Inequality in Life Expectancy Across Small Areas in the United States and Colombia

Natalia Serna, PhD

Corresponding author, E-mail: nserna@stanford.edu, Phone number: 608-215-7841, Affiliation: Department of Health Policy, Stanford University, Address: 615 Crothers Way, Stanford, CA, 94305

Iván Mejía-Guevara, PhD

E-mail: gauss75@gmail.com, Affiliation: Research Consultant, Palo Alto, CA, 94305

Paul Rodríguez-Lesmes, PhD

E-mail: paul.rodriguez@urosario.edu.co, Affiliation: Department of Economics, Rosario University, Bogotá, Colombia

Eran Bendavid, MD MS

E-mail: ebd@stanford.edu, Affiliation: Department of Medicine, Stanford University, Stanford, CA, 94305

Grant Miller, PhD

E-mail: ngmiller@stanford.edu, Affiliation: Department of Health Policy, Stanford University, Stanford, CA, 94305

Abstract

Background. Inequality in life expectancy within high-income countries has received considerable attention in the last decade, but much less is known about this topic within lower-income countries. This paper compares inequality in life expectancy at birth in Colombia, a country with unusually rich small-area census data, with similar measures of inequality in the U.S.

Methods. We used publicly available census tract-level data from Colombia's 2018 population census, which report population counts and deaths by five-year age groups as well as measures of economic position. For comparison, we draw on existing census tract-level estimates of life expectancy at birth for the U.S. produced by the Centers for Disease Control and Prevention. We estimated age-specific mortality rates for each Colombian census tract using the TOPALS model (Tool for Projecting Age-Specific Rates Using Linear Splines). Our primary outcome is life expectancy at birth, which we compute by applying standard demographic methods to the estimated age-specific mortality schedules.

Results. There is a moderate gap of 2 to 5 years in average life expectancy at birth between the U.S. and Colombia, but Colombians in the bottom 1% of the life expectancy distribution live 15 years less than their American counterparts. We investigate the role of income in explaining this inequality in life expectancy, finding that the income gradient in life expectancy at birth is flatter in Colombia than in the U.S.—and that the extremely low life expectancies in Colombia are present across all income groups.

Conclusion. Our results suggest that factors other than income play a larger role in longevity in Colombia (and potentially in other lower-income countries like Colombia) than they do in the U.S., underscoring the importance of broader social determinants in shaping health outcomes in lower-income countries.

Keywords: Life expectancy, Household income, Social determinants of health

1 Background

Within high-income countries, a robust body of research documents substantial inequality in life expectancy, with income playing a central role (Kinge et al., 2019; Currie and Schwandt, 2016; Mortensen et al., 2016; Mackenbach et al., 2008). For example, in the U.S., the difference in life expectancy at age 40 between the top and bottom one percent of the income distribution is 15 years for men and 10 years for women (Chetty et al., 2016). This income gradient in life expectancy has drawn considerable attention in public health and policy circles, pointing to critical social determinants of health.

In contrast, far less is known about inequality in life expectancy in low- and middle-income countries (LMICs). Existing research has examined cross- and within-country differences in life expectancy with a focus on individual cities or age groups (Rosero-Bixby, 2008; Bilal et al., 2019). But the shape and dispersion of socioeconomic gradients in life expectancy across the population remain poorly characterized in these settings.

Understanding this relationship is especially important in LMICs, where income inequality is often substantially greater than in high-income countries. In Brazil and Chile, for example, earnings in the top decile of the income distribution are roughly 30 times higher than those of the bottom half of the population – a disparity nearly three times larger than in the U.S. (Chancel et al., 2022). Such extreme income inequality raises the possibility that socioeconomic gradients in longevity may differ in both magnitude and shape from those documented in high-income countries. Whether they do carries direct policy stakes: if income strongly determines survival, then individual transfers and tax benefits or government transfers to municipalities are plausible instruments for narrowing longevity gaps across individuals and places, respectively; if not, such policies may prove ineffective, and interventions addressing broader social determinants of health may be warranted, particularly in LMICs marked by violence and internal displacement (Singh et al., 2021).

This study addresses this gap in the literature by comparing within-country inequality in life expectancy at birth, and its income gradient, between a middle-income country and the U.S. The focus is on Colombia, whose 2018 population census provided, for the first time,

census tract-level data on both socioeconomic status and age-specific mortality.[1] Colombia combines high income inequality with universal health insurance (Davalos et al., 2024) – opposite to the U.S. – making it well-suited for testing whether the income-longevity gradient documented in the U.S. generalizes to lower-income settings.

In this study we aimed to estimate life expectancy at birth for each census tract in Colombia; compare the resulting distribution with that of the U.S.; and investigate how differences within and between countries vary across the income distribution.

2 Methods

2.1 Data and Statistical Methods for Colombia

We used publicly available data from the 2018 Colombian population census, provided by the National Administrative Department of Statistics (DANE), covering all 504,996 census blocks (*manzanas*) in Colombia's 1,121 municipalities. For each census block, the data report population and deaths in 2017 by sex and 5-year age group, as well as socioeconomic status (SES), a 0-6 classification of residential properties assigned by municipal governments to assess utility fees and subsidies, based on housing characteristics and access to public services.

We grouped census blocks within each of Colombia's 33 states into clusters comparable in size to U.S. census tracts using K-means clustering on geographic centroid coordinates and SES decile (see details in Appendix A). We set the number of blocks within each cluster to 100.

For each resulting cluster, hereafter "census tract," we estimated 5-year age-specific mortality schedules using de Beer (2012)'s relational TOPALS spline model, which fits a linear spline to the ratios of observed age-specific mortality rates to a standard schedule. Estimation proceeded in four nested steps: state, urban/rural area, socioeconomic decile, and census tract. Each step's estimates served as the standard schedule for the next, initialized with

[1] We discuss the advantages of using census data over vital statistics records in Appendix D. We also show that our results are robust to excluding uncertified deaths from the census data (Appendix Figure 10), which makes mortality counts more comparable to those reported in vital statistics.

the national schedule from the 2024 UN World Population Prospects. We then computed life tables and life expectancy at birth using standard demographic methods. Appendix A provides details.

Sensitivity analyses. We conducted several analyses to assess the robustness of the Colombian estimates. Estimated life expectancy distributions showed accurate in- and out-of-sample fit (Appendix Figures 4–8). Life expectancy estimates were similar when the TOPALS relational model was replaced with a D-Spline model (Schmertmann, 2021) or a Log-Quadratic model (Wilmoth et al., 2012) (Appendix F), although the Log-Quadratic model yielded higher mortality rates at ages 65 and older, likely reflecting its calibration to data from high-income countries. Results were also similar using clusters of 50 or 200 census blocks rather than 100 (Appendix Figure 11) and after excluding uncertified deaths (Appendix Figure 10). Finally, 95% CIs computed by simulation from the asymptotic distribution of the TOPALS parameters showed that the long left tail of the Colombian distribution was not attributable to estimation uncertainty.

2.2 Data and Statistical Methods for the United States

For the U.S., we obtained census tract-level estimates of life expectancy at birth for 2010–2015 from the U.S. Small Area Life Expectancy Estimates Project (USALEEP) and matched them to 2011–2015 American Community Survey estimates of median household income at the census tract level. Household income is defined as total pre-tax money income received over the previous 12 months by the householder and all other household members aged 15 years or older. Because USALEEP estimates are already reported at the tract level, no clustering was required.

2.3 Measures of Economic Position

Because SES is unavailable for U.S. census tracts and household income is unavailable for most Colombian census tracts, we assessed the comparability of the two measures in Bogotá,

where both exist.[2] Household income in Bogotá comes from DANE's 2011 National Quality of Life Survey and is measured as pre-tax monetary income received in the past month by all household members aged 12 or older, annualized to match the U.S. measure. In Bogotá, SES and household income are strongly positively correlated, have similar dispersion, and produce similar gradients in life expectancy at birth (Appendix Figure 2, Panels A–C), supporting their use as comparable markers of economic position.

3 Results

Of Colombia's 504,996 census blocks, 397,381 had nonmissing SES data; the mean block population was 92 residents, and 76% of blocks recorded zero deaths (Table 1). Mortality rates estimated at this scale would be dominated by sampling noise, since a single death in a block of 92 residents implies a crude death rate more than ten times the national average. Aggregating blocks into larger units before estimation addresses this problem. Implementing our K-means clustering approach yielded 4,127 clusters with a mean population of 8,846 and a mean of 42 deaths, reducing the proportion of units with zero deaths to 1%. The U.S. comprised 63,828 census tracts with a mean population of 4,442. Mean life expectancy at birth was 78.4 years in the U.S. vs 74.1 years in Colombia, with less dispersion across geographic units in the U.S. (SD, 3.95 vs 6.50 years in Colombia).

We observed substantial variation in life expectancy at birth across census tracts within the largest cities of each country (Figure 1). Life expectancy at birth ranged from 41 to 84 years in Bogotá and from 59 to 94 years in New York. In Bogotá, census tracts with the lowest life expectancy (dark red) were concentrated in the urban periphery. We observed similar patterns in other major cities in both countries (Barranquilla, Bucaramanga, and Manizales in Colombia; Houston, Phoenix, and Philadelphia in the U.S.; Appendix Figure 12). The within-city variation points to local determinants of mortality that operate below the national or regional levels. Appendix Figure 13 shows maps for entire countries.

The distribution of life expectancy at birth across census tracts differed markedly between countries. The distribution for Colombia (red) was shifted left relative to the U.S. (white)

[2] Bogotá represents approximately 20% of the Colombian population and 10% of all census blocks.

and exhibited a longer left tail, including census tracts with life expectancy at birth below 50 years (Figure 2, Panel A). Low life expectancies in Colombia were not concentrated in rural census tracts; the left tail of the distribution was similar in urban and rural areas (Appendix Figure 14, Panel A). However, unlike the U.S., these low life expectancies could be attributed to Colombia's long history of internal armed conflict and drug trade wars, which have displaced large populations into urban areas and depressed rural ones.

Life expectancy by percentile of the within-country distribution was flatter in the U.S. than in Colombia (Figure 2, Panel B): the gap between the top and bottom percentiles was 20 years in the U.S. vs 34 years in Colombia. The between-country gap in life expectancy at birth also varied sharply across the distribution (Panel C). For the top 90% of the distribution, the difference between the two countries was 5 years or less; in the left tail, the gap widened rapidly, reaching 15 years [95% CI, 13.2–16.4] at the bottom percentile.

3.1 Life Expectancy by Measures of Economic Position

Because measures of economic position vary across census tracts, they are natural candidates for explaining the small-area geographic variation in life expectancy at birth. Our findings show that the association of life expectancy at birth with household income was steeper in the U.S. than the association with SES in Colombia (Figure 3). A 1-percentile increase of a census tract's position in the income distribution was associated with 0.032 (SE, 0.005) additional years of life in the U.S. relative to Colombia. The difference in gradients was not attributable to the Colombian SES measure being coarser than household income, nor to Colombian census tracts being more rural than U.S. tracts. First, the U.S.–Bogotá difference in gradients, which compares household income in both settings, exceeded the U.S.–Colombia difference (Appendix Figure 3), which is the opposite of what a coarseness artifact would produce. Second, restricting Colombia to urban census tracts yielded an SES gradient in life expectancy at birth of similar magnitude to that for all census tracts (Appendix Figure 14, Panel B).

Figure 4 presents the distribution of life expectancy at birth by household income decile in the U.S. and by SES decile in Colombia. In Colombia, median life expectancy increased weakly with SES, and the gradient was driven mainly by census tracts in the lower range

of each decile (Panel A). Even conditional on SES decile, the distribution retained a long left tail: census tracts with low life expectancy were present in all deciles. In the U.S., by contrast, gains across income deciles occurred throughout the within-decile distribution, and within-decile variation was smaller than in Colombia (Panel B). Household income explained 36% of the variation in life expectancy across census tracts in the U.S. while SES explained only 7% in Colombia (R^2 from country-specific regressions).

Panel C compares the life expectancy gap distributions of the two countries separately for the top (blue) and bottom (red) income/SES deciles. To construct this figure, we first ranked census tracts in the bottom income/SES decile within each country into percentiles of life expectancy at birth. Then, at each percentile of the life expectancy distribution, we computed the difference between the U.S. and Colombian values. We repeated the same procedure for census tracts in the top income/SES decile.

Two patterns stand out. Between-country differences in life expectancy at birth were largest at the bottom of the life expectancy distribution across income/SES deciles. At the first percentile of life expectancy, census tracts in the U.S outlived Colombian census tracts by 10 years [95% CI, 6.6–12.6] in the red line and by and 28 years [95% CI, 23.8–32.1] in the blue line; by the 20th percentile these gaps had fallen to roughly 0 and 4 years, respectively. Second, the U.S. advantage is systematically larger in the top income decile than in the bottom decile, and in the bottom decile it even reverses: for example, at the 90th percentile of the life expectancy distribution, low-SES Colombian tracts outlived low-income U.S. tracts by 3 years [95% CI, 2.6–3.3].

4 Discussion

This study makes two main contributions. First, it shows that comparisons of mean life expectancy either across countries or within countries may miss important disparities present in other parts of the life expectancy distribution (Bendavid et al., 2025). The finding that the gap in first-percentile life expectancy at birth between Colombia and the U.S. was constant across income groups suggests that policies promoting upward economic mobility alone, such as income transfers or tax deductions, are unlikely to reduce the cross-country gap (consistent

with Agarwal et al., 2024).

If the pattern of long left tail in life expectancy observed in Colombia is also present in other lower-income countries after controlling for income, this means that other factors may play a larger role in health inequality compared to the U.S. (Schwandt et al., 2021). In Appendix Figure 15 we show gradients in life expectancy at birth by other characteristics in both Colombia and the U.S., identifying potential contributors to these differences (including violent deaths and forced migration). Policies focused on public safety and reintegration of displaced populations may thus deserve attention as health policy in such settings, pointing to an important direction for future research on the determinants of mortality in lower-income countries.

Second, small-area life expectancy estimates permit examination of the income-longevity association for the entire population rather than only for individuals reporting income through the public finance system. For example, studies in high-income countries document substantial disparities in life expectancy at age 40 by income because mortality data can typically be linked to income records only for these age groups (Chetty et al., 2016; Kinge et al., 2019). However, income does not influence survival only in adulthood (Udalova et al., 2022).

Although the focus on Colombia and the U.S. was motivated by differences in economic inequality and the availability of small-area mortality records, these findings may extend to other settings with high economic inequality. Countries characterized by substantial economic informality, younger populations, and exposure to conflict (Singh et al., 2021) may exhibit an income-longevity relationship similar to Colombia's. Improving survival in such settings may require moving beyond policies targeting economic inequality alone and toward those addressing broader social determinants of health.

Limitations. This study has limitations. First, the Colombian socioeconomic measure is not identical to U.S. household income; however, where both measures were available, they were highly correlated and produced indistinguishable life expectancy gradients. Second, the estimates for the two countries cover different periods and were constructed using different methods. Yet, the cross-country contrasts of interest are large and remain stable across a

range of sensitivity analyses in Colombia, making it unlikely that they are attributable to these methodological differences. Finally, findings from a single middle-income country may not generalize to all low- and middle-income settings, although the mechanisms examined, such as economic informality, violence, and internal displacement, are common to many.

5 Conclusion

In this cross-sectional study, the gap in average life expectancy at birth between Colombia and the United States was modest, but it reached 15 years at the bottom of the life expectancy distribution and persisted across all income levels. Income was a substantially weaker predictor of survival in Colombia than in the U.S. These findings suggest that reducing survival disparities in lower-income countries may require policies that address broader social determinants of health rather than income alone.

References

Agarwal, S. D., B. Lê Cook, and J. B. Liebman (2024): "Effect of cash benefits on health care utilization and health: a randomized study," *JAMA*, 332, 1455–1463.

Bendavid, E., A. K. Littleton, I. Mejía-Guevara, M. Fenyo, and G. Miller (2025): "A Global Analysis of Within-Country Health Inequalities," *JAMA Health Forum*, 6, e253611–e253611.

Bilal, U., M. Alazraqui, W. T. Caiaffa, N. Lopez-Olmedo, K. Martinez-Folgar, J. J. Miranda, D. A. Rodriguez, A. Vives, and A. V. Diez-Roux (2019): "Inequalities in Life Expectancy in Six Large Latin American Cities from the SALURBAL Study: an Ecological Analysis," *The Lancet Planetary Health*, 3, e503–e510.

Chancel, L., T. Piketty, E. Saez, and G. Zucman (2022): "World Inequality Report," *World Inequality Lab*.

Chetty, R., M. Stepner, S. Abraham, S. Lin, B. Scuderi, N. Turner, A. Berg-

eron, and D. Cutler (2016): "The Association Between Income and Life Expectancy in the United States, 2001-2014," *JAMA*, 315, 1750–1766.

Currie, J. and H. Schwandt (2016): "Inequality in mortality decreased among the young while increasing for older adults, 1990–2010," *Science*, 352, 708–712.

Davalos, M. E., C. Rodriguez Castelan, R. Munoz Moreno, M. R. Thomas, and P. Siegenthaler (2024): "Trayectorias: Prosperidad y Reducción de la Pobreza en el Territorio Colombiano (Spanish)," *World Bank*.

de Beer, J. (2012): "Smoothing and projecting age-specific probabilities of death by TOPALS," *Demographic Research*, 27, 543–592.

Kinge, J., J. Modalsli, S. Øverland, H. Gjessing, G. Kristian, C. Tollånes, A. Knudsen, V. Skirbekk, B. Strand, S. Håberg, and S. Vollset (2019): "Association of Household Income with Life Expectancy and Cause-Specific Mortality in Norway, 2005-2015," *JAMA*, 321, 1916–1925.

Mackenbach, J. P., I. Stirbu, A.-J. R. Roskam, M. M. Schaap, G. Menvielle, M. Leinsalu, and A. E. Kunst (2008): "Socioeconomic inequalities in health in 22 European countries," *New England journal of medicine*, 358, 2468–2481.

Mortensen, L. H., J. Rehnberg, E. Dahl, F. Diderichsen, J. I. Elstad, P. Martikainen, D. Rehkopf, L. Tarkiainen, and J. Fritzell (2016): "Shape of the association between income and mortality: a cohort study of Denmark, Finland, Norway and Sweden in 1995 and 2003," *BMJ open*, 6, e010974.

Rosero-Bixby, L. (2008): "The exceptionally high life expectancy of Costa Rican nonagenarians," *Demography*, 45, 673–691.

Schmertmann, C. (2021): "D-splines: Estimating rate schedules using high-dimensional splines with empirical demographic penalties," *Demographic Research*, 44, 1085–1114.

Schwandt, H., J. Currie, M. Bär, J. Banks, P. Bertoli, A. Bütikofer, S. Cattan, B. Z.-Y. Chao, C. Costa, L. González, et al. (2021): "Inequality in mortality

between Black and White Americans by age, place, and cause and in comparison to Europe, 1990 to 2018," *Proceedings of the National Academy of Sciences*, 118, e2104684118.

Singh, N. S., A. Ataullahjan, K. Ndiaye, J. K. Das, P. H. Wise, C. Altare, Z. Ahmed, S. Sami, C. Akik, H. Tappis, et al. (2021): "Delivering health interventions to women, children, and adolescents in conflict settings: what have we learned from ten country case studies?" *The Lancet*, 397, 533–542.

Udalova, V., V. Bhatia, and M. Polyakova (2022): "Association of family income with morbidity and mortality among US lower-income children and adolescents," *JAMA*, 328, 2422–2430.

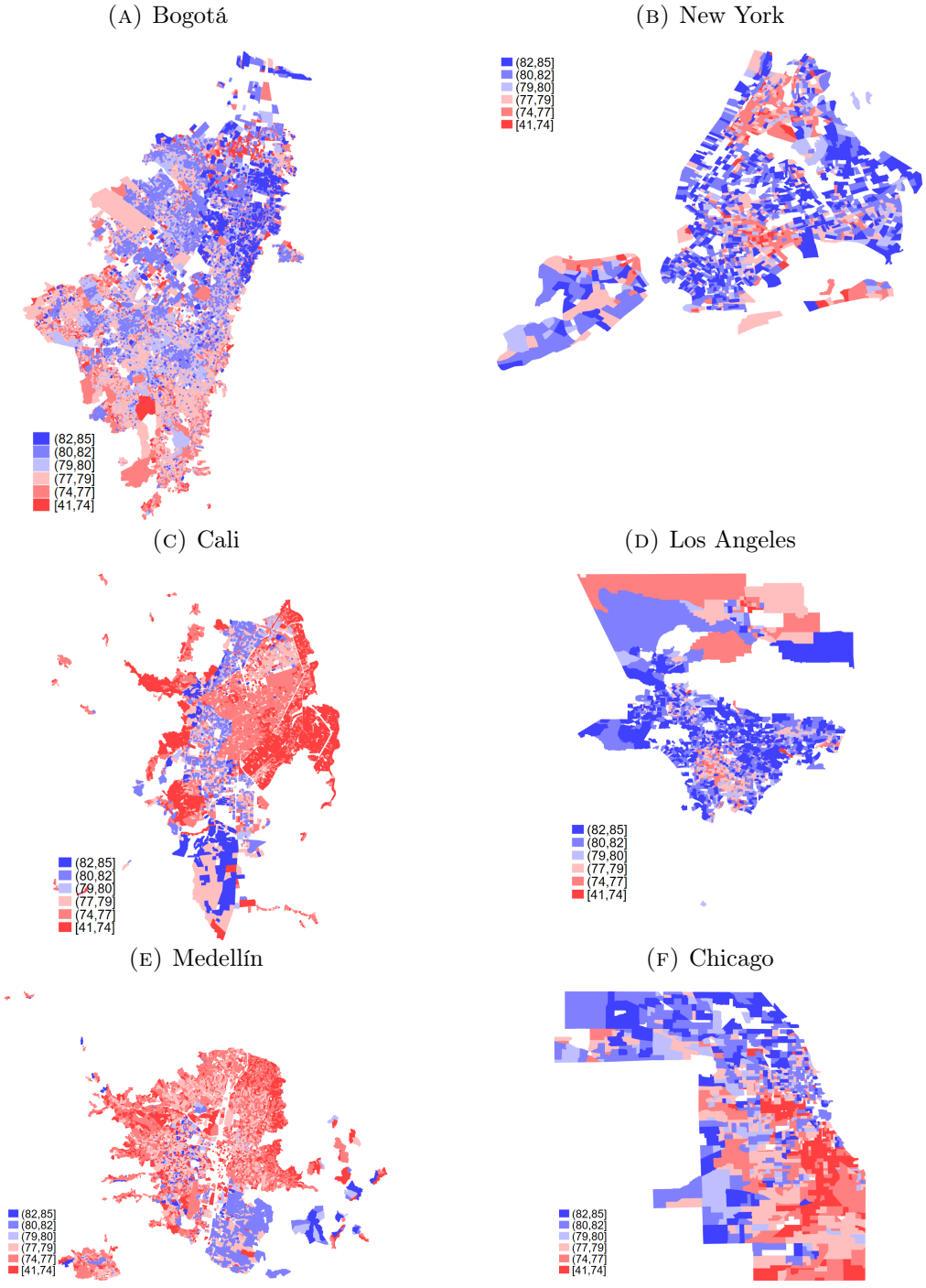
Wilmoth, J., S. Zureick, V. Canudas-Romo, M. Inoue, and C. Sawyer (2012): "A flexible two-dimensional mortality model for use in indirect estimation," *Population Studies*, 66, 1–28.

TABLE 1: Summary Statistics

Variable	Colombia		USA
	Census blocks	Clusters	Census tracts
	(1)	(2)	(3)
Life expectancy	74.31	74.10	78.40
	(5.91)	(6.50)	(3.95)
Number of deaths	0.44	42.00	—
	(1.55)	(47.01)	
Fraction with zero deaths	0.76	0.01	—
	(0.43)	(0.12)	
SES	1.79	1.76	—
	(0.91)	(0.95)	
Median per capita income (1000s)	—	—	57.86
			(28.34)
Total population	91.87	8,845.71	4,442.35
	(139.44)	(9,912.42)	(2,063.55)
Number of census blocks	397,381	96.29	
		(74.25)	
Number of census tracts		4,127	63,828

Note: Table presents the mean and standard deviation in parenthesis of the main variables in our study. In column (1) an observation is a Colombian census block conditional on those with non-missing SES information. In column (2) an observation is a cluster of Colombian census blocks, which we denote in the main text as census tract. In column (3) an observation is a. U.S. census tract. Total U.S. census tract population is obtained from the 5-year American Community Survey estimates for 2011-2015. Socieconomic status (SES) information is not available for U.S. census tracts and median per capita income is not available for Colombia census blocks.

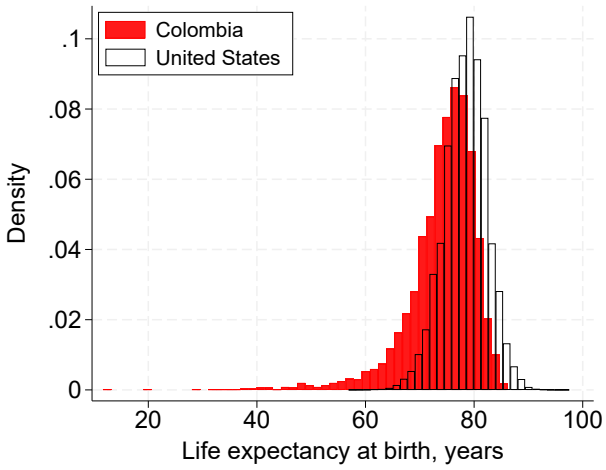
FIGURE 1: Geographic Distribution of Life Expectancy at Birth in Colombia and the United States



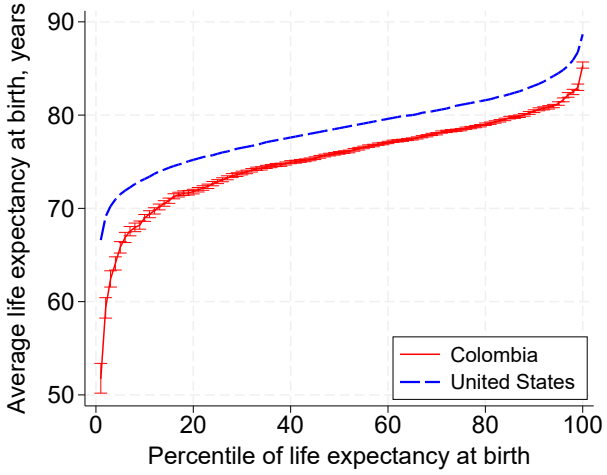
Note: Figure presents the geographic distribution of life expectancy at birth in Colombia for the cities of Bogotá, Cali and Medellín (Panels A, C, E respectively) and in the U.S. for the cities of New York, Los Angeles, and Chicago (Panels B, D, F, respectively). Darker blues represent higher values and darker reds represent lower values.

FIGURE 2: Distribution of Life Expectancy at Birth in Colombia and the United States

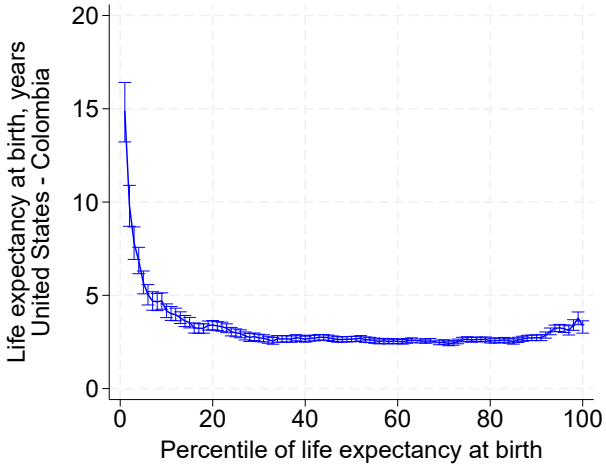
(A) Distribution of life expectancy at birth



(B) Mean life expectancy at birth by percentile

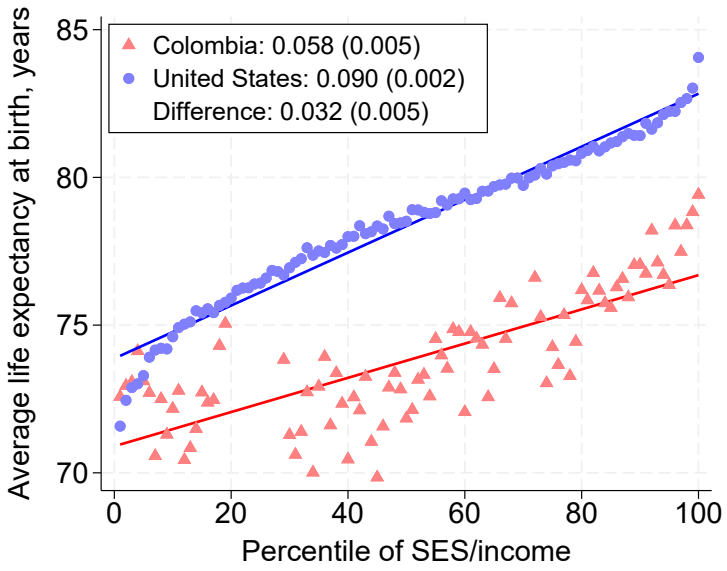


(C) Between-country gap in life expectancy at birth



Note: Panel A presents the distribution of life expectancy at birth across census tracts in Colombia (red) and in the U.S. (white). Panel B presents the mean life expectancy at birth by percentile of the distribution in Colombia (red) and the U.S. (gray). Panel C presents the gap in mean life expectancy at birth between the U.S. and Colombia for census tracts at each percentile of the distribution of life expectancy at birth. 95% confidence intervals reflect uncertainty in Colombian estimates only and were computed by simulation.

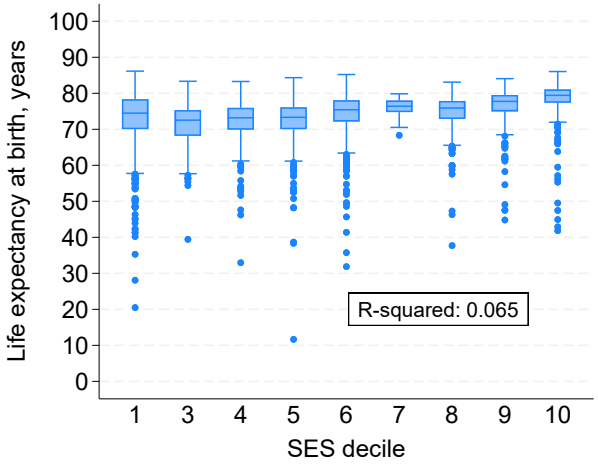
FIGURE 3: Mean Life Expectancy by SES/Income Decile in Colombia and the United States



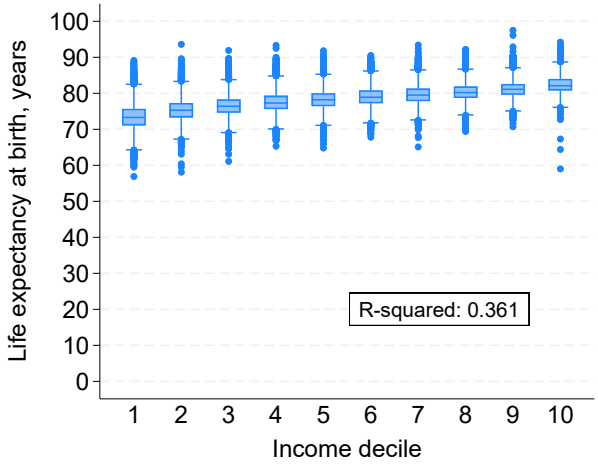
Note: Figure presents mean life expectancy at birth in Colombia (red) by percentile of socioeconomic status (SES) and in the U.S. (gray) by percentile of household income. Straight lines correspond to linear fits. Socioeconomic status in Colombia is a 0-6 categorization of residential units with the purpose of charging differential utilities fees. The text box in the figure reports the coefficient and standard error in parenthesis of a linear regression of mean life expectancy at birth on SES/household income percentile estimated separately for each country. The text box also reports the difference in slopes between the two countries and its standard error in parenthesis. To compute the difference in slopes, we estimate a linear regression pooling the data from the two countries. The dependent variable is mean life expectancy at birth, and the independent variables are SES/income percentile, a dummy variable for the U.S., and the interaction between SES/income percentile and the U.S. dummy variable. The difference in slopes is the coefficient on this interaction.

FIGURE 4: Mortality and Life Expectancy by Income in Colombia and the United States

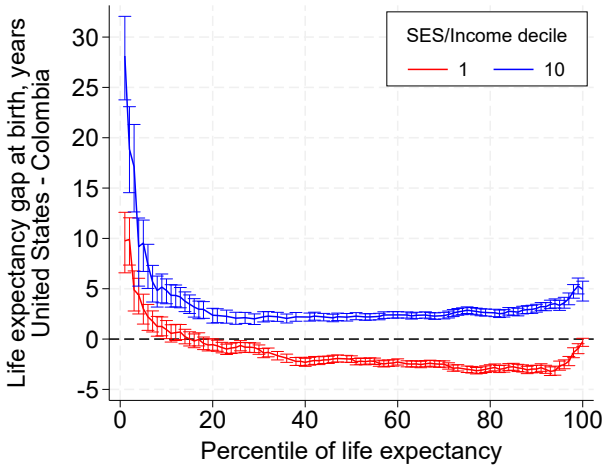
(A) Life expectancy distribution by Socioeconomic Status in Colombia



(B) Life expectancy distribution by income in the United States



(C) Life expectancy gap at birth between the United States and Colombia by economic position



Note: Panels A and B present the distribution of life expectancy at birth conditional on deciles of socioeconomic status (SES) in Colombia and on deciles of median household income in the U.S., respectively. The center line in each box represents the median of the distribution. The bottom and top of each box represent the 25th and 75th percentiles, respectively. The bottom whisker is the 25th percentile minus 1.5 times the interquartile range. The upper whisker is the 75th percentile plus 1.5 times the interquartile range. Socioeconomic status in Colombia is a 0-6 categorization of residential units with the purpose of charging differential utilities fees. Each panel reports the R-squared of a linear regression of life expectancy at birth on median household income in the U.S. and on socioeconomic status in Colombia. An observation in these regressions is a census tract. Panel C presents the gap between the U.S. and Colombia at every percentile of the distribution of life expectancy at birth conditional on top and bottom income deciles. 95% confidence intervals reflect uncertainty in Colombian estimates only and were computed by simulation.

Declarations

Ethics approval and consent to participate. Not applicable.

Consent for publication. All authors approve of the content of this article and consent for publication.

Availability of data and materials. The datasets generated and/or analyzed during in this study are available in the Harvard Dataverse https://doi.org/10.7910/DVN/BMDFWG

Competing interests. The authors declare no competing interests.

Funding. This study received no external funding.

Authors' contributions. N.S. and G.M. conceived and designed the study. P.R.-L. and N.S. assembled and curated the data. N.S. and I.M.-G. performed the statistical analyses. N.S. drafted the manuscript. E.B., G.M., I.M.-G., and P.R.-L. revised the manuscript for important intellectual content. G.M. supervised the project. All authors approved the final version.

Acknowledgements. Eran Bendavid acknowledges support from the National Institutes of Health, HD116234 and HD104835. We thank José Manuel Aburto, Hiram Beltrán-Sánchez, Maria Polyakova, and Hannes Schwandt for their valuable feedback and comments. All errors are our own.

Supplementary Materials

Appendix A Methodology Details

A.1 Clustering of census blocks

To group Colombian census blocks, we implemented K-means clustering analysis (Bishop, 2006) using R software on a sample of 403,916 census blocks after excluding 101,080 blocks that lacked socioeconomic status data. Census blocks without socioeconomic status information are those without houses or with buildings for which the estrato classification does not apply, such as shopping centers, parking lots, or parks. The clustering methodology involved four steps: (1) randomly selecting initial cluster centroids; (2) assigning each census block to the nearest centroid based on the Euclidean distance of standardized socioeconomic status, and longitude and latitude coordinates; (3) recalculating centroids as the mean of all census blocks within each cluster, and (4) iterating steps 2-3 until the centroids stabilized or the maximum number of iterations was reached.

Because the optimal number of clusters (k) and the number of census blocks per cluster were initially unknown, we approximated k by dividing the total number of census blocks by the desired cluster size, which we set at 100. We chose this cluster size to approximate the mean population size of a U.S. census tract (generally between 2,500 and 8,000 residents), preserve variability of age-specific mortality rates across census tracts, and avoid over-smoothing these mortality rates. In Appendix Figure 11 we show that our main results are robust to different cluster sizes. We initialized the algorithm with 50 random starts and a maximum of 100 iterations to ensure consistency.

A.2 Age-specific mortality rates

To estimate age-specific mortality rates for each cluster we combined Poisson regressions with the TOPALS model. We used a version of TOPALS adapted by Gonzaga and Schmertmann (2016) to our Colombian data. Specifically, we considered mortality and population counts for the following age groups: 0–4, 5–9, 10–14, . . . , 70–74, 75–79, 80–84, and 85 or older. Then,

we fitted logarithmic mortality rates for these age groups by adding a linear spline function with seven parameters $\alpha_0 \ldots \alpha_6$ to the standard mortality schedule. The seven parameters accommodate a variety of shapes of the standard mortality schedule in small geographic areas. We estimated these parameters by maximizing a penalized Poisson likelihood function of the number of age-specific deaths, conditional on age-specific population. The Poisson regression approach has proven to be effective for estimating mortality rates for specific age groups where there are no death counts. Gonzaga and Schmertmann (2016) provide more details on this approach.

We implemented our methodology in four nested steps as follows. First, we obtained Colombia's standard age-specific mortality schedule (measured on the logarithmic scale) in 2024 from the United Nations' World Population Prospect (United Nations, 2024). Using this standard schedule, we estimated age-specific mortality rates for every Colombian state using the TOPALS model. Second, we estimated age-specific mortality rates for rural/urban areas within each state using as standard mortality schedule the state-level results from the first step (see Appendix Figure 4). Third, we estimated age-specific mortality rates by rural/urban area, state, and socioeconomic status decile using as standard mortality schedule the results from the second step (see Appendix Figures 5 to 7). Finally, the mortality rates in the third step were used as standard schedules to fit age-specific mortality rates for all census tracts within rural/urban area, state, and socioeconomic status decile. We applied standard demographic methods (Preston et al., 2000) to estimate life expectancy at birth from these resulting age-specific mortality rates.

A.3 Confidence Intervals

For each census tract, we obtained the variance-covariance matrix of the TOPALS offset parameters by inverting the Hessian of the penalized Poisson log-likelihood at its maximum. One thousand offset vectors were then drawn from the corresponding multivariate normal distribution, each draw was converted to a log-mortality schedule and a life expectancy at birth, and the 2.5th and 97.5th percentiles of the resulting life expectancy draws form the interval.

Appendix B Comparison of Socioeconomic Status and Household Income in Colombia

We complement the 2018 Colombian census data with information from DANE's 2011 National Quality of Life Survey (*Encuesta Nacional de Calidad de Vida*, or ECV). The ECV characterizes living conditions for a representative sample of Colombians, and importantly, includes a direct measure of household income for households in Bogotá. The 2011 ECV wave was the only survey wave reporting income by census block in this city. To perform analyses of life expectancy by income, we link census blocks in Bogotá from the 2018 population census with census blocks from 2011 ECV.

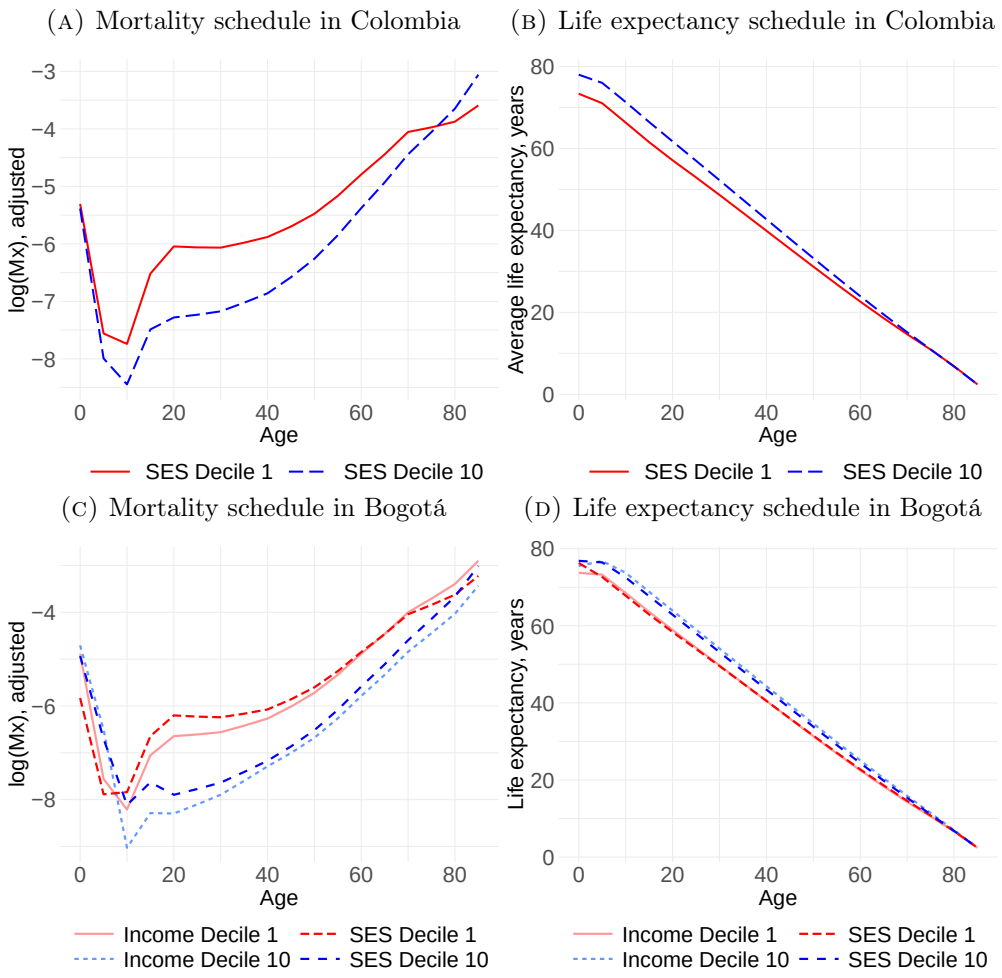
Besides our main clustering approach that groups census blocks by geographic distance and socioeconomic status, for Bogotá we additionally clustered census blocks by mean household income and geographic distance. Hence, only for the city of Bogotá we have two sets of clusters: one controlling for socioeconomic status that results in 388 clusters with a mean number of people per cluster equal to 18,387 and one controlling for household income resulting in 50 clusters with a mean number of people per cluster equal to 24,907.

In Appendix Figure 1 we compare age-specific life expectancy estimates by household income and SES. Our goal is to determine whether disparities in life expectancy at birth by income closely match those by SES in Bogotá, and whether life expectancy patterns in Bogotá by SES align with those in the entire country. Panels A and B depict the mean log mortality rate and corresponding mean life expectancy estimates by 5-year age groups among census tracts in the top and bottom deciles of SES in Colombia, respectively. Panels C and D present plots for Bogotá that are otherwise analogous to Panels A and B, stratified by both socioeconomic decile and income decile. In general, the distributions in Bogotá by SES are similar to those for the entire country. Moreover, within Bogotá, the distributions by SES show broadly similar patterns with respect to differences in life expectancy relative to those reported by income.

In Bogotá, Appendix Figure 2, Panel A shows that SES and income are high correlated, Panel B shows that both measures have similar dispersion, and Panel C shows that the gradient in life expectancy at birth by income is statistically indistinguishable from the

gradient by SES. Appendix Figure 3 compares the gradient in life expectancy at birth by household income in Bogotá against the U.S. finding a similar pattern as the comparison with the gradient in Colombia by SES.

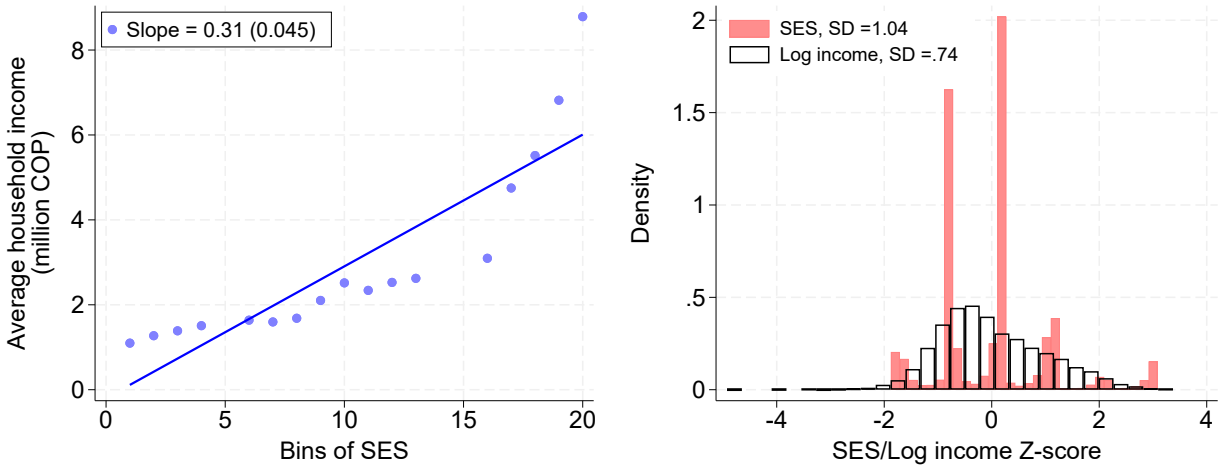
Appendix Figure 1: Mortality and Life Expectancy by Socioeconomic Status and Income in Colombia



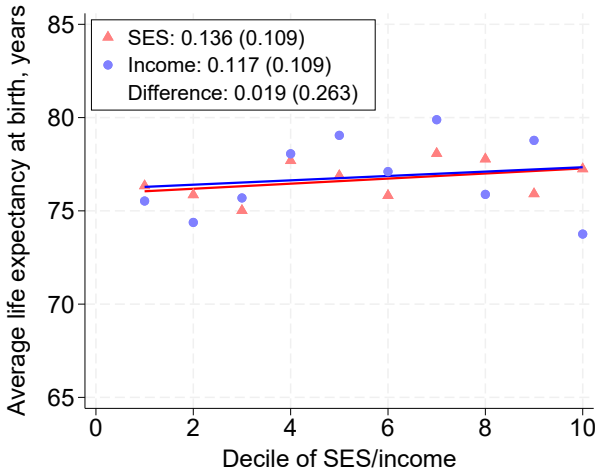
Note: Panel A shows the mean log mortality rate by 5-year age groups across census blocks in the top (blue) and bottom (red) deciles of socioeconomic status in Colombia. Panel B presents the implied mean life expectancy by 5-year age groups across census blocks in the top (blue) and bottom (red) deciles of socioeconomic status. Socioeconomic status is a 0-6 categorization of residential units with the purpose of charging differential utilities fees. Panel C shows the mean log mortality rate by 5-year age groups across census blocks in the top (solid dark blue) and bottom (solid dark red) deciles of socioeconomic status and in the top (dashed light blue) and bottom (dashed light red) deciles of household income for the city of Bogotá. Panel D presents the implied mean life expectancy by 5-year age groups across census blocks in the top (solid dark blue) and bottom (solid dark red) deciles of socioeconomic status and in the top (dashed light blue) and bottom (dashed light red) deciles of household income.

Appendix Figure 2: Comparison of SES and Household Income in Bogotá

(A) Correlation between SES and household income (B) Distribution of SES and log of household income

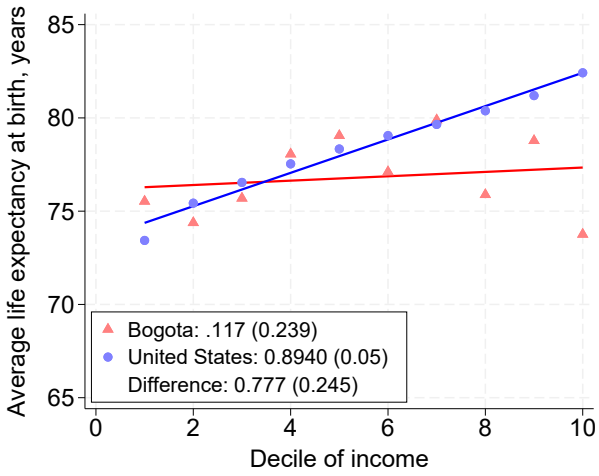


(C) SES vs Income gradient in Bogotá



Note: Panel A shows average household income by 50 equally-sized bins of SES in Bogotá. The blue line represents a linear fit and the text at the top left corner represents the coefficient and standard error of this linear regression. Panel B shows the distribution of standardized SES in red and standardized log household income in black for Bogotá. Variables are standardized by subtracting the mean and dividing by standard deviation. Panel C presents average life expectancy at birth in Bogotá by SES decile (in red) and by household income decile (in blue). Straight lines correspond to linear fits. The text box in this panel reports the coefficient and standard error in parenthesis of a linear regression of average life expectancy at birth on household income decile and on SES decile. The difference in slopes and its standard error is also reported.

Appendix Figure 3: Mean Life Expectancy by Income Percentile in Bogotá and the United States

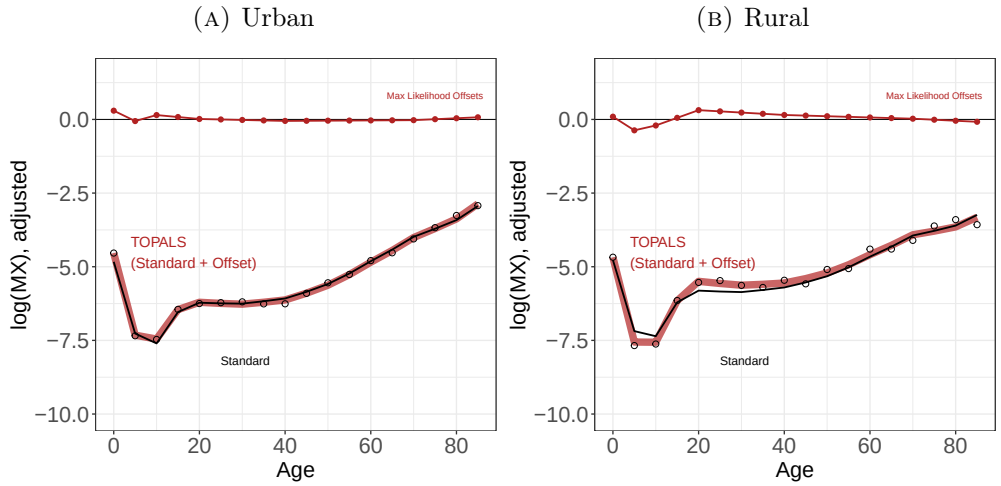


Note: Figure presents average life expectancy at birth in Bogotá (red) and in the U.S. (gray) by household income decile. Straight lines correspond to linear fits. The text box in the figure reports the coefficient and standard error in parenthesis of a linear regression of average life expectancy at birth on household income decile estimated separately for Bogotá and the U.S. The text box also reports the difference in slopes and its standard error in parenthesis. To compute the difference in slopes, we estimate a linear regression pooling the data from the two geographies. The dependent variable is average life expectancy at birth, and the independent variables are income decile, a dummy variable for the U.S., and the interaction between income decile and the U.S. dummy variable. The difference in slopes is the coefficient on this interaction.

Appendix C Clustering Methodology Fit

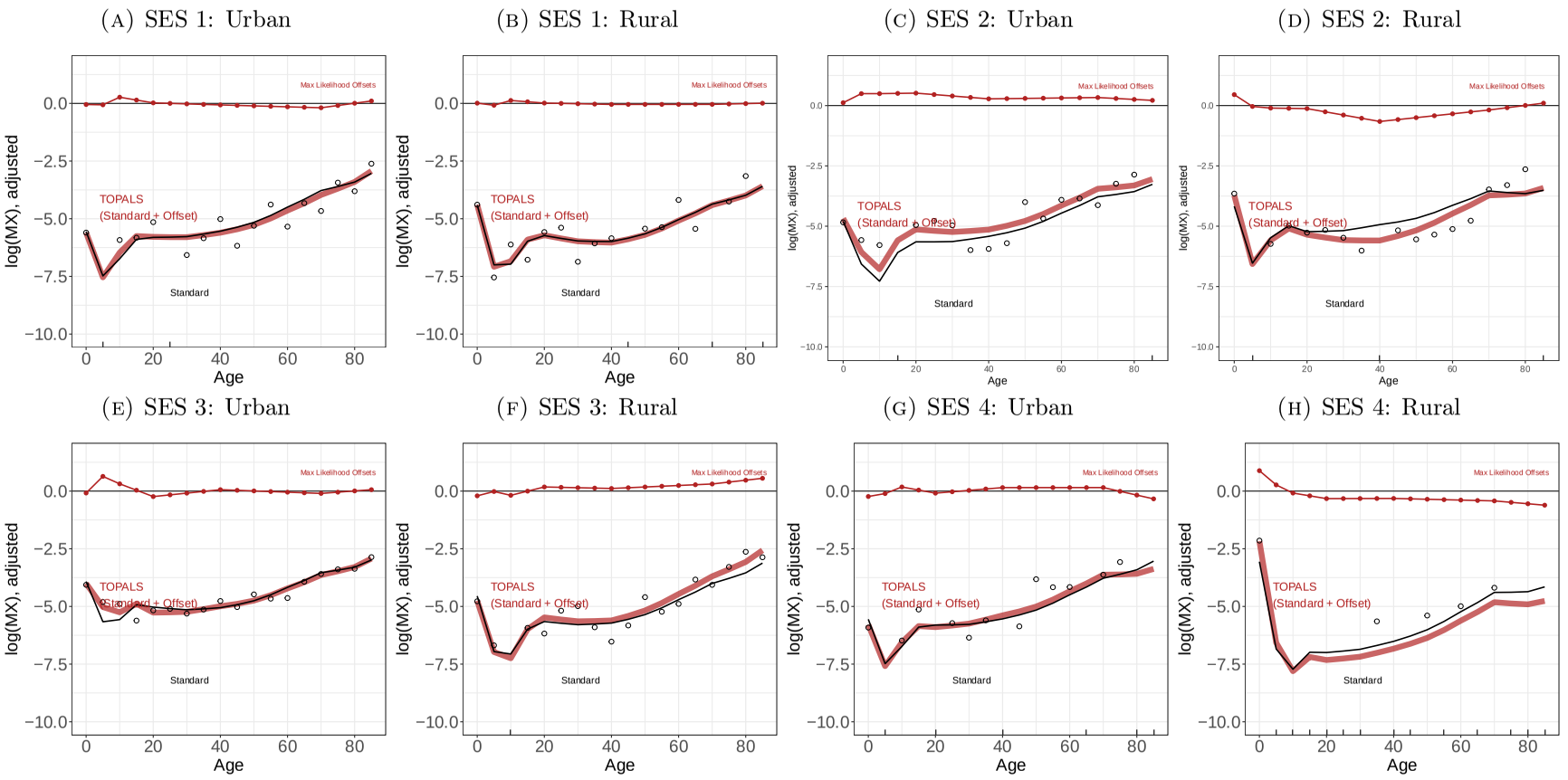
This Appendix presents measures of the fit of our clustering methodology. Appendix Figure 4 reports the fit by urban and rural regions in Colombia. Appendix Figures 5 to 7 break down the model fit by rurality and socioeconomic status decile. Finally, Appendix Figure 8 reports the out-of-sample fit of our life expectancy distributions by socioeconomic status decile. To construct these figures we first implement our K-means clustering approach setting the number of census blocks per cluster to 200. Then we randomly split each cluster into two clusters with 100 census blocks as in the main text. In the first sample we estimate the parameters of the TOPALS model (following our nested approach) and then use these estimates to predict age-specific mortality rates in the second sample.

Appendix Figure 4: Standard mortality profiles in urban and rural regions, Colombia



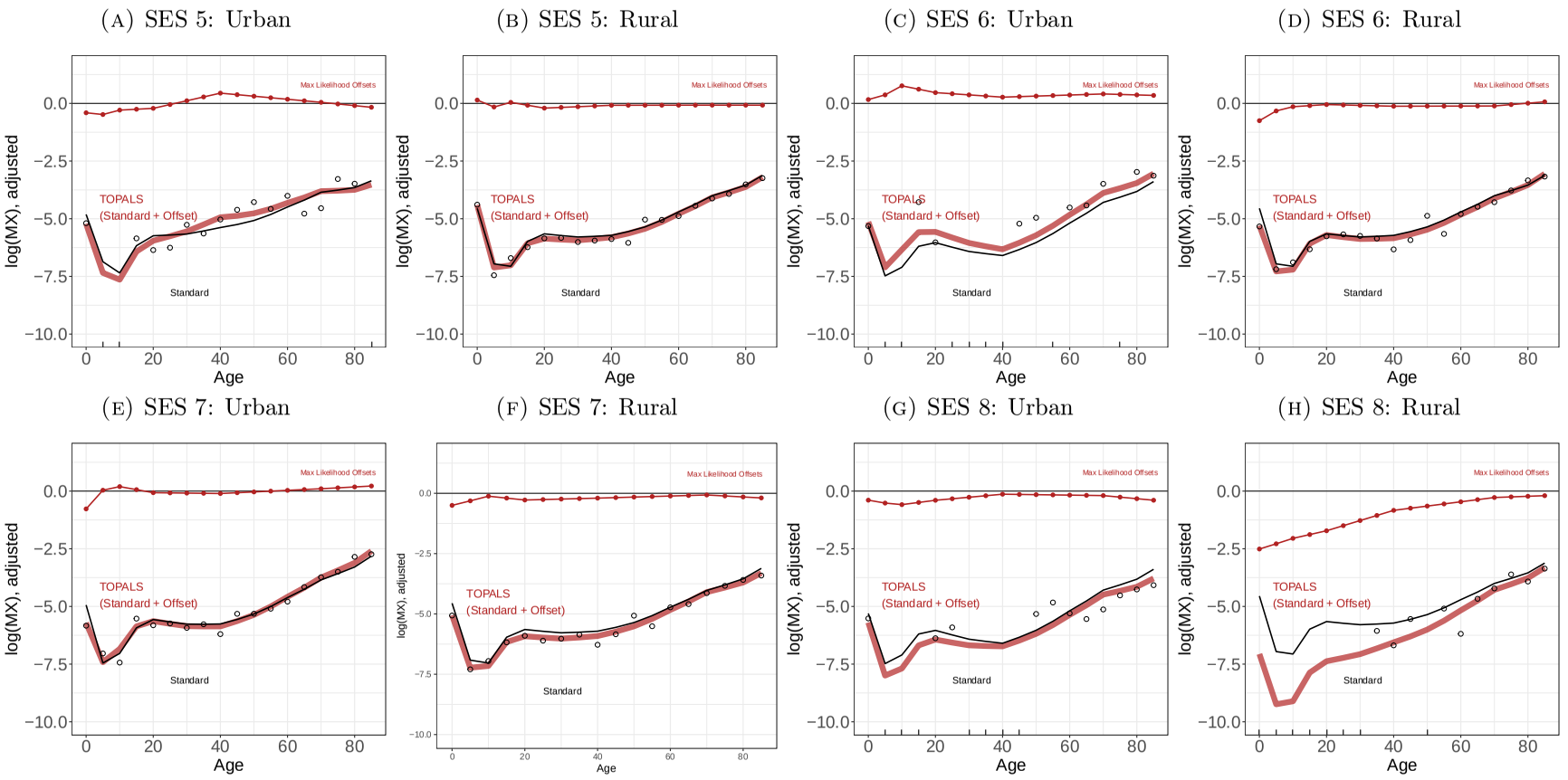
Note: Black lines in Panels A and B represent the standard mortality schedule for Colombia in 2018, derived from the WPP 2024 projections. The corresponding red lines show the fitted mortality schedules by age for urban (Panel A) and rural (Panel B) regions, obtained using the TOPALS Poisson regression model. These fitted schedules are derived by adding the standard mortality schedule to the offset values estimated by the model.

Appendix Figure 5: Mortality and Life Expectancy by Socioeconomic Status in Urban and Rural Colombia (Deciles 1-4)



Note: Each panel displays the fitted mortality schedule for urban and rural areas, stratified into deciles of socioeconomic status (SES) (deciles 1 to 4). The standard schedules (black lines) are derived from the previous estimation step, fitted for the overall urban and rural areas, as shown in Appendix Figure 4. The resulting fitted schedules were then used as standards for estimating age-specific mortality schedules using the TOPALS software for each cluster within each stratified area. Socioeconomic status is a 0-6 categorization of residential units with the purpose of charging differential utilities fees.

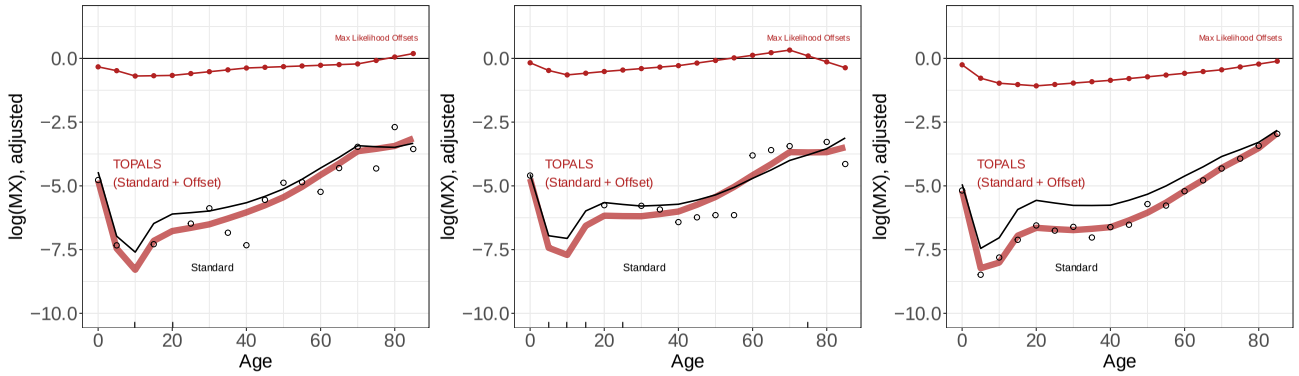
Appendix Figure 6: Mortality and Life Expectancy by Socioeconomic Status in Urban and Rural Colombia (Deciles 5-8)



Note: Each panel displays the fitted mortality schedule for urban and rural areas, stratified into deciles of socioeconomic status (SES) (deciles 5 to 8). The standard schedules (black lines) are derived from the previous estimation step, fitted for the overall urban and rural areas, as shown in Appendix Figure 4. The resulting fitted schedules were then used as standards for estimating age-specific mortality schedules using the TOPALS software for each cluster within each stratified area. Socioeconomic status is a 0-6 categorization of residential units with the purpose of charging differential utilities fees.

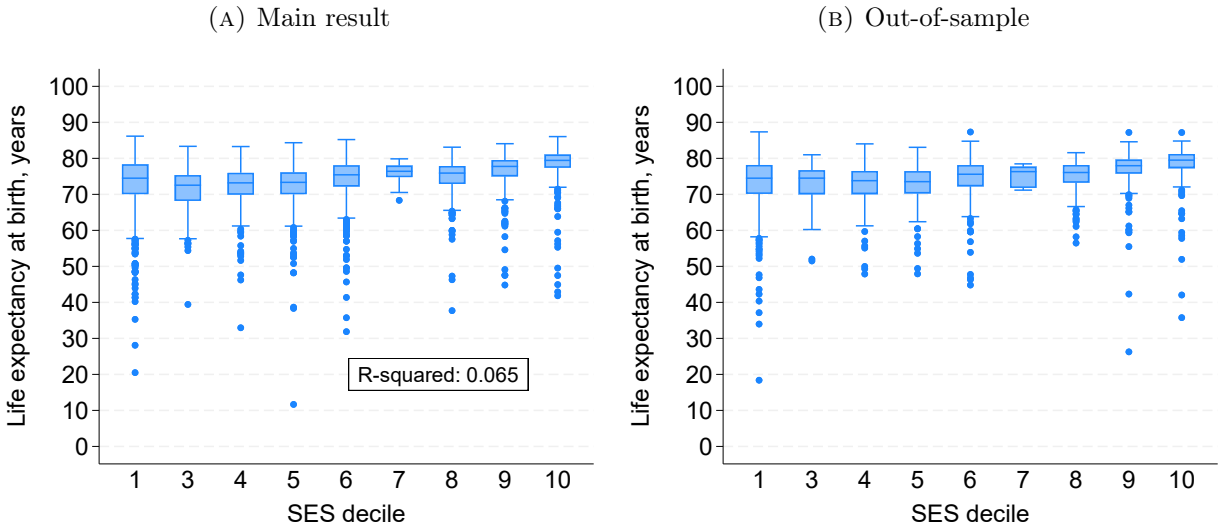
Appendix Figure 7: Mortality and Life Expectancy by Socioeconomic Status in Urban and Rural Colombia (Deciles 9-10)

(A) SES 9: Urban (B) SES 9: Rural (C) SES 10: Urban



Note: Each panel displays the fitted mortality schedule for urban and rural areas, stratified into deciles of socioeconomic status (SES) (deciles 9 to 10). The standard schedules (black lines) are derived from the previous estimation step, fitted for the overall urban and rural areas, as shown in Appendix Figure 4. The resulting fitted schedules were then used as standards for estimating age-specific mortality schedules using the TOPALS software for each cluster within each stratified area. Socioeconomic status is a 0-6 categorization of residential units with the purpose of charging differential utilities fees.

Appendix Figure 8: Out-of-Sample Fit of the TOPALS Model



Note: Panel A reproduces our main result on the distribution of life expectancy at birth by socioeconomic status decile in Colombia. Panel B presents the out of sample distribution of life expectancy at birth by socioeconomic status decile. To construct panel B we first create 200-census block clusters using K-means by minimizing the geographic distance between census blocks within a state. Then, we split each cluster randomly into two groups of equal number of census blocks. We use the first group to estimate the parameters $\alpha_0, ..., \alpha_6$ of the TOPALS model. Then, we use these parameters to predict mortality out of sample in the second group. Panel B presents the resulting distribution of life expectancy in the second group.

Appendix D Vital Statistics vs. Census Data

We rely on census data rather than vital statistics for several reasons. Although Colombian vital statistics have relatively high coverage rates (Karlinsky, 2024), they remain incomplete in certain geographic areas (Urdinola et al., 2017), among disadvantaged subpopulations (Burgard and Chen, 2014), and for some violent causes of death (Urdinola et al., 2017). In addition, the most recent census provides information at the census block level, whereas vital statistics are available only at the municipality level, limiting geographic precision.

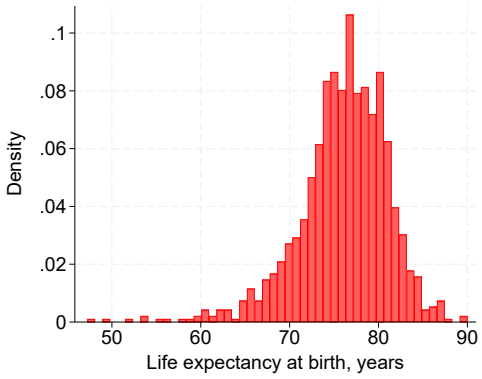
In this appendix, we compare life expectancy estimates based on vital statistics with those derived from municipality-level census data, restricting the latter to certified deaths to ensure comparability. The census reports population counts by 5-year age group for 2018 and deaths by 5-year age group for 2017, including an indicator for whether a death certificate was issued. Of the 244,461 total reported deaths, 183,389 are certified, implying that approximately 25% are uncertified. By contrast, vital statistics record 203,115 non-fetal deaths in 2017.

Appendix Figure 9 shows the distribution of life expectancy at birth across municipalities from the census data. Panel A includes all reported deaths, while Panel B includes only those with a death certificate. We find that excluding uncertified deaths leads to a systematic overestimation of life expectancy. For example, the mean life expectancy at birth is 76 years in Panel A, compared to 79 years in Panel B. The discrepancy is particularly pronounced at the lower tail of the distribution: the 1st percentile life expectancy is 60 years in Panel A but rises to 66 years in Panel B. Panel C plots the geographic distribution of the fraction of total deaths that are uncertified. These types of deaths tend to be more common in municipalities with a history of forced migration such as those in Chocó, Llanos Orientales, and the Amazon.

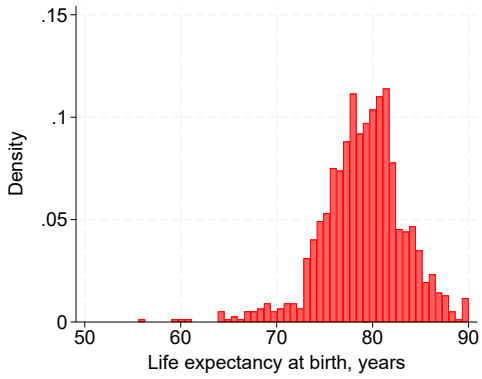
In Appendix Figure 10 we compare the distribution of life expectancy at birth in the main text to the resulting distribution after excluding uncertified deaths from the census data. The distributions are qualitatively similar and exhibit a long left tail. This suggests that the extremely low life expectancies in Colombia are not driven by the greater number of deaths reported in census relative to vital statistics records.

Appendix Figure 9: Distribution of Life Expectancy at Birth with Municipality Level Census Data

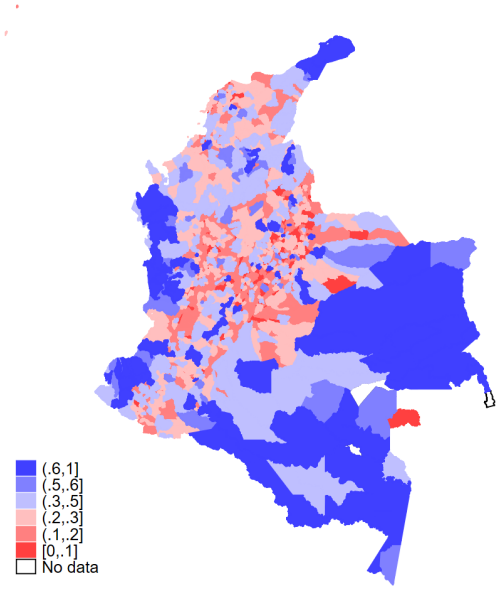
(A) All deaths



(B) Only deaths with death certificates

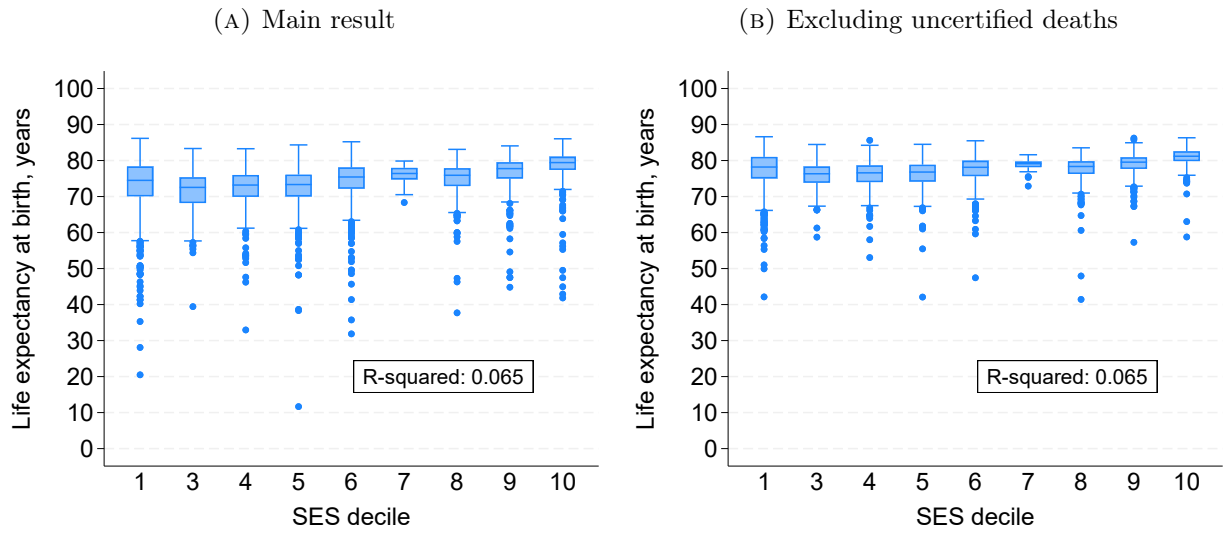


(C) Geographic distribution of uncertified deaths



Note: Figure presents the distribution of life expectancy at birth using municipality level census data from the 2018 Colombian census. Panel A presents the distribution considering all reported deaths. Panel B presents the distribution considering only deaths with a death certificate. Panel C presents the geographic distribution of the fraction of total deaths that are uncertified.

Appendix Figure 10: Life Expectancy Distribution by Socioeconomic Status in Colombia Excluding Uncertified Deaths



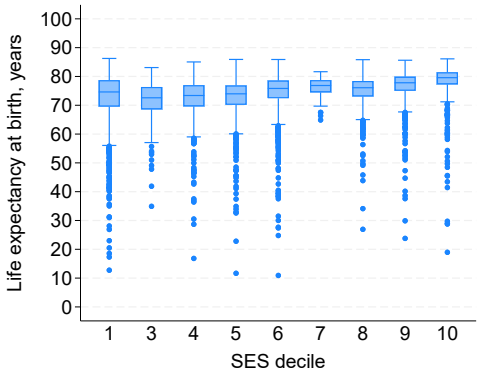
Note: Panel A reproduces our main result on the distribution of life expectancy at birth by socioeconomic status decile in Colombia. Panel B presents the distribution after excluding uncertified deaths from the census data and re-estimating the TOPALS model in the resulting data.

Appendix E Additional Results

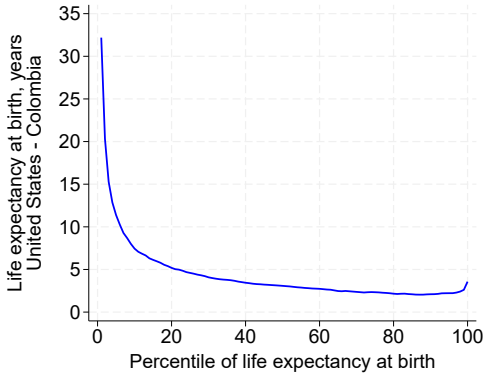
In this Appendix we present additional results and robustness checks. Appendix Figure 11 presents the distribution of life expectancy in Colombia by socioeconomic decile as well as the difference in life expectancy at birth between the U.S. and Colombia using 50-census block clusters and 200-census block clusters. Appendix Figure 12 presents the distribution of life expectancy at birth for several major cities in Colombia and the U.S. Appendix Figure 13 presents the geographic distribution of life expectancy at birth at the country level. Appendix Figure 14 presents the life expectancy at birth distribution and its gradient with respect to SES for conditional on urban/rural status. Appendix Figure 15 presents gradients in life expectancy at birth and SES by other sociodemographic characteristics of the census tracts such as the fraction of resident who identify as Black, the fraction of residents who migrated across departments and states, and the homicide rate.

Appendix Figure 11: Life Expectancy by Socioeconomic Status in Colombia Based on 50- and 200-census block Clusters

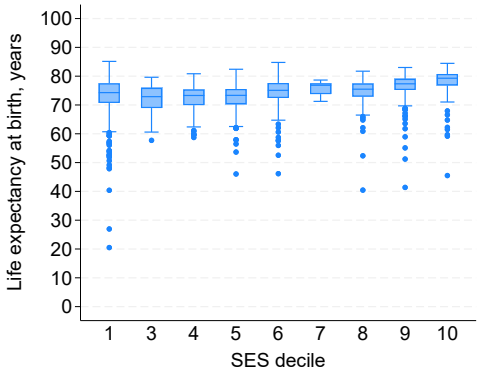
(A) Life expectancy distribution: 50-census block clusters



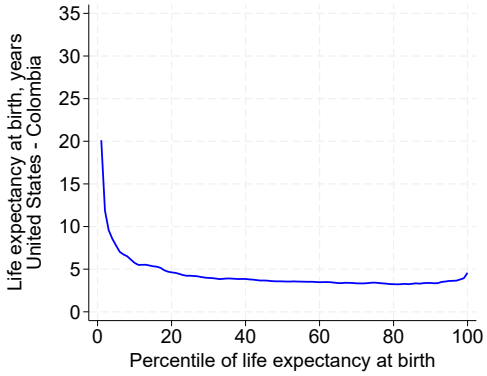
(B) U.S. vs Colombia (50-census block)



(C) Life expectancy distribution: 200-census block clusters

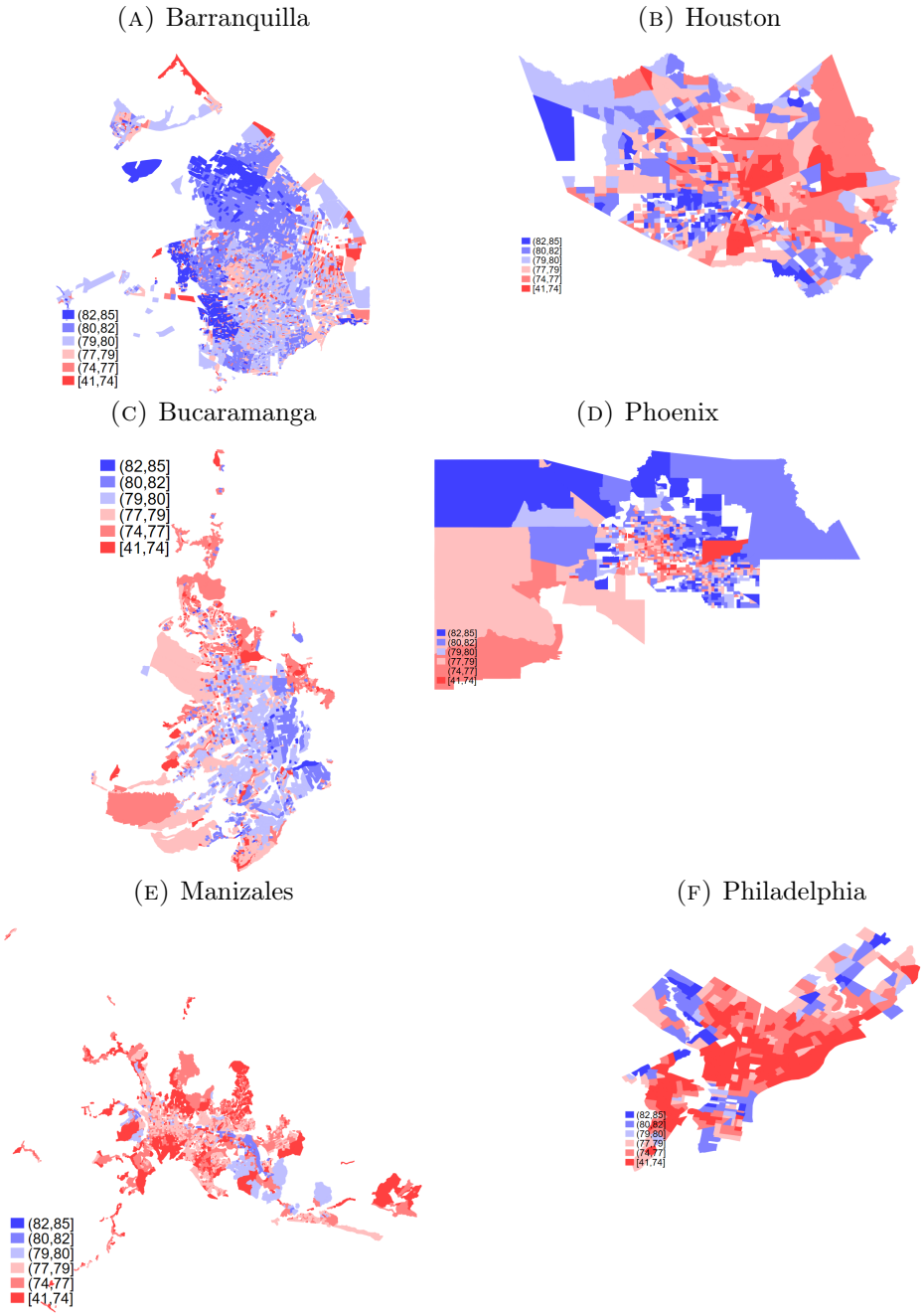


(D) U.S. vs Colombia (200-census block)



Note: Panels A and C present the distribution of life expectancy at birth conditional on each decile of socioeconomic status (SES) in Colombia. An observation is a cluster of census blocks. Clustering is performed using K-means with a minimum number of census blocks per cluster equal to 50 in Panel A and to 200 in Panel C. The center line in each box represents the 50th percentile. The bottom and top of each box represents the 25th and 75th percentile, respectively. The bottom whisker below the box is the 25th percentile minus 1.5 times the interquartile range. The upper whisker above the box is the 75th percentile plus 1.5 times the interquartile range. Socioeconomic status is a 0-6 categorization of residential units with the purpose of charging differential utilities fees. Panels B and D present the difference in life expectancy at both between the U.S. and Colombia by percentiles of the life expectancy distribution. Panel B uses 50-census block clusters in Colombia and Panel D uses 200-census block clusters.

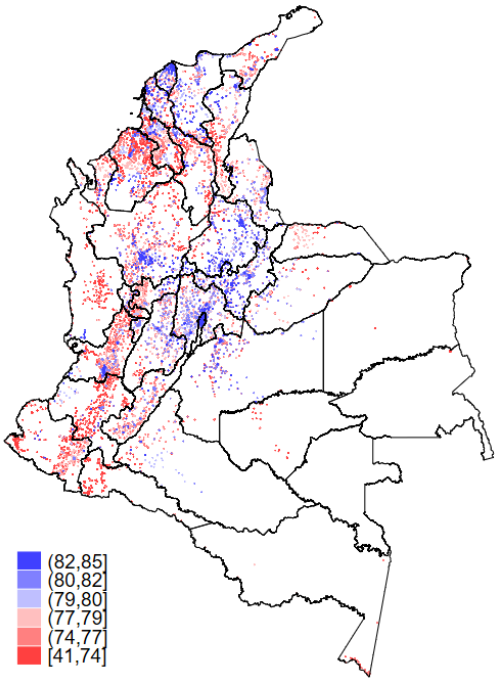
Appendix Figure 12: Geographic Distribution of Life Expectancy at Birth by City in Colombia and the United States



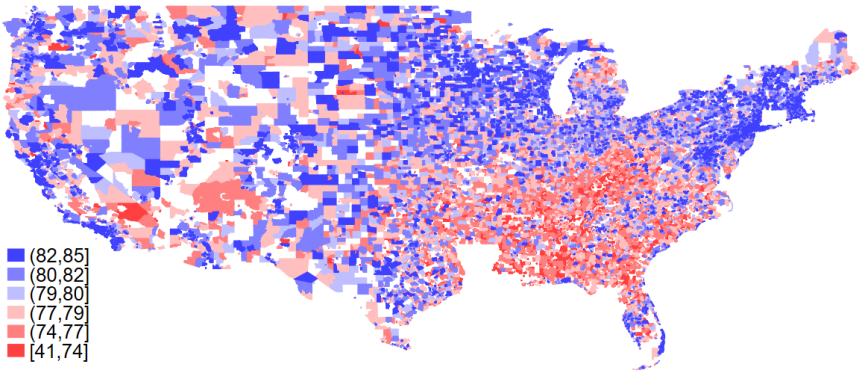
Note: Figure presents the distribution of life expectancy at birth by census tracts within the cities of Barranquilla, Bucaramanga, and Manizales in Colombia and the cities of Houston, Phoenix, and Philadelphia in the United States. Darker blues denote higher values and darker reds denote lower values.

Appendix Figure 13: Geographic Distribution of Life Expectancy at Birth by Country

(A) Colombia



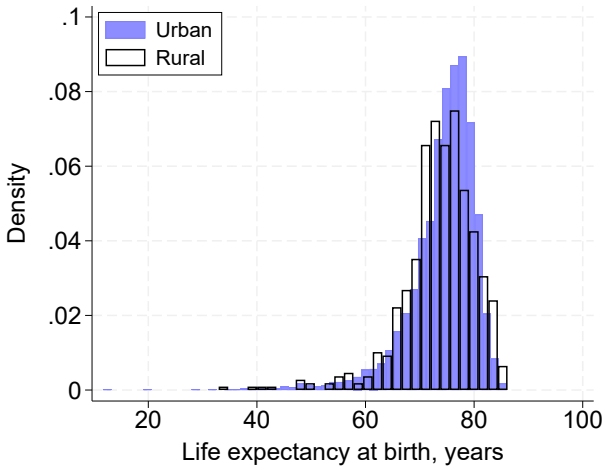
(B) United States



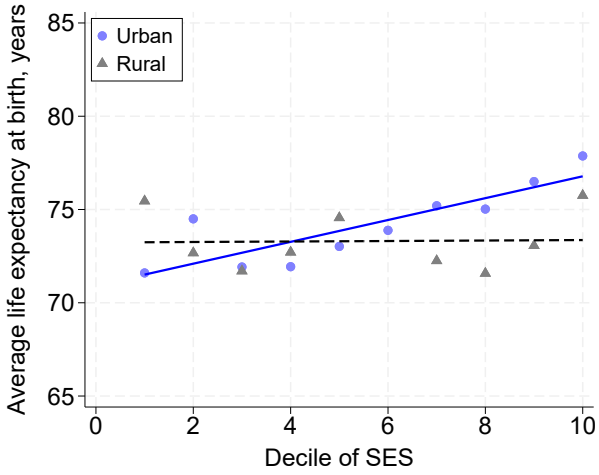
Note: Figure presents the distribution of life expectancy across census tracts in Colombia (Panel A) and in the United States (Panel B). Missing areas in the United States are due to missing life expectancy estimates by census tract from the raw Centers for Disease Control and Prevention data.

Appendix Figure 14: Distribution of Life Expectancy at Birth in Colombia by Urban/Rural Areas

(a) Distribution of life expectancy at birth



(b) Gradient in life expectancy at birth by SES

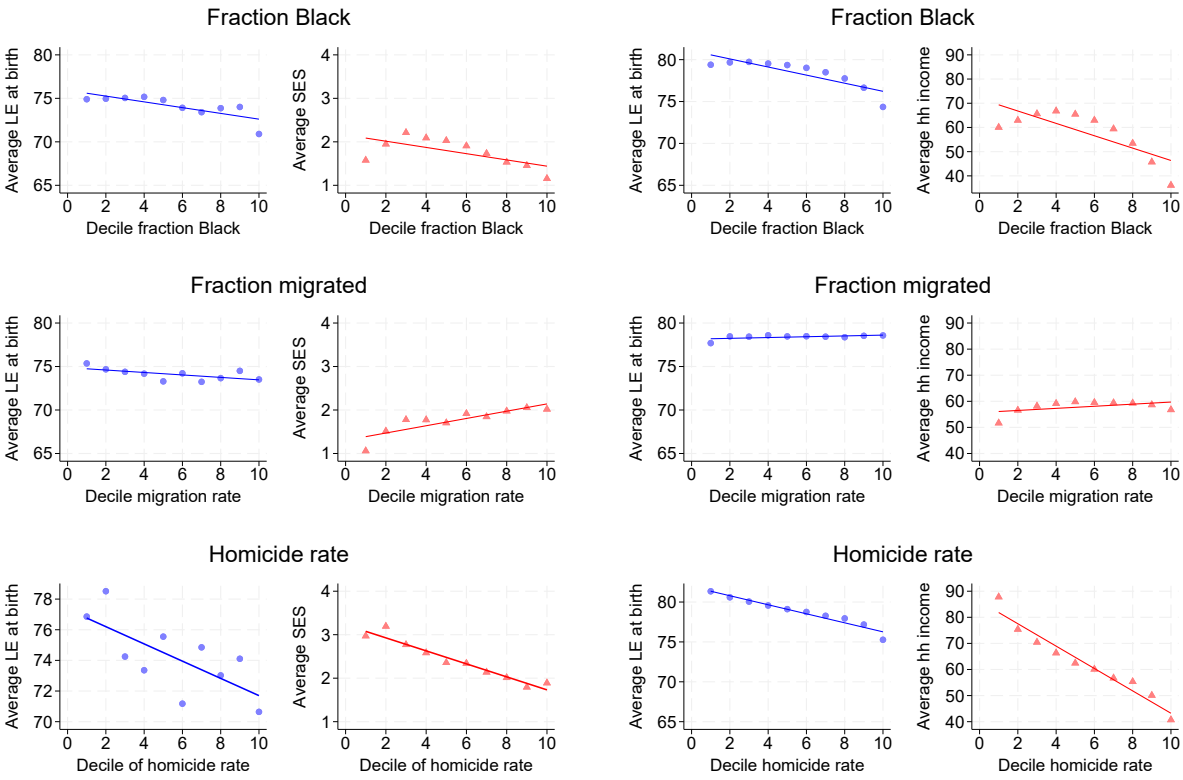


Note: Panel A presents the distribution of life expectancy at birth across urban census tracts in blue and rural census tracts in black. Panel B presents average life expectancy at birth by SES deciles conditional on urban census tracts in blue and conditional on rural census tracts in black.

Appendix Figure 15: Distribution of Life Expectancy at Birth by Deciles of Demographic Characteristics in Colombia and the United States

(A) Gradients for Colombia

(B) Gradients for the United States



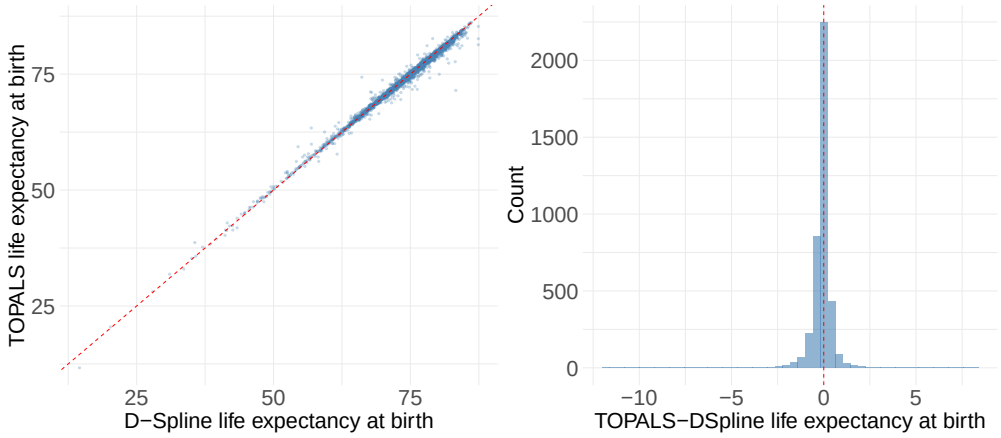
Note: The blue plots present the mean life expectancy at birth by deciles of the fraction Black in the top panel, the fraction who moved across municipalities (in Colombia) or across counties or states (in the U.S.) in the middle panel, and the homicide rate in the bottom panel. The red plots present the mean socioeconomic status (in Colombia) or median household income (in the U.S.) by deciles of the same sociodemographic characteristics. Straight lines in each plot correspond to linear fits. Columns 1 and 2 depict relations for Colombia. Columns 3 and 4 report relations for the U.S. Homicide rates in Colombia correspond to the city of Medellín for which census tract-level data are available.

Appendix F Alternative Models

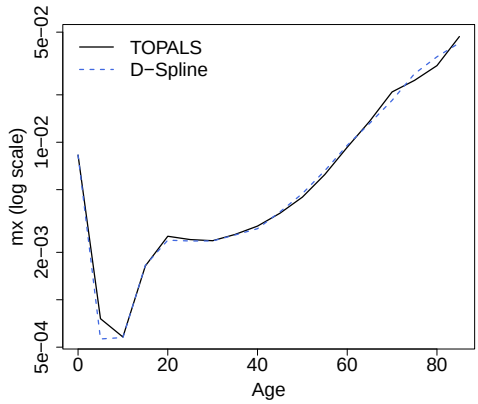
Our main estimates of age-specific mortality rates rely on the TOPALS relational model (de Beer, 2012). In this appendix, we assess the sensitivity of our results to this modeling choice by re-estimating age-specific mortality rates using two alternative approaches. The first is the D-Spline model of Schmertmann (2021), which regularizes mortality schedules using demographic patterns estimated from the Human Mortality Database rather than a single standard schedule, and which we implement following the same nested hierarchical structure as our main TOPALS estimation. The second is the Log-Quadratic model of Wilmoth et al. (2012), which characterizes the age pattern of mortality as a function of child mortality and a single shape parameter, using coefficients calibrated to a large set of national life tables. These two alternatives differ from TOPALS both in functional form and in the source of demographic regularity they impose. Appendix Figures 16 and 17 show that life expectancy estimates are nearly identical across models, with the exception that the Log-Quadratic model yields higher mortality rates at older ages. This discrepancy is likely explained by its calibration to data from high-income countries. We conclude that our findings are robust to the choice of mortality estimation model.

Appendix Figure 16: Comparison of life expectancy estimates TOPALS vs D-Spline

(A) Scatter plot of life expectancy (B) Histogram of life expectancy difference



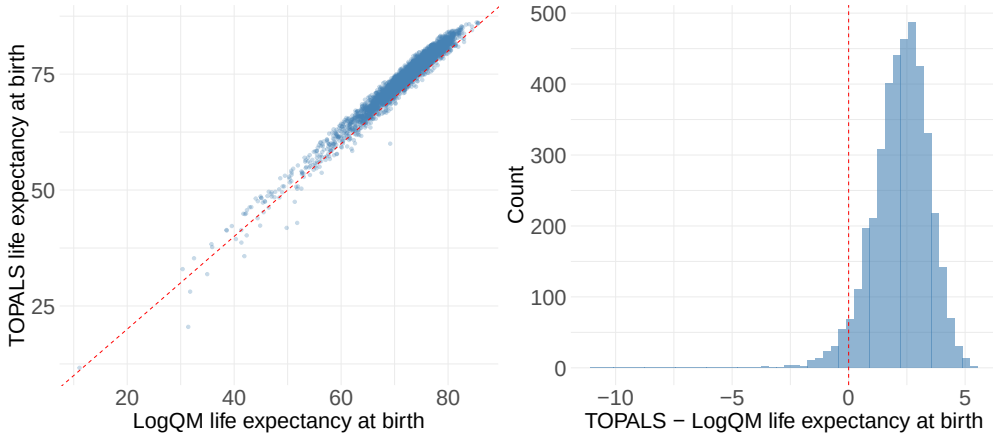
(C) Log mortality curves



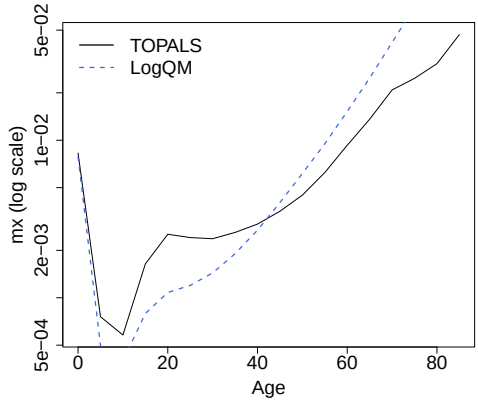
Note: Figure compares estimates from the TOPALS relational model (de Beer, 2012) with those from a D-Spline model (Schmertmann, 2021) estimated using the same hierarchical structure. Panel A presents a scatter plot of life expectancy at birth estimated by TOPALS against the corresponding D-Spline estimates, where each point is a Colombian census tract; the red dashed line is the 45-degree line. Panel B presents the distribution of the difference in estimated life expectancy at birth between the two models (TOPALS minus D-Spline). Panel C presents mean log mortality rates by age from the TOPALS model (black solid line) and the D-Spline model (blue dashed line).

Appendix Figure 17: Comparison of life expectancy estimates TOPALS vs Log-Quadratic model

(A) Scatter plot of life expectancy (B) Histogram of life expectancy difference



(C) Log mortality curves



Note: Figure compares estimates from the TOPALS relational model (de Beer, 2012) with those from a log-quadratic model (Wilmoth et al., 2012). Panel A presents a scatter plot of life expectancy at birth estimated by TOPALS against the corresponding D-Spline estimates, where each point is a Colombian census tract; the red dashed line is the 45-degree line. Panel B presents the distribution of the difference in estimated life expectancy at birth between the two models (TOPALS minus LogQM). Panel C presents mean log mortality rates by age from the TOPALS model (black solid line) and the Log-quadratic model (blue dashed line).

Supplementary References

Bishop, C. M. (2006). *Pattern Recognition and Machine Learning (Information Science and Statistics)*. Springer-Verlag, Berlin, Heidelberg.

Burgard, S. A. and Chen, P. V. (2014). Challenges of health measurement in studies of health disparities. *Social Science & Medicine*, 106:143–150.

Gonzaga, M. R. and Schmertmann, C. P. (2016). Estimating age- and sex-specific mortality rates for small areas with TOPALS regression: an application to Brazil in 2010. *Brazilian Journal of Population Studies*, 33(3):629–652.

Karlinsky, A. (2024). International completeness of death registration. *Demographic Research*, 50:1151–1170.

Preston, S., Heuveline, P., and Guillot, M. (2000). *Demography: Measuring and Modeling Population Processes*. Demography: Measuring and Modeling Population Processes. Wiley.

United Nations (2024). World population prospects 2024, online edition. *United Nations, Department of Economic and Social Affairs, Population Division*.

Urdinola, B. P., Torres Avilés, F., and Velasco, J. A. (2017). The homicide atlas in Colombia: contagion and under-registration for small areas. *Cuadernos de Geografía: Revista Colombiana de Geografía*, 26(1):101–118.