database of sibling correlations in educational outcomes that spans 94 developed and developing countries.

We find that at least 51% of the inequality in schooling can be explained by shared family background. Countries from North America have the lowest level of sibling correlations, while the opposite is true for countries in South Asia. However, there is important heterogeneity within some regions of the developing world. We also show that the average sibling correlations in schooling have been decreasing globally, driven by North America and Latin America and the Caribbean. Finally, we show that educational mobility is strongly correlated to several variables related to the economy, education, health, labor market, demography, infraestructure, and governance.

Data acknowledgment

The authors wish to acknowledge the statistical offices that provided the underlying data making this research possible: National Institute of Statistics and Censuses, Argentina; National Statistical Service, Armenia; National Bureau of Statistics, Austria; Bangladesh Bureau of Statistics, Bangladesh; Ministry of Statistics and Analysis, Belarus; National Institute for Statistics and Economic Analysis, Benin; National Institute of Statistics, Bolivia; Central Statistics Office, Botswana; Institute of Geography and Statistics, Brazil; National Institute of Statistics and Demography, Burkina Faso; National Institute of Statistics, Cambodia; Central Bureau of Census and Population Studies, Cameroon; Statistics Canada, Canada; National Institute of Statistics, Chile; National Bureau of Statistics, China; National Administrative Department of Statistics, Colombia; National Institute of Statistics and Censuses, Costa Rica; Office of National Statistics, Cuba; National Institute of Statistics, Côte d'Ivoire; National Statistics Office, Dominican Republic; National Institute of Statistics and Censuses, Ecuador; Central Agency for Public Mobilization and Statistics, Egypt; Department of Statistics and Censuses, El Salvador; Central Statistical Agency, Ethiopia; Bureau of Statistics, Fiji; National Institute of Statistics and Economic Studies, France; Federal Statistical Office, Germany; Ghana Statistical Services, Ghana; National Statistical

Office, Greece; National Institute of Statistics, Guatemala; National Statistics Directorate, Guinea; Institute of Statistics and Informatics, Haiti; National Institute of Statistics, Honduras; Central Statistical Office, Hungary; BPS Statistics Indonesia, Indonesia; Statistical Centre, Iran; Central Organization for Statistics and Information Technology, Iraq; Central Statistics Office, Ireland; Central Bureau of Statistics, Israel; National Institute of Statistics, Italy; Statistical Institute, Jamaica; Department of Statistics, Jordan; National Bureau of Statistics, Kenya; National Statistical Committee, Kyrgyzstan; Statistics Bureau, Laos; Bureau of Statistics, Lesotho; Institute of Statistics and Geo-Information Systems, Liberia; National Statistical Office, Malawi; Department of Statistics, Malaysia; National Directorate of Statistics and Informatics, Mali; Statistics Mauritius, Mauritius; National Institute of Statistics, Geography, and Informatics, Mexico; National Statistical Office, Mongolia; Department of Statistics, Morocco; National Institute of Statistics, Mozambique; Ministry of Labour, Immigration and Population, Myanmar; Central Bureau of Statistics, Nepal; National Institute of Information Development, Nicaragua; Statistics Division, Pakistan; Central Bureau of Statistics, Palestine; Census and Statistics Directorate, Panama; National Statistical Office, Papua New Guinea; General Directorate of Statistics, Surveys, and Censuses, Paraguay; National Institute of Statistics and Informatics, Peru; National Statistics Office, Philippines; Central Statistics Office, Poland; National Institute of Statistics, Portugal; U.S. Bureau of the Census, Puerto Rico; National Institute of Statistics, Romania; Federal State Statistics Service, Russia; National Institute of Statistics, Rwanda; Government Statistics Department, Saint Lucia; National Agency of Statistics and Demography, Senegal; Statistics Sierra Leone, Sierra Leone; Statistical Office of the Republic of Slovenia, Slovenia; Statistics South Africa, South Africa; National Bureau of Statistics, South Sudan; National Institute of Statistics, Spain; Central Bureau of Statistics, Sudan; General Bureau of Statistics, Suriname; Federal Statistical Office, Switzerland; Bureau of Statistics, Tanzania; National Statistical Office, Thailand; National Institute of Statistics, Togo; Central Statistical Office, Trinidad and Tobago; Turkish Statistical Institute, Turkey; Bureau of Statistics, Uganda; State Committee

of Statistics, Ukraine; Office of National Statistics, United Kingdom; Bureau of the Census, United States; National Institute of Statistics, Uruguay; National Institute of Statistics, Venezuela; General Statistics Office, Vietnam; Central Statistics Office, Zambia; and National Statistics Agency, Zimbabwe.

References

Ahsan, M. N., Emran, M. S., Jiang, H., Han, Q., & Shilpi, F. (2023). Growing up together: Sibling correlation, parental influence, and intergenerational educational mobility in developing countries. *Policy Research working paper 10285*.

Ahsan, M. N., Emran, M. S., Jiang, H., & Shilpi, F. (2023). Making the most of coresident data: Credible evidence on intergenerational mobility with sibling correlation. *Available at SSRN 4666149*.

Alesina, A., Hohmann, S., Michalopoulos, S., & Papaioannou, E. (2021). Intergenerational Mobility in Africa. *Econometrica*, *89*(1), 1–35.

Alesina, A., Hohmann, S., Michalopoulos, S., & Papaioannou, E. (2023). Religion and educational mobility in africa. *Nature*, *618*(7963), 134–143.

Asher, S., Novosad, P., & Rafkin, C. (2021). Intergenerational Mobility in India: Estimates from New Methods and Administrative Data. *Working Paper. Available at: http://paulnovosad.com/pdf/anr-india-mobility.pdf (accessed November 2021))*.

Björklund, A., & Jäntti, M. (2012). How Important is Family Background for Labor-economic Outcomes? *Labour Economics*, *19*(4), 465–474.

Björklund, A., & Jäntti, M. (2020). Intergenerational Mobility, Intergenerational Effects, Sibling Correlations, and Equality of Opportunity: A Comparison of Four Approaches. *Research in Social Stratification and Mobility*, *70*.

Björklund, A., Jäntti, M., & Lindquist, M. J. (2009). Family background and income during the rise of the welfare state: Brother correlations in income for Swedish men born 1932-1968. *Journal of Public Economics*, *93*(5-6), 671–680.

Björklund, A., Lindahl, L., & Lindquist, M. J. (2010). What more than parental income, education and occupation? An exploration of what Swedish siblings get from their parents. *B.E. Journal of Economic Analysis and Policy*, *10*(1).

Björklund, A., & Salvanes, K. G. (2011). Education and Family Background: Mechanisms and Policies. In E. A. Hanushek, S. J. Machin, & L. Woessmann (Eds.), *Handbook of the economics of education* (Vol. 3,

pp. 201–247). Amsterdam: North Holland.

Bredtmann, J., & Smith, N. (2018). Inequalities in Educational Outcomes: How Important Is the Family? *Oxford Bulletin of Economics and Statistics*, *80*(6), 1117–1144.

Card, D., Domnisoru, C., & Taylor, L. (2022). The Intergenerational Transmission of Human Capital: Evidence from the Golden Age of Upward Mobility. *Journal of Labor Economics*, *40*(S1), S39–S95.

Chetty, R., Hendren, N., Kline, P., & Saez, E. (2014). Where is the Land of Opportunity? The Geography of Intergenerational Mobility in the United States. *The Quarterly Journal of Economics*, *129*(4), 1553–1623.

Conley, D., & Glauber, R. (2008). All in the family? Family Composition, Resources, and Sibling Similarity in Socioeconomic Status. *Research in Social Stratification and Mobility*, *26*(4), 297–306.

Derenoncourt, E. (2022). Can You Move to Opportunity? Evidence from the Great Migration. *American Economic Review*, *112*(2), 369–408.

Grätz, M., Barclay, K. J., Wiborg, Ø. N., Lyngstad, T. H., Karhula, A., Erola, J., ... Conley, D. (2021). Sibling similarity in education across and within societies. *Demography*, *58*(3), 1011–1037.

Hällsten, M., & Thaning, M. (2018). Wealth vs . education , occupation and income: Unique and overlapping influences of SES in intergenerational transmissions. *The Department of Sociology Working Paper Series, Stockholm University*(35).

Marks, G. N., & Mooi-Reci, I. (2016). The declining influence of family background on educational attainment in Australia: The role of measured and unmeasured influences. *Social Science Research*, *55*, 171–185. Retrieved from http://dx.doi.org/10.1016/j.ssresearch.2015.10.002

Mazumder, B. (2008). Sibling Similarities, Differences and Economic Inequality. *Journal of Population Economics*, *21*(3), 685–701.

Munoz, E. (2024). The geography of intergenerational mobility in latin america and the caribbean. *IDB Working Paper Series*(1620).

Munoz, E., & Siravegna, M. (2023). When Measure Matters: Coresidence Bias and Intergenerational Mobility Revisited. *IDB Working Paper 01469*.

Raaum, O., Salvanes, K. G., & Sørensen, E. (2006). The neighbourhood is not what it used to be. *Economic Journal*, *116*(508), 200–222.

Ruggles, S., Cleveland, L., Lovaton, R., Sarkar, S., Sobek, M., Burk, D., ... Lee, J. (2024). *Integrated Public Use Microdata Series, International: Version 7.5 [dataset].* Minneapolis, MN: IPUMS: Minnesota Population Center.

Schnitzlein, D. (2014). How Important Is the Family? Evidence from Sibling Correlations in Permanent Earnings in the USA, Germany, and Denmark. *Journal of Population Economics*, *27*(1), 69–89.

Sieben, I., Huinink, J., & De Graaf, P. M. (2001). Family background and sibling resemblance in educational

attainment trends in the former FRG, the former GDR, and the Netherlands. *European Sociological Review*, *17*(4), 401–430.

Sobek, M., & Kennedy, S. (2009). The development of family interrelationship variables for international census data. *Minneapolis,"University of Minnesota Population Center Working Paper Series*(2009-02).

Van der Weide, R., Lakner, C., Mahler, D. G., Narayan, A., & Gupta, R. (2024). Intergenerational mobility around the world: A new database. *Journal of Development Economics*, *166*, 103167.

Appendices

The appendix provides additional tables and figures, and other relevant information.

A Identification of siblings

The microdata of each census in the study are organized at the household level, so it is possible to link those individuals who live in the same household at the time of the interview through a variable that, using 62 different values, details the relationship between each member of the household and the head of the household. This information allows us to identify groups of siblings in two ways. First, we use the variables *pernum_pop* and *pernum_mom*, which are variables constructed by IPUMS (Sobek & Kennedy, 2009) providing the identity within the household of the probable father and probable mother of each household member (for example, for individuals classified as children of the head of the household, these variables will identify the head as either probable father or probable mother, depending on their sex). All household members with the same probable father are considered as siblings. Similarly, all household members with the same probable mother are considered siblings when the information about the probable father is missing. Second, among those with missing parental information (i.e., missing values for *pernum_pop* and *pernum_mom*), we identify groups of siblings using the information reported in the variable *related*, which describes the individual's relationship to the household head. Heads of the households and those classified as siblings or step-siblings of the head of the household form a group of siblings. Similarly, spouses or partners and those classified as siblings or step-siblings of the spouse/partner of the head of the household form a group of siblings.

Table A1: Overview of previous estimates available in the literature

Country	Study	Sibling type	Estimate
Australia	Marks and Mooi-Reci (2016)[2]	Mixed	0.34-0.58
Denmark	Bredtmann and Smith (2018)[1]	Mixed	0.33
Denmark	Bredtmann and Smith (2018)[1]	Brothers	0.31
Denmark	Bredtmann and Smith (2018)[1]	Sisters	0.39
Finland	Grätz et al. (2021)	Mixed	0.36
Germany	Schnitzlein (2014)[1]	Brothers	0.66
Germany	Schnitzlein (2014)[12]	Sisters	0.55
Netherlands	Sieben, Huinink, and De Graaf (2001)[12]	Mixed	0.45
Norway	Raaum, Salvanes, and Sørensen (2006)[2]	Brothers	0.42
Norway	Raaum et al. (2006)[2]	Sisters	0.46-0.47
Norway	Björklund, Lindahl, and Lindquist (2010)[1]	Mixed	0.41
Norway	Björklund and Salvanes (2011)[2]	Mixed	0.40-0.42
Norway	Grätz et al. (2021)	Mixed	0.41
Sweden	Björklund, Jäntti, and Lindquist (2009)[2]	Brothers	0.46-0.48
Sweden	Björklund and Jäntti (2012)[2]	Sisters	0.39
Sweden	Björklund and Jäntti (2012)[3]	Brothers	0.46
Sweden	Björklund and Jäntti (2012)[1]	Sisters	0.40
Sweden	Hällsten and Thaning (2018)[2]	Mixed	0.38
Sweden	Grätz et al. (2021)	Mixed	0.44
UK	Grätz et al. (2021)	Mixed	0.42
USA	Conley and Glauber (2008)[2]	Mixed	0.63
USA	Mazumder (2008)[12]	Mixed	0.60
USA	Grätz et al. (2021)	Mixed	0.51

Notes: [1] Estimates compiled in Björklund and Jäntti (2020). [2] Estimates compiled in Grätz et al. (2021). Several estimates because of reporting estimates on different cohorts. [3] Björklund and Jäntti (2020) reports .43 but the paper has .46.

Figure A1: Comparison of estimates using years of schooling to those obtained with levels of education after imputing years of schooling

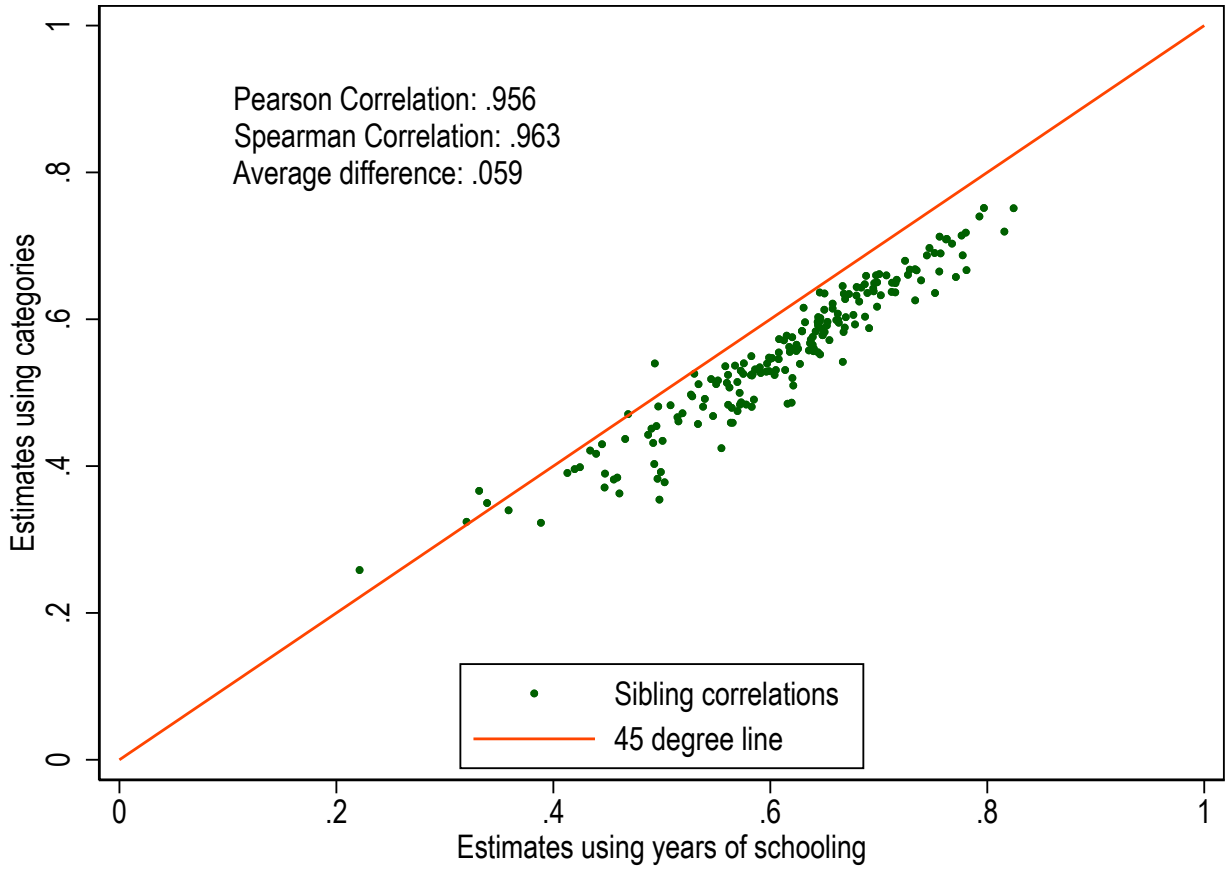


Figure A2: Sensitivity of sibling correlations to controlling for age

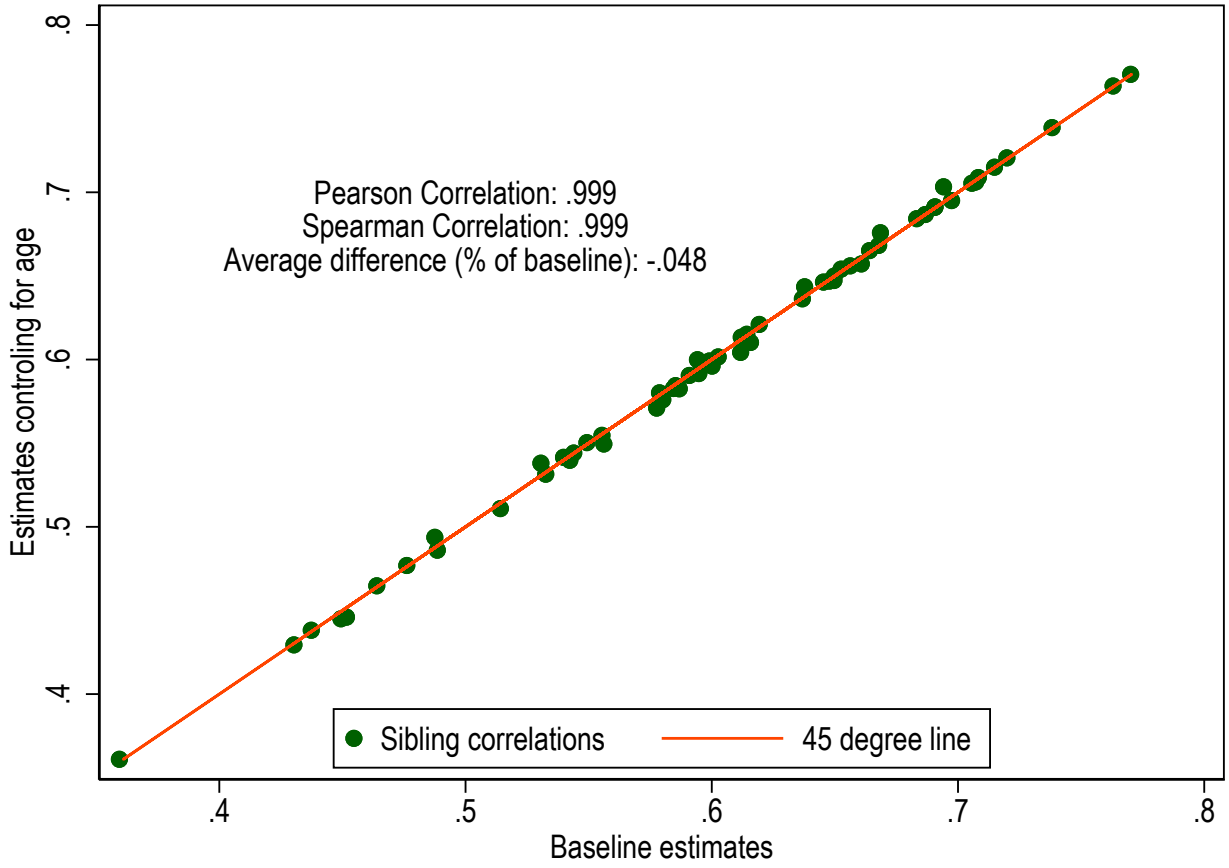


Figure A3: Sensitivity of sibling correlations to age cutoff

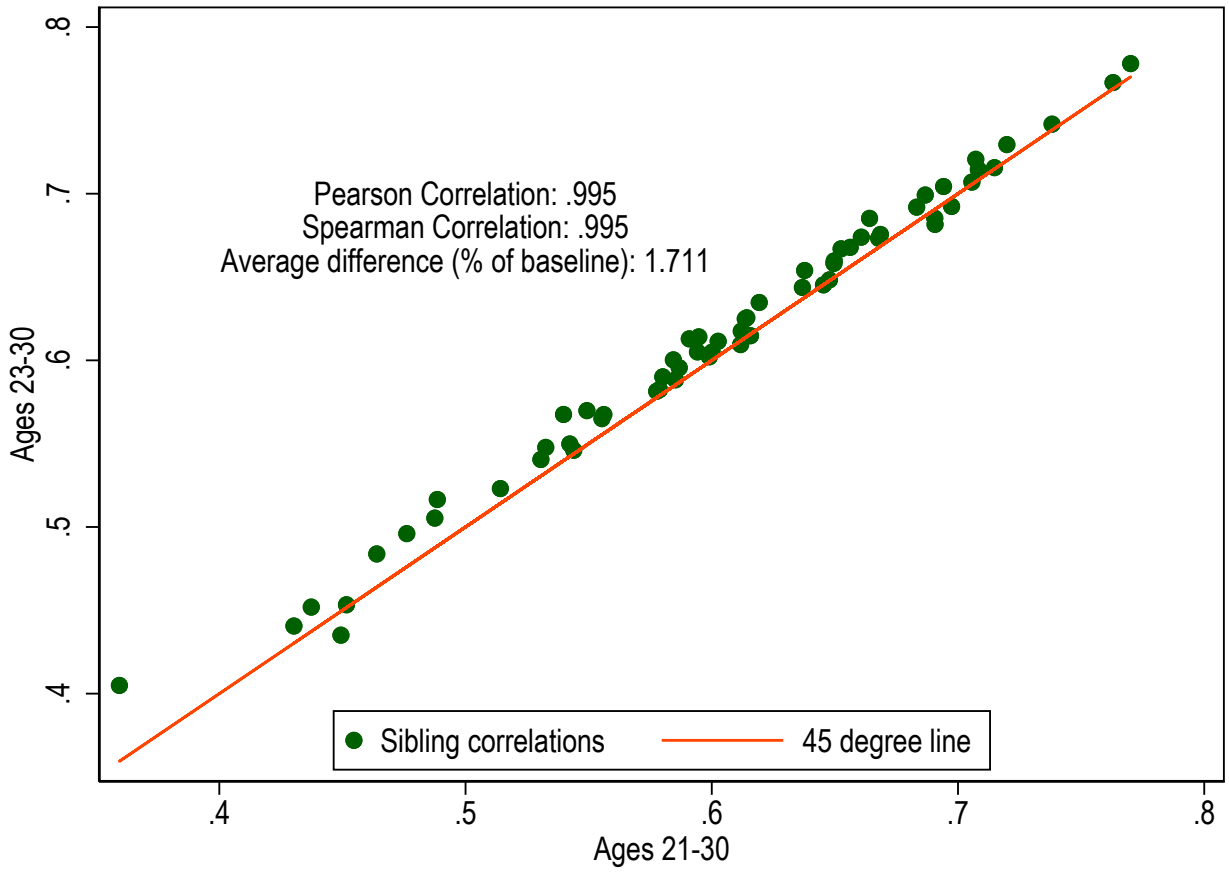


Figure A4: Sibling correlations over time estimated from years of schooling

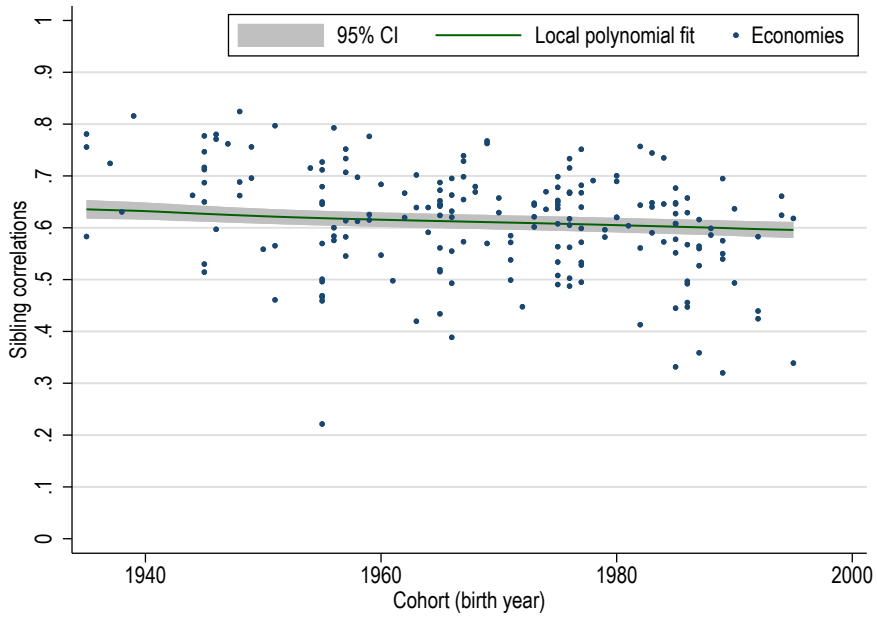


Figure A5: Sibling correlations over time (economies with at least four censuses)

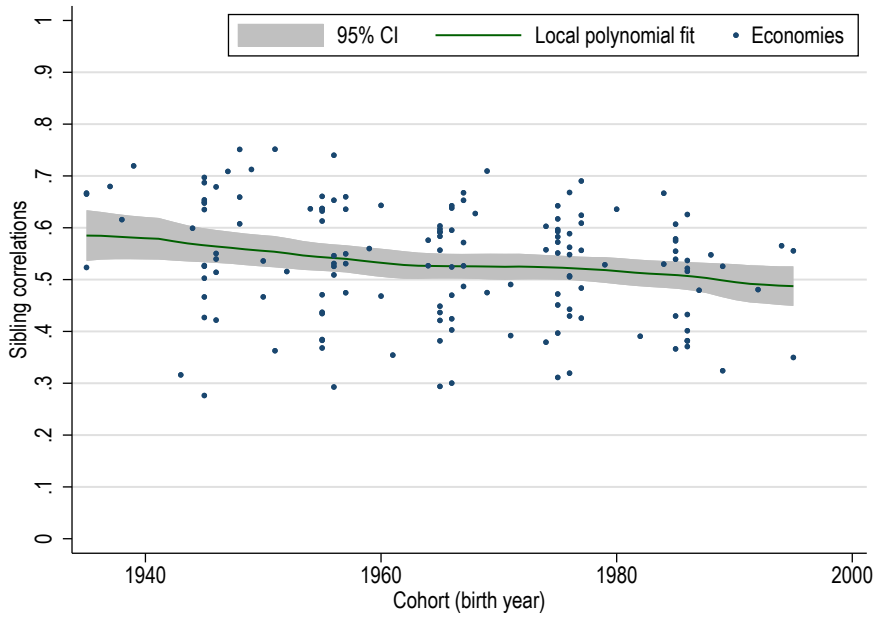


Figure A6: Sibling correlations over time by income level

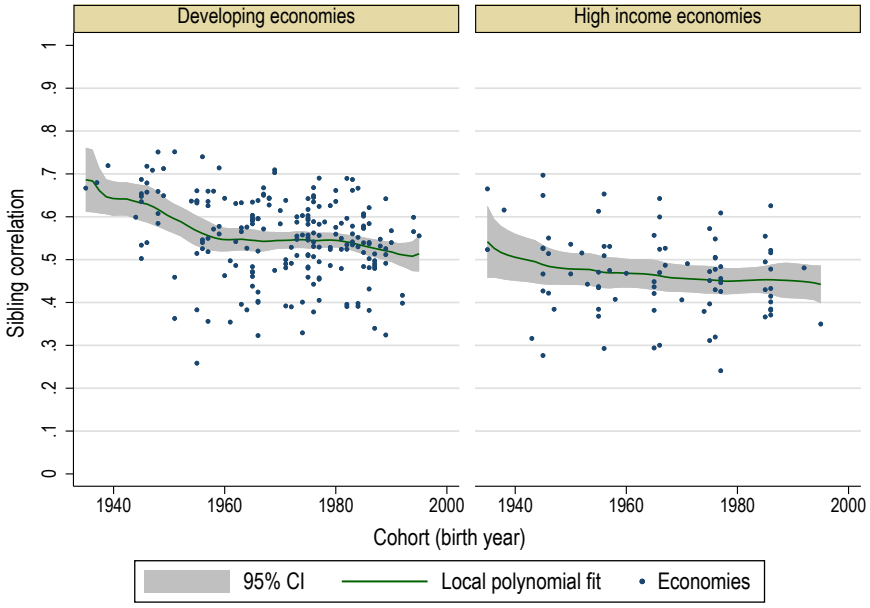


Figure A7: Sibling correlations over time by region

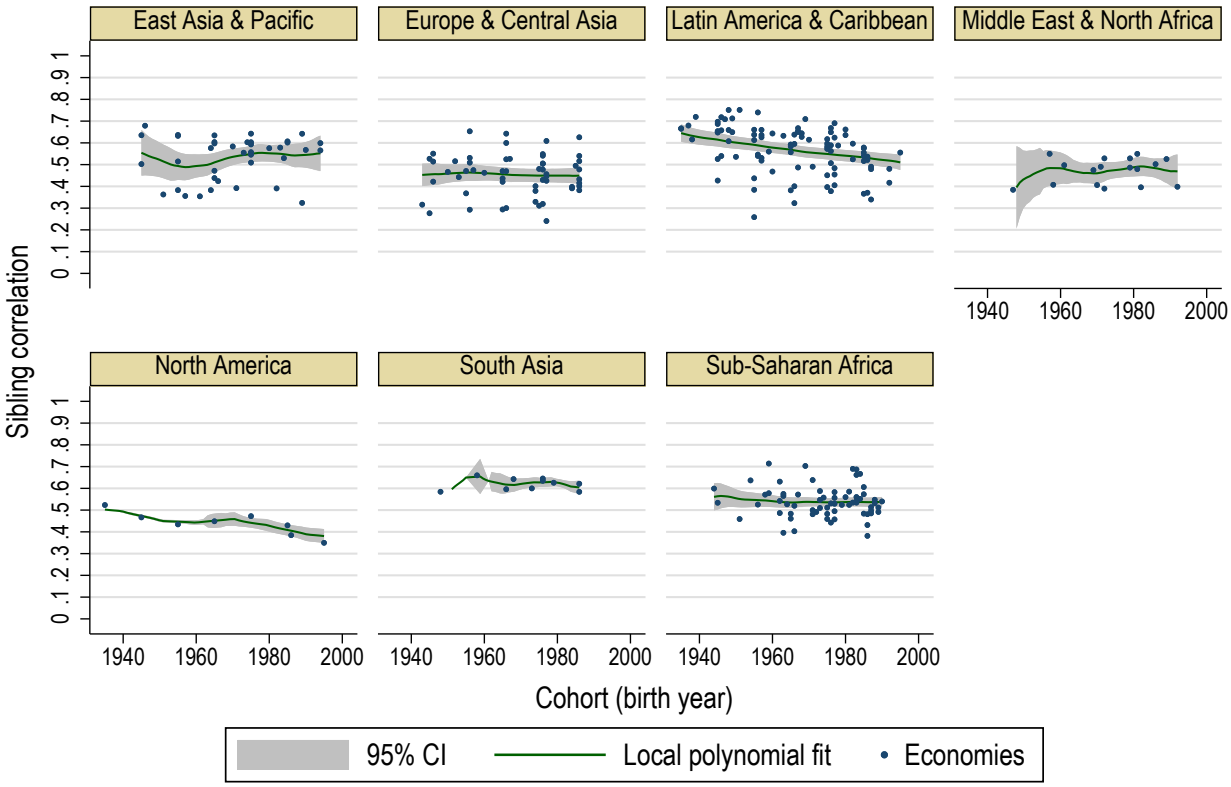


Table A2: Country coverage by income level

	Levels of education		Years of schooling	
Income group/Region	Countries	Population	Countries	Population
High income	21	60.61	7	26.18
Developing countries	73	90.97	57	57.10
East Asia & Pacific	12	98.62	10	30.55
Europe & Central Asia	4	39.79	0	0.00
Latin America & Caribbean	20	99.89	19	99.79
Middle East & North Africa	6	69.05	2	10.56
South Asia	4	96.99	3	85.28
Sub-Saharan Africa	27	78.66	23	69.27
World	94	85.17	64	51.20

Notes: The table shows the number of countries covered by the data set (with information about levels of education as well as years of schooling) and the population share that they account for. The country classification follows World Bank's income grouping and geographic regions for fiscal year 2025 available at: https://datahelpdesk.worldbank.org/knowledgebase/articles/906519-world-bank-country-and-lending-groups.

Table A3: Sibling correlations by income level

Region	Average	First cohort	Last cohort	Countries
High income	0.47	0.48	0.43	21
Developing countries	0.55	0.55	0.54	73
East Asia & Pacific	0.53	0.54	0.56	12
Europe & Central Asia	0.45	0.43	0.45	4
Latin America & Caribbean	0.59	0.61	0.54	20
Middle East & North Africa	0.49	0.49	0.50	6
South Asia	0.62	0.62	0.61	4
Sub-Saharan Africa	0.53	0.54	0.55	27

Notes: The table reports simple averages of our estimates from each region by income level. Columns "Average", "First cohort", and "Last cohort", respectively, use all the birth cohorts available for each country, only the first cohort (or census sample) available from each country, and the latest birth cohort (or census sample) available from each country. Column "Countries" reports the number of countries with data for each region.