Beyond the Income Gradient: Comparing Life Expectancy Gaps in Lower-Income Countries and the United States

Natalia Serna,
Department of Health Policy, Stanford University, Stanford, CA, 94304

Iván Mejía-Guevara,
Research Consultant, Palo Alto, CA, 94304

Paul Rodríguez-Lesmes,
Department of Economics, Rosario University, Bogotá, Colombia

Eran Bendavid,
Department of Medicine, Stanford University, Stanford, CA, 94304

Grant Miller,
Department of Health Policy, Stanford University, Stanford, CA, 94304

Abstract

This paper compares the income gradient in life expectancy at birth between lower-income countries and the United States (US). We focus on Colombia, a country with substantial economic inequality (more than the US) and rich small-area census data comparable to the US. We find that the income gradient in life expectancy is smaller in Colombia than in the US, with a moderate gap between the two countries of 2 to 5 years across all income deciles. However, we also find that Colombians in the bottom 1% of the life expectancy distribution live 25 years less than their American counterparts, and that these extremely low life expectancies are present across all income groups. Our results suggest that factors other than income play a larger role in life expectancy in Colombia than the US, underscoring the importance of broader social determinants in shaping health outcomes in lower-income countries.

Keywords

Mortality, Life Expectancy, Socioeconomic status, Low- and Middle-Income Countries.

Significance

In many high-income countries like the United States, there is a strong relationship between income and life expectancy at birth. However, whether or not income is as strong a predictor of survival in lower-income countries, many of which have a more economic inequality, remains unknown. This paper uses granular data from Colombia (and from the United States), documenting that some of the population experiences extremely low life expectancies across the entire income distribution. Despite comparable average life expectancy in the two countries, Colombians in the first percentile of the life expectancy distribution live 25 years less than Americans–and this difference persists across income levels. These findings suggest factors beyond income may play a greater role in longevity in lower-income countries.

1 Introduction

Within high-income countries, a robust body of research documents a strong relationship between economic inequality and disparities in life expectancy (e.g., Currie and Schwandt, 2016). For example, in the United States (US), 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). Similarly, the corresponding life expectancy gaps at age 40 in Norway are 13.8 years for men and 8.4 years for women (Kinge et al., 2019). This income gradient in life expectancy has drawn considerable attention in public health and policy debates, pointing to critical social determinants of health.

However, the question of whether or not income is as strongly related to life expectancy in low- and middle-income countries (LMICs) remains open. Economic inequality in LMICs is often more pronounced than in high-income settings (Lustig, 2015; Gasparini and Cruces, 2021). For instance, in countries like Brazil and Chile, earnings in the top decile of the income distribution are 30 times greater than in the bottom 50 percent—a disparity three times larger than in the US (Chancel et al., 2022). Yet despite this stark income inequality in many LMICs, there is limited evidence on whether or not they correspond to equally steep disparities in life expectancy.

This paper examines if the income gradient in life expectancy present in the US generalizes to lower-income countries, particularly ones with greater economic inequality. We focus our analysis in Colombia—a country with both a high degree of economic inequality (Davalos et al., 2024) and unusually rich small-area census data.[1] We first compute small-area life expectancy at birth in Colombia and compare with existing small-area estimates in the US, confirming that average life expectancy in the US (78 years) is moderately higher than in Colombia (74 years). Next, we examine the full distribution of life expectancy within each country—strikingly, revealing that in the bottom 1% of the distribution, Americans live 25 years longer than Colombians. In this bottom percentile, life expectancy at birth in Colombia ranges from 20-40 years of life.[2]

We then proceed to investigate how differences between the two countries vary across the income distribution (investigating if the income gradient in life expectancy differs between the two countries). Surprisingly, we find that the difference in average life expectancy at birth between top and bottom income deciles is smaller in Colombia (5 years) than in the US (9 years). Moreover, the very low life expectancies that we observe in Colombia (between 20 and 40 years) are generally distributed evenly across income deciles. Overall, the share of variation in life expectancy at birth explained by income (measured as the R-squared in linear regressions) equals 36% in the US but only 7% in Colombia. These results show that severe health disadvantages in a country like Colombia are not easily explained by income.

Our findings contribute to the growing body of research on determinants of health in LMICs, particularly in Latin America. Some studies have focused on factors such as race and ethnicity (e.g., Costa et al., 2022), educational attainment (e.g., Bilal et al., 2019; Amo-Adjei et al., 2018),

[1] We use census data rather than vital statistics for several reasons. First, although Colombian vital statistics have high coverage rates (Karlinsky, 2024), they are also incomplete, particularly in some geographic areas (Urdinola et al., 2017), among some disadvantaged sub-populations (Burgard and Chen, 2014), and due to some violent causes (Urdinola et al., 2017). Second, recent Colombian census data is available at the census block level, while vital statistics are only available at the municipio level. Third, in our census data, respondents report that about 25% of deaths are not recorded with death certificates (authors' calculations—see Appendix C). In Appendix C, we show that excluding deaths reported in the census, but also reported to lack death certificates, leads to a systematic overestimation of life expectancy.

[2] To avoid missing population counts, we cluster census blocks in Colombia (see Materials and Methods). In Appendix Figure 7, we show that our estimates of life expectancy at birth in the bottom percentiles of the distribution in Colombia are robust to using different cluster sizes.

and health insurance coverage (e.g., Moreno et al., 2021). Others have analyzed health disparities by income using cross-country comparisons (e.g., Ferre, 2016). Building on this prior work, we use rich census block level data from Colombia to document differences in life expectancy relative to the US, highlighting the persistence of exceptionally poor health outcomes in a country like Colombia across the entire income distribution.

2 Results

We use detailed census block data from the 2018 Colombian population census and the 2015 American Community Survey to investigate patterns of age-specific mortality and life expectancy. In our analysis, we cluster census blocks in Colombia by geographic distance within a state to address missing population counts by age group. For ease of exposition, we refer to these census block clusters as simply "census tracts." We also stratify census tracts by a measure of socioeconomic status in Colombia (following the Colombian government's "estrato" classification system) and by household income in the US (see Materials and Methods). There is no corresponding measure of socioeconomic status for US census tracts, and income is not available for all Colombian census tracts. However, Appendix A shows that socioeconomic status and income are highly correlated across census tracts in Colombia where both measures are available (and functionally comparable for our purposes). Hence, for expositional purposes, we refer to socioeconomic status in Colombia as "income" for simplicity.[3]

Figure 1 first shows the geographic distribution in life expectancy at birth and household income for the largest cities in each country, Bogotá and New York City. The geographic patterns in life expectancy generally align closely with patterns of household income within each city: in the left panels, blue census tracts have a life expectancy at birth of 80 years or more, and these closely correspond with traditionally more affluent census tracts in the right panels. In Bogotá, the maps also show that census tracts with extremely low life expectancies (depicted in dark red) are located towards the periphery of the city. Appendix Figures 8 to 11 present similar maps for other major cities in the two countries (Cali, Medellín, Barranquilla, Bucaramanga, and Manizales in Colombia; and Los Angeles, Chicago, Houston, Phoenix, and Philadelphia in the US).

We next examine the distribution of life expectancy within each country. Figure 2, Panel A presents the distribution of life expectancy at birth across census tracts in Colombia (red) and in the US (white). The mass of the life expectancy distribution in Colombia generally falls to the left of the mass for the US. The distribution for Colombia also has a much longer left tail, with census tracts in which life expectancy at birth falls below 60 years. Panel B then shows life expectancy at birth by percentile of the distribution in each country. Overall, the profile of life expectancy across percentiles is much flatter in the US than in Colombia: the within-country gap between top and bottom percentiles of the life expectancy distribution is 20 years in the US and 44 years in Colombia.

To show differences between Colombia and the US across percentiles more clearly, Panel C plots the difference in life expectancy at birth between the two countries by percentile of the distribution. For the top 80% of the life expectancy distribution, the gap between the two countries is 5 years or less. However, the gap grows rapidly in the left tail, rising to a difference of 25 years in the bottom percentile.

Finally, we investigate how life expectancy at birth varies by household income in Colombia

[3] Both socioeconomic status and income are available for Bogotá, which represents about 20% of the Colombian population and 10% of all census blocks in Colombia. Appendix A shows that focusing on Bogotá, we find comparable results using both socioeconomic status and income.

and in the US—and differentially so between the two countries. First, Figure 3 shows average life expectancy at birth in each country by percentile of household income. We find that the income gradient is moderately steeper in the US than in Colombia. The slope difference reported in the figure suggests that increasing the position of the census tract in the income distribution by 1 percentile is associated with 0.026 more years of life in the US than in Colombia.

Then, Figure 4 Panels A and B show the distribution of life expectancy at birth in each country by household income decile—and also within each decile. The center horizontal line in each box depicts median life expectancy in that decile, and the top and bottom of each box represent the 75th and 25th percentiles, respectively. The top whisker shows the 75th percentile plus 1.5 times the interquartile range, and the bottom whisker shows the 25th percentile of the life expectancy distribution minus 1.5 times the interquartile range. Finally, and importantly, individual dots represent census tracts with life expectancies beyond the range of the whiskers.

Panel A shows that median life expectancy at birth in Colombia is weakly increasing in household income, and that this modest gradient is driven mainly by census tracts in the lower range of each decile (in particular, census tracts between the bottom whisker and the 25th percentile). Strikingly, even conditional on income decile, the distribution of life expectancy at birth in Colombia still has a long left tail—there are census tracts with very low life expectancies present in all income deciles. In contrast, Panel B shows very different patterns of life expectancy at birth in the US across the income distribution. The gains in life expectancy across income deciles occur among all census tracts (rather than being limited to less healthy census tracts, as in Colombia). Moreover, within each income decile, there is relatively less variation in life expectancy at birth in the US than in Colombia. Formalizing some of these variance differences between Colombia and the US, Panels A and B also report how much variation in life expectancy at birth is explained by household income in each country (summarized by R-squared values from linear regressions of life expectancy at birth on household income, estimated separately for each country). Consistent with the idea that income is a relatively stronger determinant of life expectancy in the US, we find that income explains 36% of the variation in the US, but only 7% of the variation in life expectancy in Colombia—nearly a 5-fold difference.

Panel C shows the difference in life expectancy at birth between US and Colombian census tracts, separately for the top and bottom income deciles in each country and across percentiles of the life expectancy distribution. The figure highlights that life expectancy differences between the two countries are very large at the bottom of the life expectancy distribution, regardless of household income. In the bottom percentile of the life expectancy distribution, Americans in the bottom income decile live 30 years longer, and in the top income decile, this difference is nearly the same (Americans live 25 years longer).

Notably, Panel C also shows that the difference between the two countries in life expectancy at birth is smaller at the lower end of the income distribution than at the higher end (i.e., the red line is always below the blue line). Finally, the gap between countries in the bottom income decile even reverses above the 40th percentile of the life expectancy distribution. For example, in the 99th percentile of life expectancy at birth, Colombians live 3 years longer than Americans.

3 Discussion

This paper provides evidence that inequality in life expectancy at birth within LMICs is not explained by income to the same extent as in the United States, despite lower-income countries exhibiting greater economic inequality. Colombia, which ranks second in economic inequality among OECD countries, trailing only Brazil (Davalos et al., 2021), offers a compelling case. Comparing

Colombia with the US, we find a substantial gap in first-percentile life expectancy at birth between the two countries that is mostly constant across income deciles. Within the bottom income decile, Americans live 25 years longer than Colombians and within the top income decile they live 30 years longer. In contrast, when focusing on average life expectancy, we observe only a moderate and stable gap between the two countries equal to 2-5 years across income deciles.

Our results make two important contributions. First, they indicate that a comparison of average life expectancy either across countries or within countries may miss important disparities present in other parts of the life expectancy distribution. Documenting disparities along the distribution of life expectancy is important to identify the types of policies that can more effectively reduce these gaps. For example, our finding that the gap in first-percentile life expectancy between the US and Colombia remains constant across income groups suggests that policies aimed at promoting upward economic mobility, such as income transfers or tax deductions, are unlikely to substantially improve life expectancy among Colombians relative to Americans.

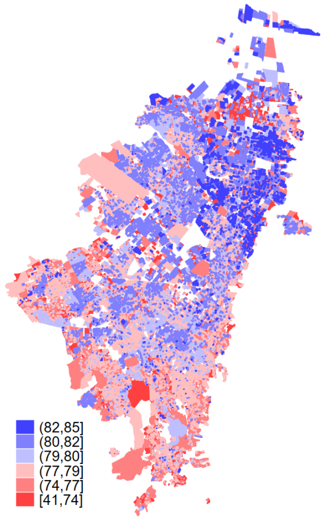
This finding also opens avenues for future research examining determinants of mortality in lower-income countries. The longer left tail and the higher variance in life expectancy at birth in Colombia indicate that health inequalities are driven by more than just income. In Appendix E we explore gradients in life expectancy at birth with respect to other sociodemographic characteristics in Colombia, finding some potential culprits for observed patterns such as violent deaths or forced migration. Thus, policies focused on public safety and social reintegration for forcibly displaced populations may be more effective in improving survival within LMICs.

Second, the use of small-area life expectancy estimates enables us to examine the relationship between income and life expectancy for the entire population, not only for those who earn an income. Existing studies in high-income countries document substantial disparities in life expectancy by income among individuals over age 40, as mortality data can typically be linked to income records only for this group (Chetty et al., 2016; Kinge et al., 2019). However, this does not imply that income influences survival only in adulthood. Although individual-level data improve precision, our comparison of the distribution of life expectancy at birth across census tracts provides a more comprehensive view of how income and other conditions at the census tract influence survival across all age groups (Shanahan et al., 2022).

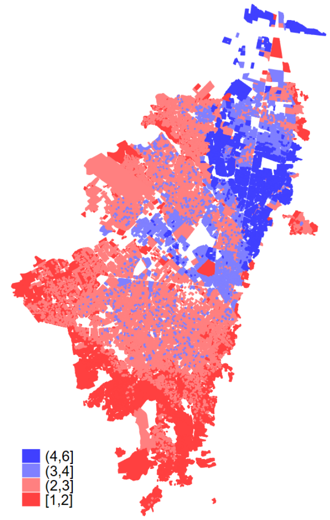
Finally, even though our focus on Colombia and the US is motivated by both the stark difference in economic inequality between the two countries and the availability of small-area mortality records, our results likely extend to other LMICs. Many of these countries share structural characteristics, such as high informality, younger populations, and exposure to conflict, that may generate similar patterns in the relationship between income and life expectancy. Ultimately, improving survival in LMICs will require moving beyond policies that target economic inequality to those that address the broader social and structural determinants of health.

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

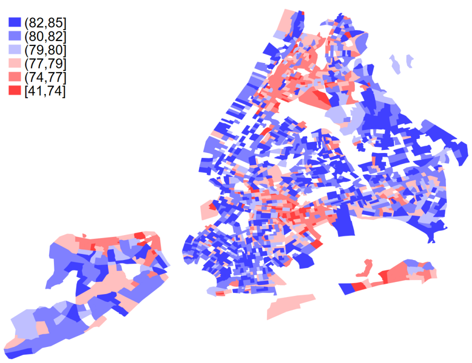
(a) Life expectancy at birth in Bogotá



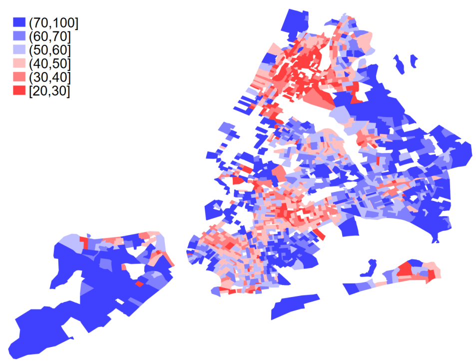
(b) Socioeconomic status in Bogotá



(c) Life expectancy at birth in NYC



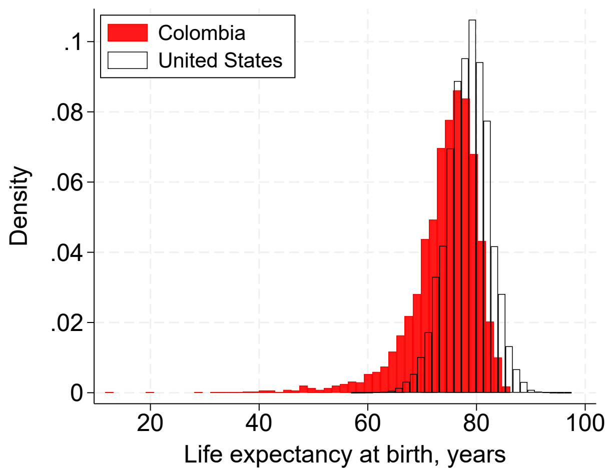
(d) Median household income in NYC



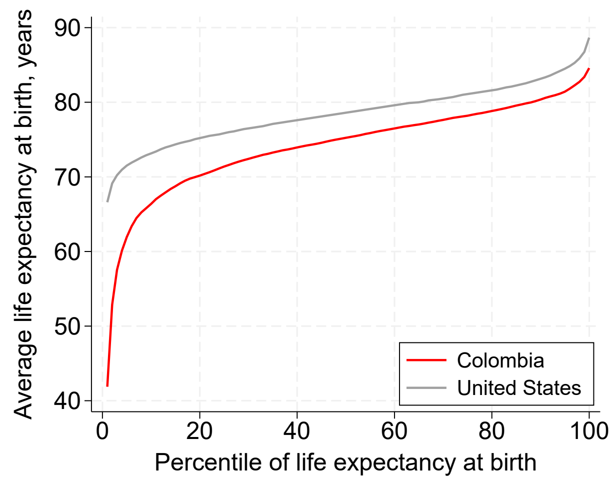
Note: Figure presents the geographic distribution of life expectancy at birth and socioeconomic status in Bogotá in Panels A and B, respectively. Panels C and D present the geographic distributions of life expectancy at birth and median household income in New York City, respectively. Darker blues represent higher values and darker reds represent lower values.

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

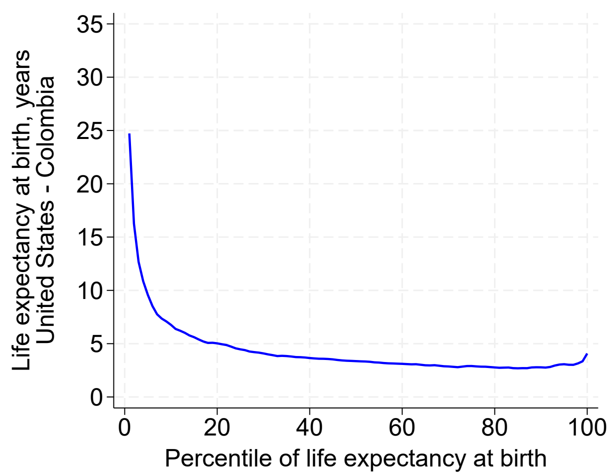
(a) Distribution of life expectancy at birth in Colombia and the United States



(b) Average life expectancy at birth by percentile

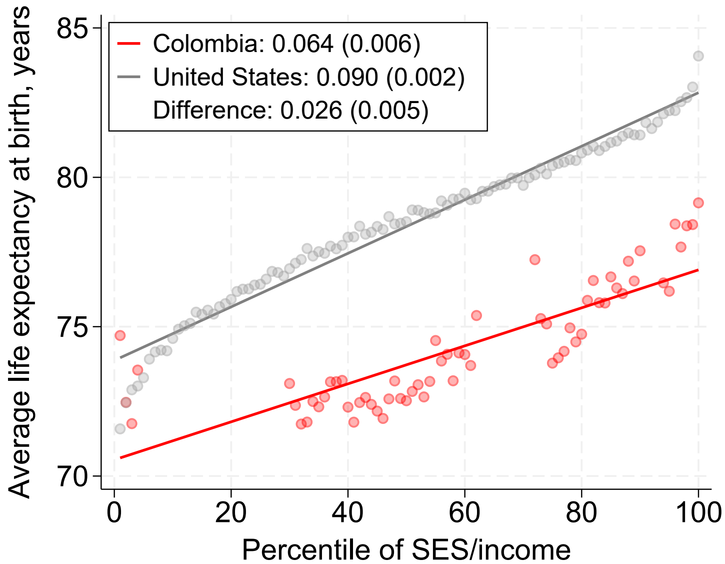


(c) Gap in life expectancy at birth between the United States and Colombia



Note: Panel A presents the distribution of life expectancy at birth across census tracts in Colombia (red) and in the US (white). Panel B presents the average life expectancy at birth by percentile of the distribution in Colombia (red) and the US (gray). Panel C presents the gap in life expectancy at birth between the US and Colombia for census tracts at each percentile of the distribution of life expectancy at birth.

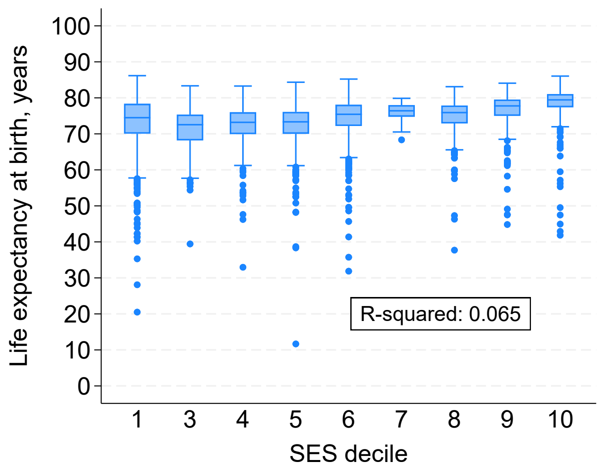
Figure 3: Average life expectancy by income percentile in Colombia and the United States



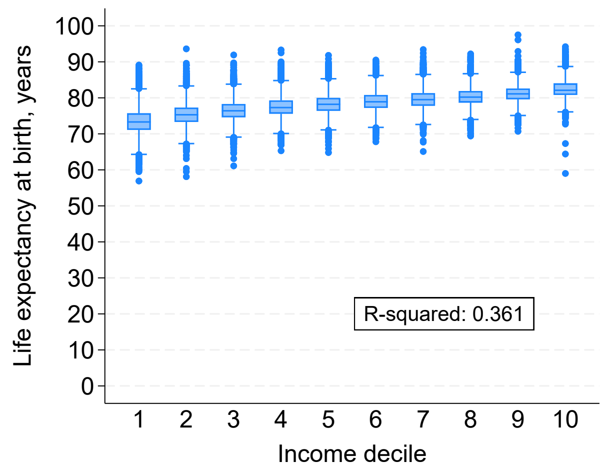
Note: Figure presents average life expectancy at birth in Colombia (red) by percentile of socioeconomic status (SES) and in the US (gray) by percentile of household income. Straight lines correspond to linear fits. The text box in the figure reports the coefficient and standard error in parenthesis of a linear regression of average life expectancy at birth on SES/household income percentile separately by country. The text box also reports the difference in slopes between the two countries. We estimate a pooled linear regression of average life expectancy on SES/income percentile, a dummy variable for the US, and their interaction. The difference in slopes is the coefficient on the interaction.

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

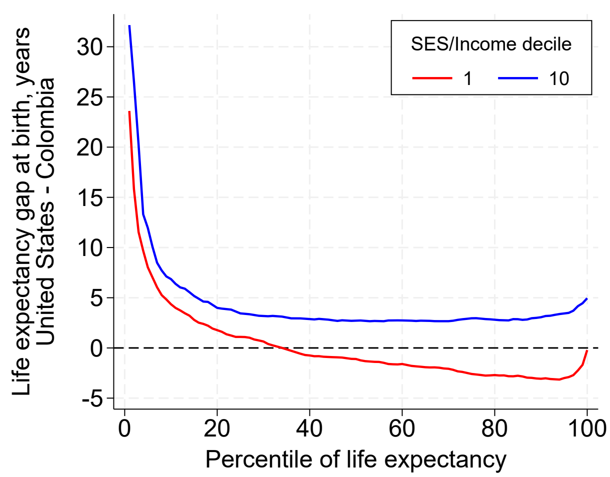
(a) Life expectancy distribution by Socioeconomic Status in Colombia



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



(c) Life expectancy gap at birth between the United States and Colombia by income



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 US, respectively. 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. Each panel reports the R-squared of a linear regression of life expectancy at birth on median household income in the US and on socioeconomic status in Colombia. An observation in these regressions is a census tract. Panel C presents the gap between the US and Colombia at every percentile of the distribution of life expectancy at birth conditional on top and bottom income deciles.

References

Amo-Adjei, J., K. Aduo-Adjei, C. Opoku-Nyamah, and C. Izugbara (2018): "Analysis of Socioeconomic Differences in the Quality of Antenatal Services in Low and Middle-Income Countries (LMICs)," *PloS one*, 13, e0192513.

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.

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

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

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

Costa, J. C., O. J. Mujica, G. Gatica-Domínguez, S. Del Pino, L. Carvajal, A. Sanhueza, S. Caffe, C. G. Victora, and A. J. Barros (2022): "Inequalities in the Health, Nutrition, and Wellbeing of Afrodescendant Women and Children: A Cross-Sectional Analysis of Ten Latin American and Caribbean Countries," *The Lancet Regional Health–Americas*, 15.

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., P. Dudine, J. G. Araujo Lozano, and X. V. Del Carpio (2021): "Building an Equitable Society in Colombia," *Open Knowledge Repository, World Bank.*

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.

Ferre, J. C. (2016): "Economic Inequalities in Latin America at the Base of Adverse Health Indicators," *International Journal of Health Services*, 46, 501–522.

Gasparini, L. and G. Cruces (2021): "The Changing Picture of Inequality in Latin America: Evidence for Three Decades," *UNDP LAC Working Paper Series.*

Gonzaga, M. R. and C. P. Schmertmann (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, 629–652.

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

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.

LUSTIG, N. (2015): "Most Unequal on Earth," *Finance and Development, International Monetary Fund*, 15.

MORENO, X., L. LERA, F. MORENO, AND C. ALBALA (2021): "Socioeconomic Inequalities in Life Expectancy and Disability-Free Life Expectancy among Chilean Older Adults: Evidence from a Longitudinal Study," *BMC geriatrics*, 21, 1–7.

PRESTON, S., P. HEUVELINE, AND M. GUILLOT (2000): *Demography: Measuring and Modeling Population Processes*, Demography: Measuring and Modeling Population Processes, Wiley.

SHANAHAN, K. H., S. SUBRAMANIAN, K. J. BURDICK, M. C. MONUTEAUX, L. K. LEE, AND E. W. FLEEGLER (2022): "Association of neighborhood conditions and resources for children with life expectancy at birth in the US," *JAMA network open*, 5, e2235912–e2235912.

UNITED NATIONS (2024): "World Population Prospects 2024, Online Edition," *United Nations, Department of Economic and Social Affairs, Population Division.*

URDINOLA, B. P., F. TORRES AVILÉS, AND J. A. VELASCO (2017): "The homicide atlas in Colombia: contagion and under-registration for small areas," *Cuadernos de Geografía: Revista Colombiana de Geografía*, 26, 101–118.

4 Materials and Methods

4.1 Data for the United States

For the US, we obtain 2010-2015 estimates of life expectancy at birth by census tract from the National Center for Health Statistics at the Centers for Disease Control and Prevention (CDC). These estimates were constructed by the US Small Area Life Expectancy Estimates Project (USALEEP). We match census tracts in the USALEEP file with the 5-year American Community Survey (ACS) estimates between 2011-2015 for median household income, population who identifies as Black, and population who migrated between states or counties within the last year. We also match census tracts in the USALEEP file to 2023 census tract homicide rates from the CDC.

4.2 Data for Colombia

We use detailed, publicly available census block (*manzana* in Spanish) data from the 2018 Colombian population census provided by the Colombian national statistical agency (*Departamento Administrativo Nacional de Estadística*, or DANE). Census blocks are groupings of built or unbuilt lots delimited by public roads in areas defined as urban across Colombia's 1,121 municipalities (there are 504,996 census blocks in total).

For each census block, we use information on the number of individuals by sex and 5-year age group, the number of deaths in 2017 by sex and 5-year age group, and the number of houses by socioeconomic status (*estrato*). Estrato is a polychotemous 0-6 categorization of individual properties (0 is the lowest, 6 is the highest) made by each municipal government for the purpose of assessing utility fees, utility subsidies, and property taxes. It is an index comprised of housing characteristics including type of housing (house or apartment), square footage, number of bedrooms, and key utilities and conditions including access to roads, access to sidewalks, terrain conditions, and provision of public services (like water and electricity).

We complement the 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.

Finally, we obtain publicly available data from the Police Department in Medellín reporting geo-located homicides from 2009 to 2018. We use these data to produce gradients in life expectancy at birth and socioeconomic status by homicide rate in this city. We match each homicide to the 2018 census block polygon where they occurred. Then, to compute the homicide rate, we calculate the average number of homicides per year from 2009 to 2018 and divide by the total census block population reported in the 2018 census.

4.2.1 Clustering of census blocks

Colombian census blocks are very small geographic units. Approximately 80% of them have zero deaths in 2017 due to low risk and small population with sparse data. These characteristics of the 2018 census data result in unstable event/exposure rates, making it difficult to identify mortality patterns by age. To address these issues, we created clusters of census blocks based on both geographical distance and socioeconomic status within each Colombian state (there are 33 states in the country). We used a K-means clustering approach, an unsupervised machine learning algorithm that groups data points into a predefined number of clusters (Bishop, 2006).

We implemented this cluster analysis using R software, starting with a sample size of 403,916 census blocks after excluding 101,080 blocks that lacked socioeconomic status data (these tend to be census blocks without houses or with other constructions for which the estrato classification does not apply, such as shopping centers, parking lots, parks, or unbuilt blocks). The K-means clustering process involved four main steps: (1) randomly selecting initial cluster centroids, (2) assigning each data point to the nearest centroid based on Euclidean distance, (3) recalculating centroids as the mean of all points within each cluster, and (4) iterating this process 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. This cluster size was chosen after testing various options to strike a balance between preserving variability across census blocks and avoiding excessively large clusters that might resemble municipalities. The algorithm was initialized with 50 random starts and a maximum of 100 iterations to ensure consistency.

To generate clusters that are as homogeneous as possible, we modified the distance function used in the assignment step by incorporating both the socioeconomic status of each census block and the geographic distance between blocks—calculated using their longitude and latitude coordinates—within each Colombian state. As a result, we obtained 4,127 clusters, where the mean number of census blocks per cluster was 96, and the mean number of people per cluster was 8,845 with a standard deviation of 9,912.

For census blocks in Bogotá we additionally implemented the clustering analysis modifying the distance function in the assignment step by incorporating census block average household income and geographic distance between census blocks. This means that 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.

4.2.2 Age-specific mortality and life expectancy across clusters

After identifying clusters of census blocks throughout the country, denoted in the text as "census tracts," we estimated age-specific mortality schedules for each cluster, drawing from empirical patterns observed in a larger geographic area to which each cluster belongs. To fit local mortality rates, we employed a flexible approach combining Poisson regressions with the TOPALS relational model. TOPALS (Tool for Projecting Age-Specific Rates Using Linear Splines), developed by de Beer (2012), uses a linear spline to model the ratios between age-specific death probabilities and a smooth standard age schedule. This algorithm generates complete schedules of age-specific rates by making mathematical adjustments to a prespecified standard schedule.

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: $0–4, 5–9, 10–14, \ldots, 70–74, 75–79, 80–84,$ and 85+. We then 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 schedule, allowing flexibility to accommodate a variety of schedule shapes in small areas. These parameters were estimated by maximizing a penalized Poisson likelihood function for age-specific deaths, conditional on age-specific exposure. This Poisson regression approach has proven to be effective in addressing the 'zeros' for specific age groups where there are no death counts. More details of this approach are available at Gonzaga and Schmertmann (2016).

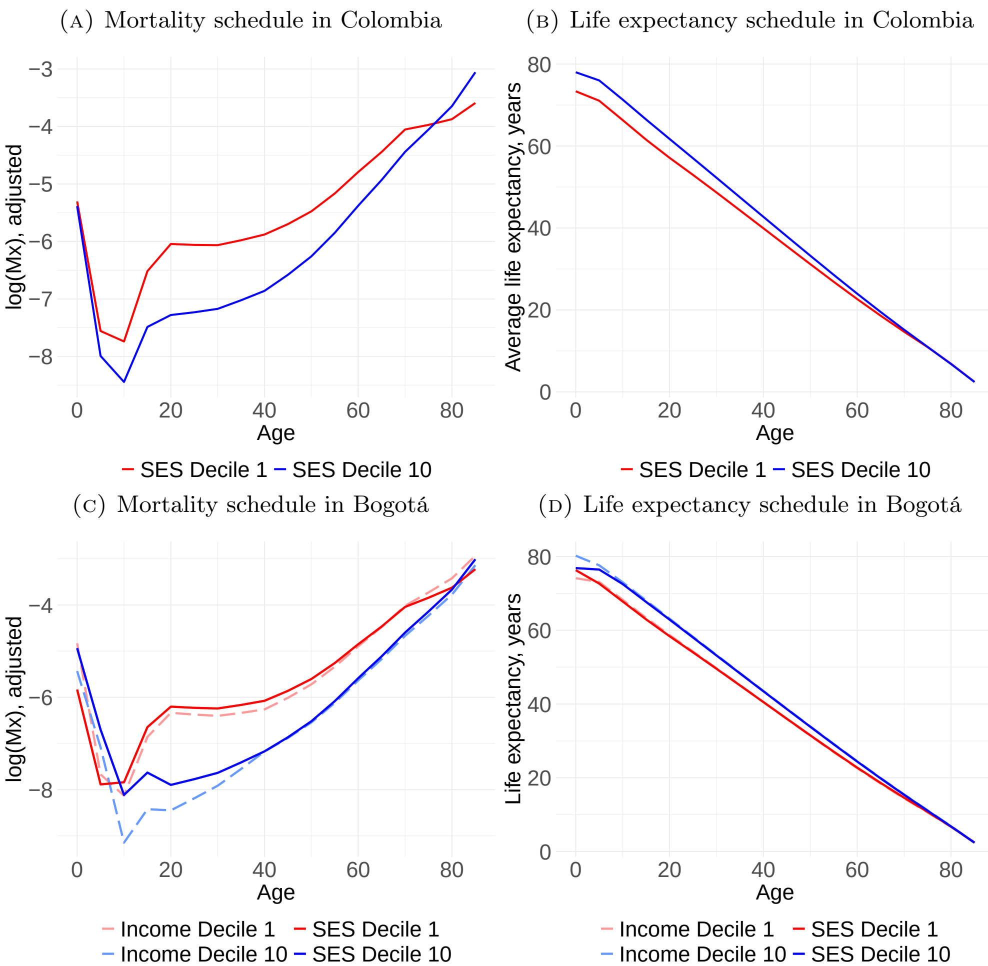
We adopted an approach that considers nested geographic areas to enhance the accuracy of the method. At each level, we used mortality patterns as standard schedules and applied the TOPALS algorithm at each stage to fit mortality schedules for smaller geographic areas, capturing variations in mortality across different regions. The initial standard schedule consisted of age-specific mortality rates on the logarithmic scale for Colombia in 2018, obtained from World Population Prospects 2024 (United Nations, 2024). Using this schedule, we separately fitted age profiles for rural and urban areas (see Appendix Figure 2 in the Supplementary Material). The resulting profiles were then used as standard schedules to fit age profiles for rural and urban regions stratified by Colombian states. We then applied these schedules to estimate mortality rates stratified by rural and urban region, Colombian state, and socioeconomic deciles (see Appendix Figures 3 to 5 in the Supplementary Material). Finally, the resulting profiles were used as standard schedules to fit age-specific mortality schedules for all clusters within these regions throughout the country.

Once we have constructed age profiles for each cluster in the sample, we apply standard demographic methods to construct life tables (Preston et al., 2000), along with the corresponding life expectancy estimates for all age groups in each cluster.

Supplementary Materials

Appendix A Comparison of Socioeconomic Status and Household Income

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



Note: Panel A shows the average 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 average 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 average 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 average 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.

In this Appendix we compare age-specific life expectancy estimates by household income and socioeconomic status in Colombia. We obtain household income from the 2011 Colombian National Quality of Life Survey, which is only reported for census tracts in Bogotá. Our goal is to determine whether disparities in life expectancy at birth by income closely match those by socioeconomic status in Bogotá, and whether life expectancy patterns in Bogotá by socioeconomic status align with those in the entire country.

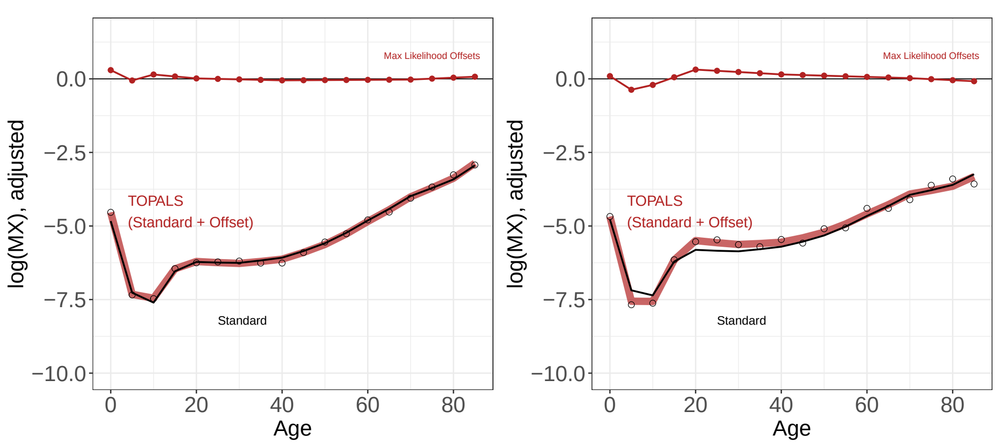
Appendix Figure 1, Panels A and B depict the average log mortality rate and corresponding average life expectancy estimates by 5-year age groups among census tracts in the top and bottom deciles of socioeconomic status in Colombia, respectively. Panels C and D present plots for Bogotá that are otherwise analogous to Panels A and B, stratified by both socioeconomic decile and income decile. In general, the distributions in Bogotá by socioeconomic status are similar to those for the entire country. Moreover, within Bogotá, the distributions by socioeconomic status show broadly similar patterns with respect to differences in life expectancy relative to those reported by income.

Appendix B Clustering Methodology

This Appendix presents results of the fit of our clustering methodology. Appendix Figure 2 reports the fit by urban and rural regions of Colombia. Appendix Figures 3 to 5 break down the model fit by rurality and socioeconomic status decile.

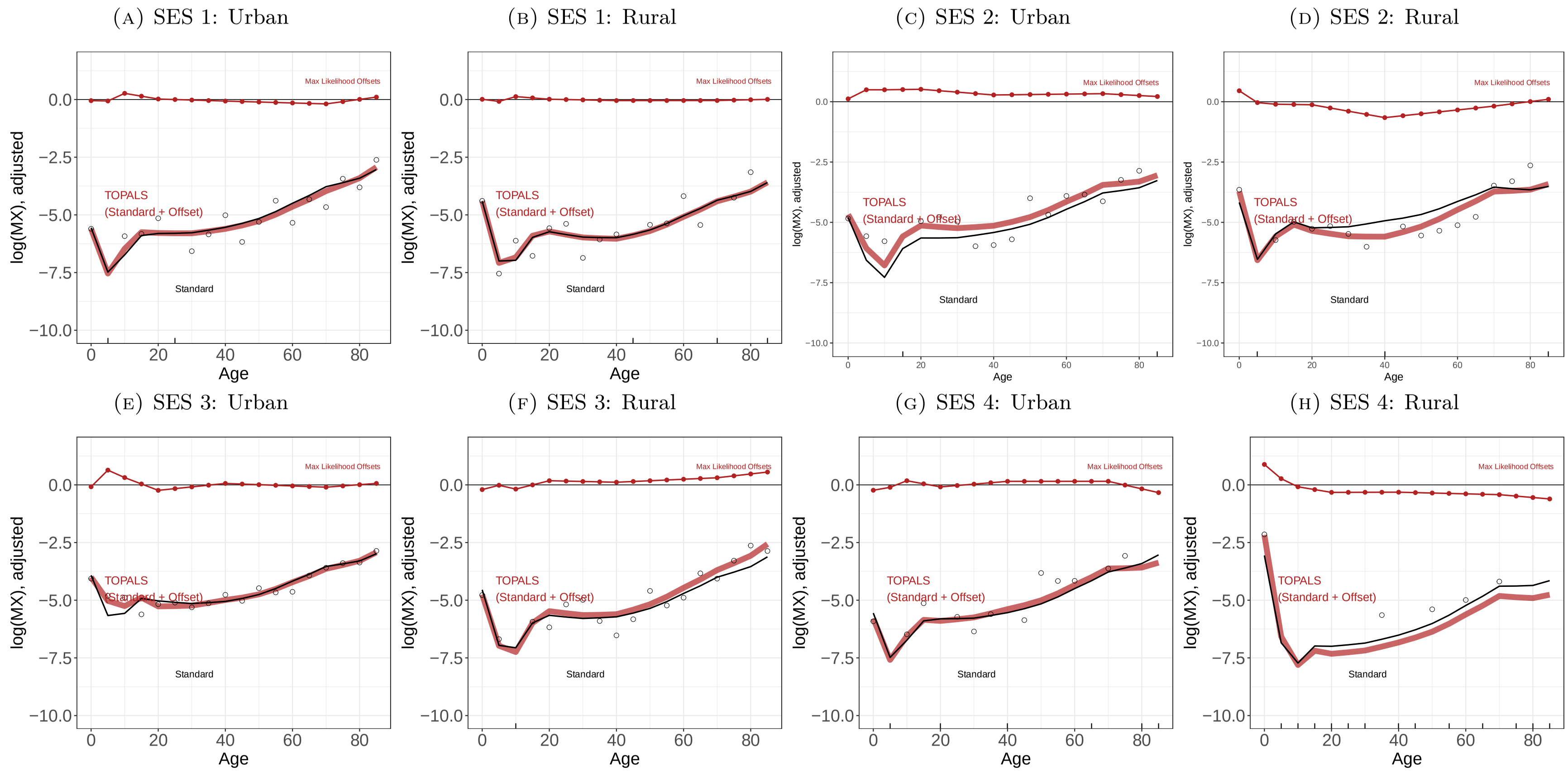
Appendix Figure 2: Standard mortality profiles in urban and rural regions, Colombia

(a) Urban (b) Rural



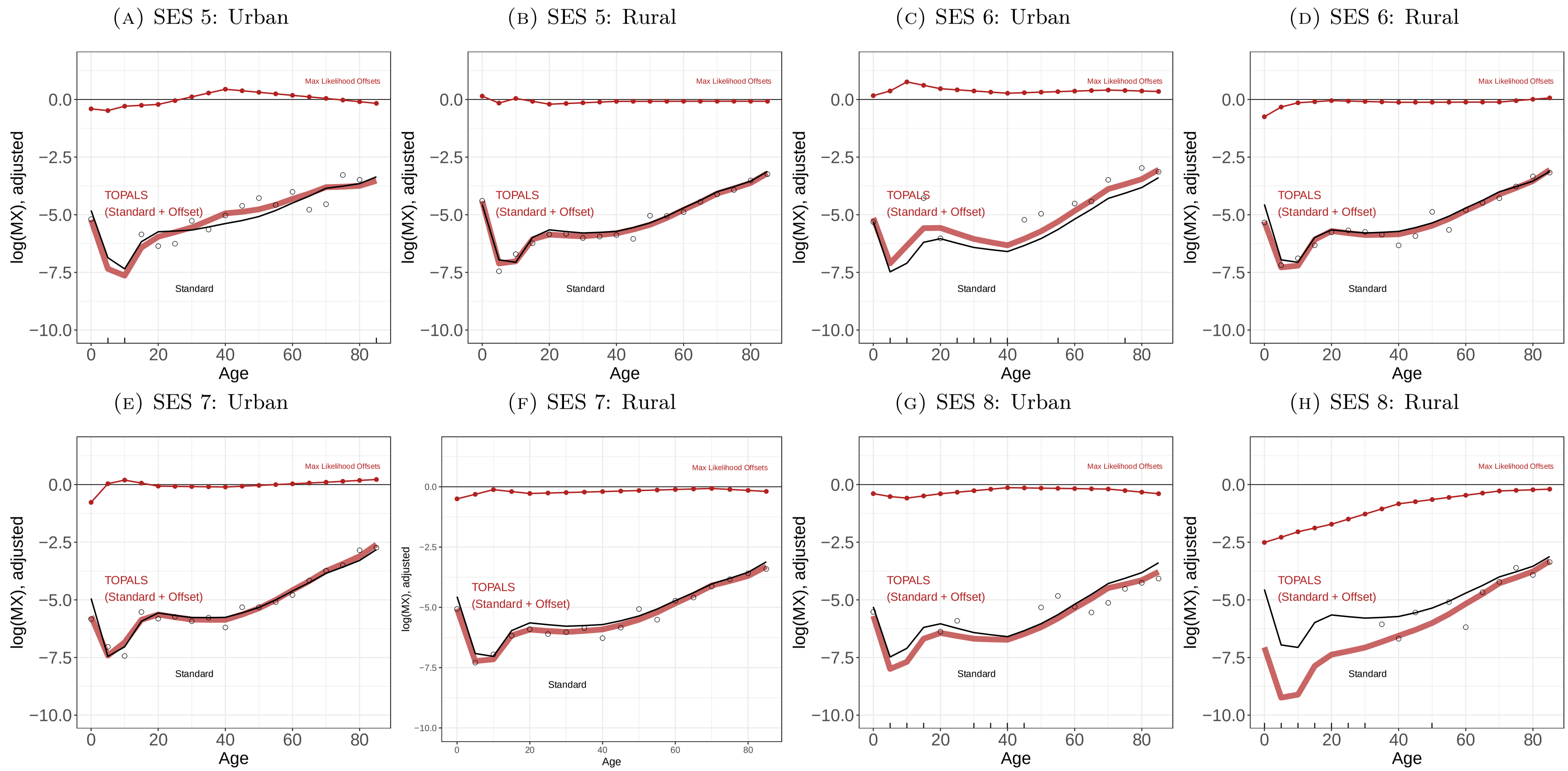
Note: Black lines in Panels A and B represent the standard mortality schedule for Colombia in 2018, derived from the WPP 2024 projections. The corresponding red lines show the fitted mortality schedules by age for urban (Panel A) and rural (Panel B) regions, obtained using the TOPALS Poisson regression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

Appendix Figure 3: Mortality and Life Expectancy by Socioeconomic Status in Urban and Rural Colombia (Deciles 1-4)



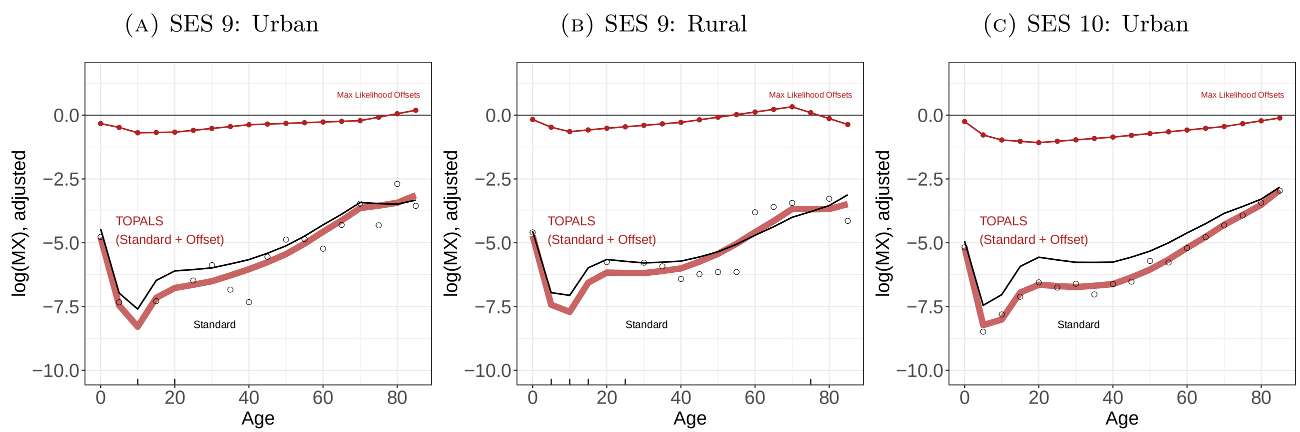
Note: Each panel displays the fitted mortality schedule for urban and rural areas, stratified into deciles of socioeconomic status (SES) (deciles 1 to 4). The standard schedules (black lines) are derived from the previous estimation step, fitted for the overall urban and rural areas, as shown in Appendix Figure 2. 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 4: 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 2. 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 5: Mortality and Life Expectancy by Socioeconomic Status in Urban and Rural Colombia (Deciles 9-10)



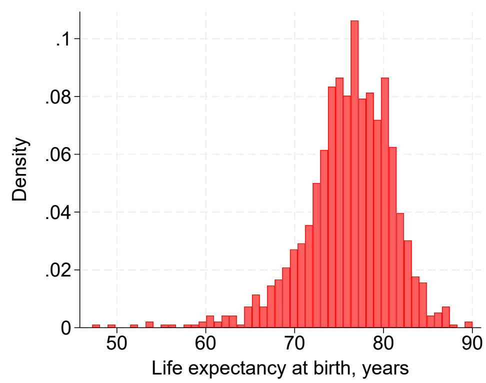
Note: Each panel displays the fitted mortality schedule for urban and rural areas, stratified into deciles of socioeconomic status (SES) (deciles 9 to 10). The standard schedules (black lines) are derived from the previous estimation step, fitted for the overall urban and rural areas, as shown in Appendix Figure 2. 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 C Vital Statistics vs. Census Data

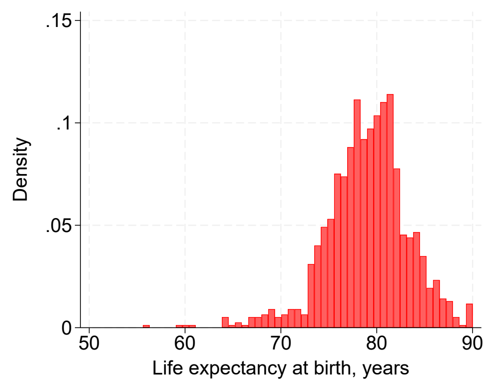
We obtain publicly available 2018 census data at the municipality level from the National Administrative Department of Statistics (DANE). These data include the population by 5-year age group for 2018 and the number of deaths by 5-year age group for 2017. The data also report the number of deaths disaggregated by whether a death certificate was issued. Of the 244,461 total reported deaths, 183,389 have an associated death certificate. Hence, around 25% of deaths are uncertified. In comparison, the number of non-fetal deaths reported in vital statistics for 2017 is 203,115.

Appendix Figure 6: Distribution of Life Expectancy at Birth with Municipality Level Census Data

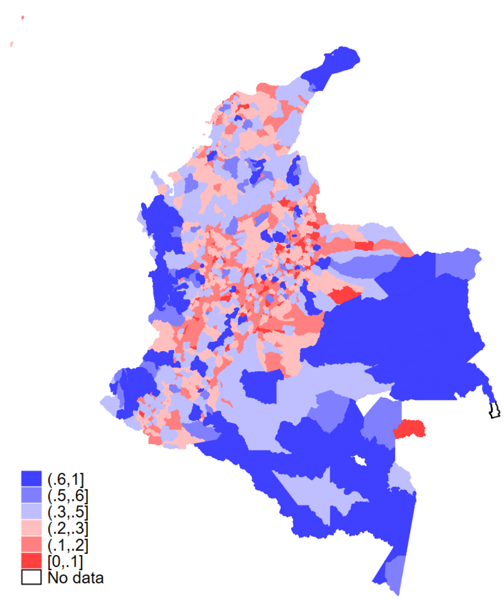
(a) All deaths



(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.

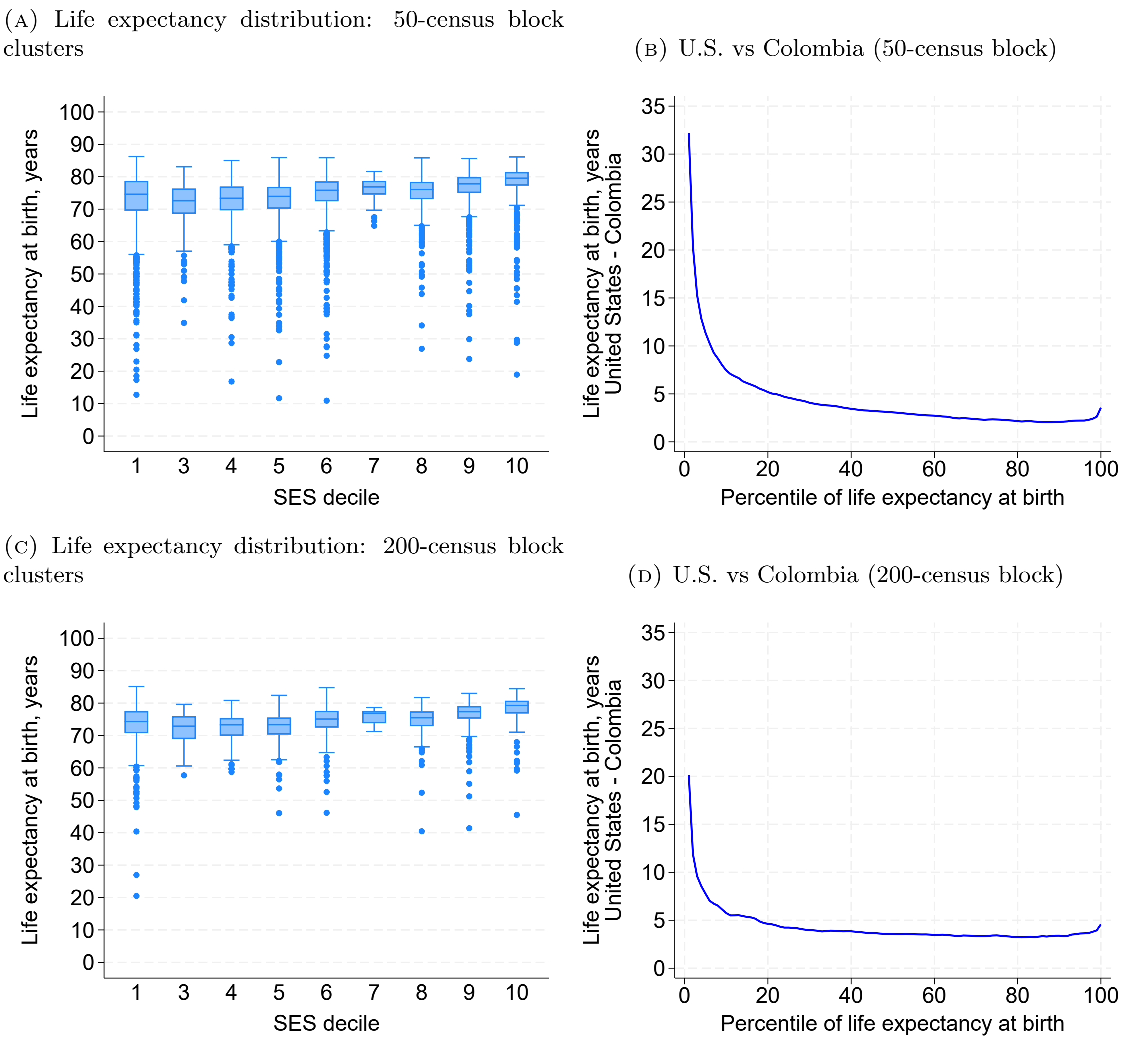
Appendix Figure 6 shows the distribution of life expectancy at birth across municipalities from the census data. Panel A includes all reported deaths, while Panel B includes only those with a death certificate. We find that excluding uncertified deaths leads to a systematic overestimation of life expectancy. For example, the average 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.

Appendix D Additional Results

In this Appendix we present additional results and robustness checks. Appendix Figure 7 presents the distribution of life expectancy in Colombia by socioeconomic decile as well as the difference in life expectancy at birth between the US and Colombia using 50-census block clusters and 200-census block clusters. Appendix Figures 8-11 present the distribution of life expectancy at birth and socioeconomic status/household income for several major cities in Colombia and the US.

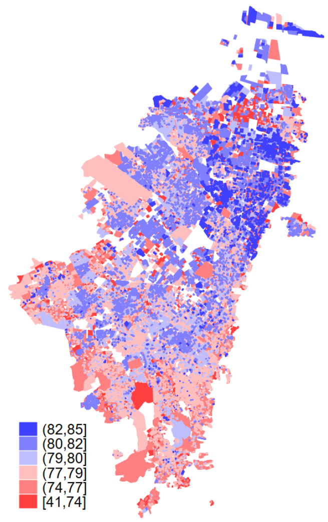
Appendix Figure 7: Life Expectancy by Socioeconomic Status in Colombia Based on 50- and 200-census block Clusters



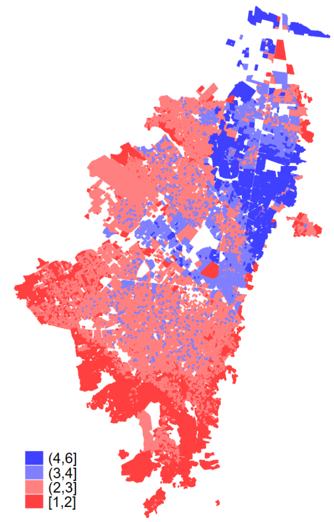
Note: Panels A and C present the distribution of life expectancy at birth conditional on each decile of socioeconomic status (SES) in Colombia. An observation is a cluster of census blocks. Clustering is performed using K-means with a minimum number of census blocks per cluster equal to 50 in Panel A 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 8: Geographic Distribution of Life Expectancy at Birth and Socioeconomic Status in Colombia - Part 1

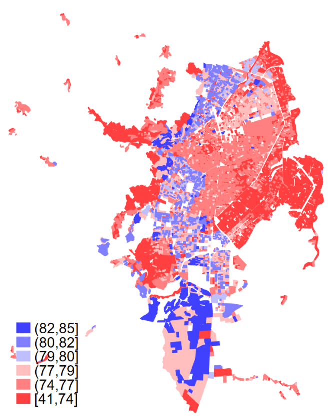
(a) LE Bogotá



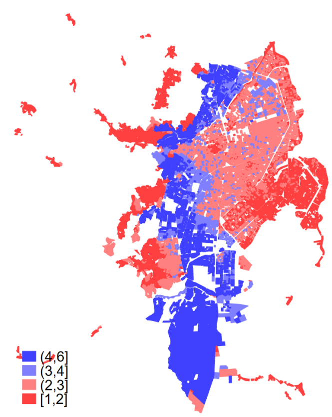
(b) SES Bogotá



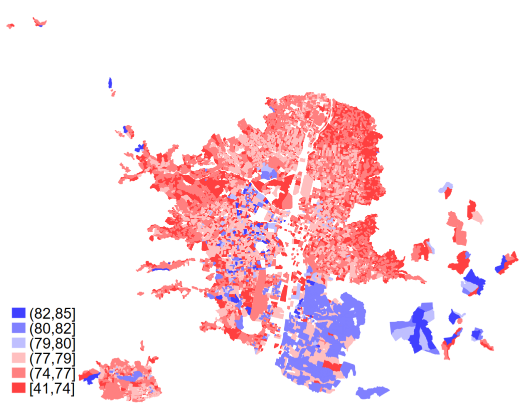
(c) LE Cali



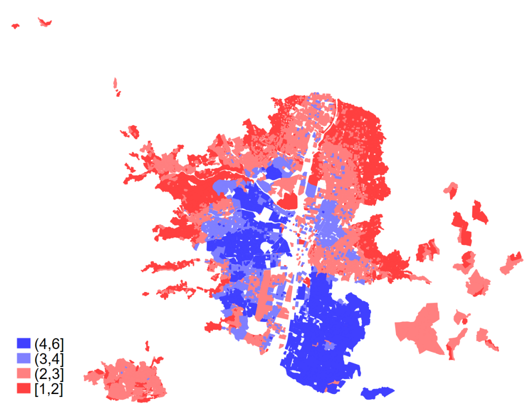
(d) SES Cali



(e) LE Medellín



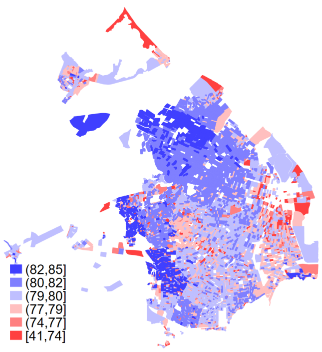
(f) SES Medellín



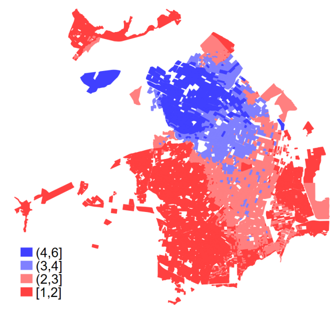
Note: Figure presents the distribution of life expectancy at birth and socioeconomic status (SES) by census tracts within the cities of Bogotá, Cali, and Medellín. Darker blues denote higher values and darker reds denote lower values.

Appendix Figure 9: Geographic Distribution of Life Expectancy at Birth and Socioeconomic Status in Colombia - Part 2

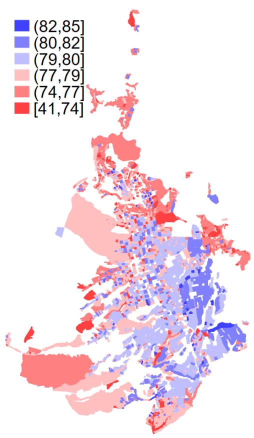
(a) LE Barranquilla



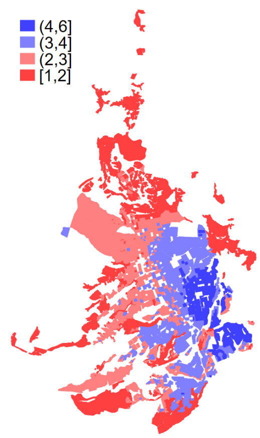
(b) SES Barranquilla



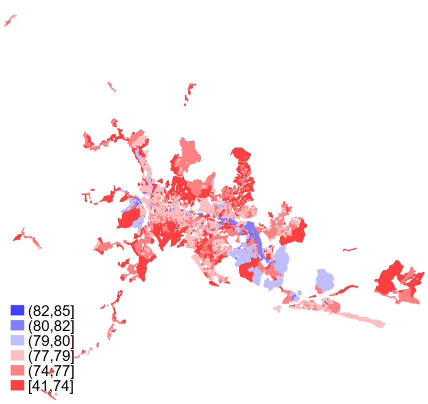
(c) LE Bucaramanga



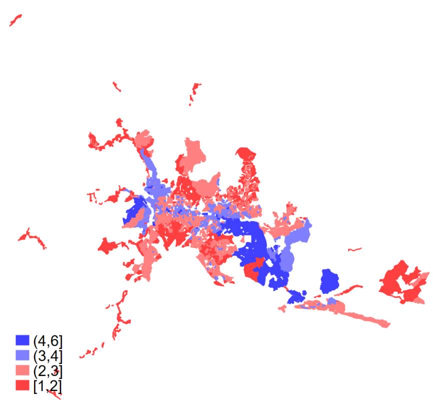
(d) SES Bucaramanga



(e) LE Manizales



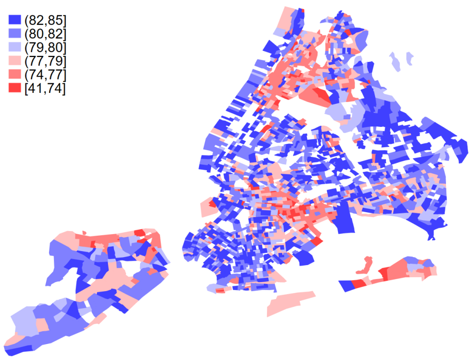
(f) SES Manizales



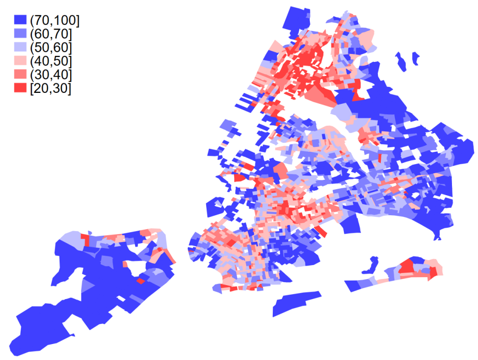
Note: Figure presents the distribution of life expectancy at birth and socioeconomic status (SES) by census tracts within the cities of Barranquilla, Bucaramanga, and Manizales. Darker blues denote higher values and darker reds denote lower values.

Appendix Figure 10: Geographic Distribution of Life Expectancy at Birth and Median Household Income in the US - Part 1

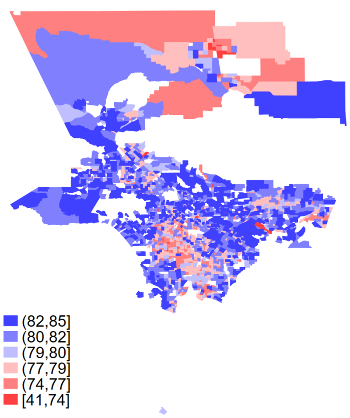
(a) Life expectancy NYC



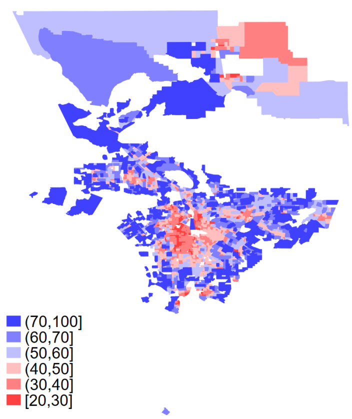
(b) Median household income NYC



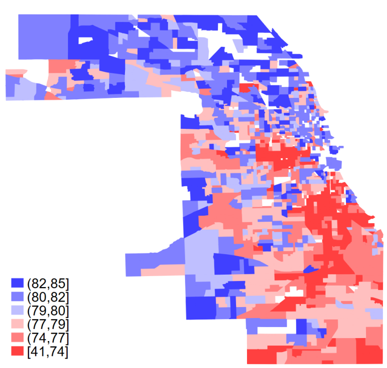
(c) Life expectancy Los Angeles, CA



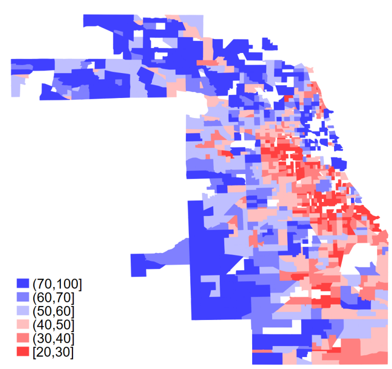
(d) Median household income, Los Angeles, CA



(e) Life expectancy Chicago, IL



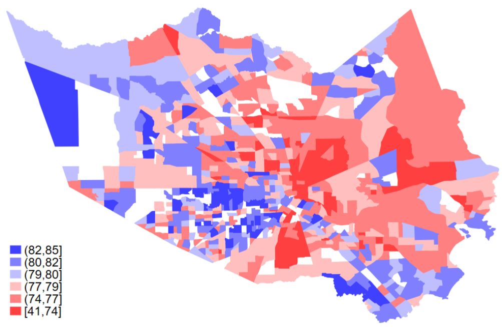
(f) Median household income, Chicago, IL



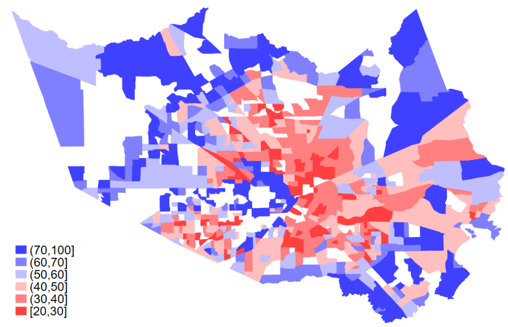
Note: Figure presents the distribution of life expectancy at birth and median household income by census tracts within the cities of New York City, Los Angeles, and Chicago. Darker blues denote higher values and darker reds denote lower values.

Appendix Figure 11: Geographic Distribution of Life Expectancy at Birth and Median Household Income in the US - Part 2

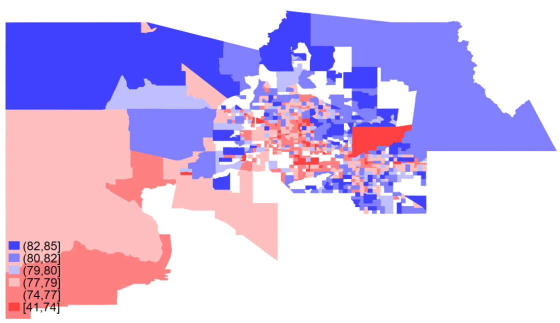
(a) Life expectancy, Houston, TX



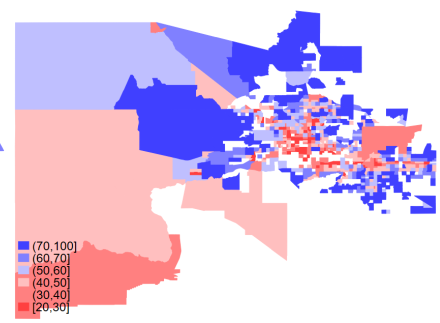
(b) Median household income, Houston, TX



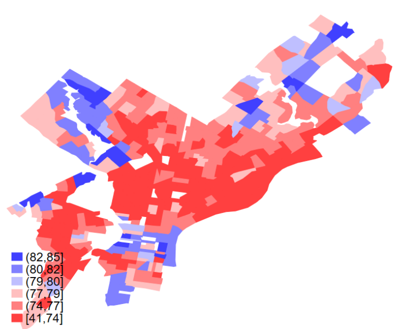
(c) Life expectancy, Phoenix, AZ



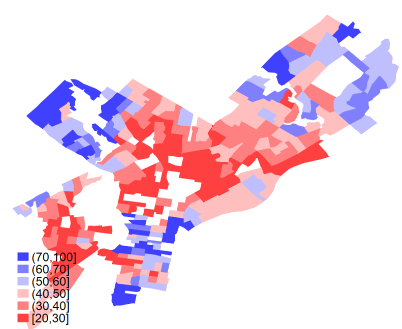
(d) Median household income, Phoenix, AZ



(e) Life expectancy, Philadelphia, PA



(f) Median household income, Philadelphia, PA



Note: Figure presents the distribution of life expectancy at birth and median household income by census tracts within the cities of Houston, Phoenix, and Philadelphia. Darker blues denote higher values and darker reds denote lower values.

Appendix E Additional Life Expectancy Gradients

In this appendix we explore the gradient in life expectancy at birth by additional sociodemographic characteristics beside socioeconomic status. For the analysis in Colombia, we use census tract sociodemographic characteristics reported in the 2018 population census and geo-located homicide records which are available only for the city of Medellín. For the analysis in the US, we use census tract sociodemographic characteristics from the 2011-2015 American Community Survey Estimates and homicide records from the Center of Disease Control and Prevention, accessed through https://data.cdc.gov/Injury-Violence/Mapping-Injury-Overdose-and-Violence-Census-Tract/4day-mt2f/about_data. We consider the following characteristics: fraction of the population who identifies as Black, fraction of the population who migrated across municipalities in Colombia or across counties and states in the US during the last year, and homicide rate.

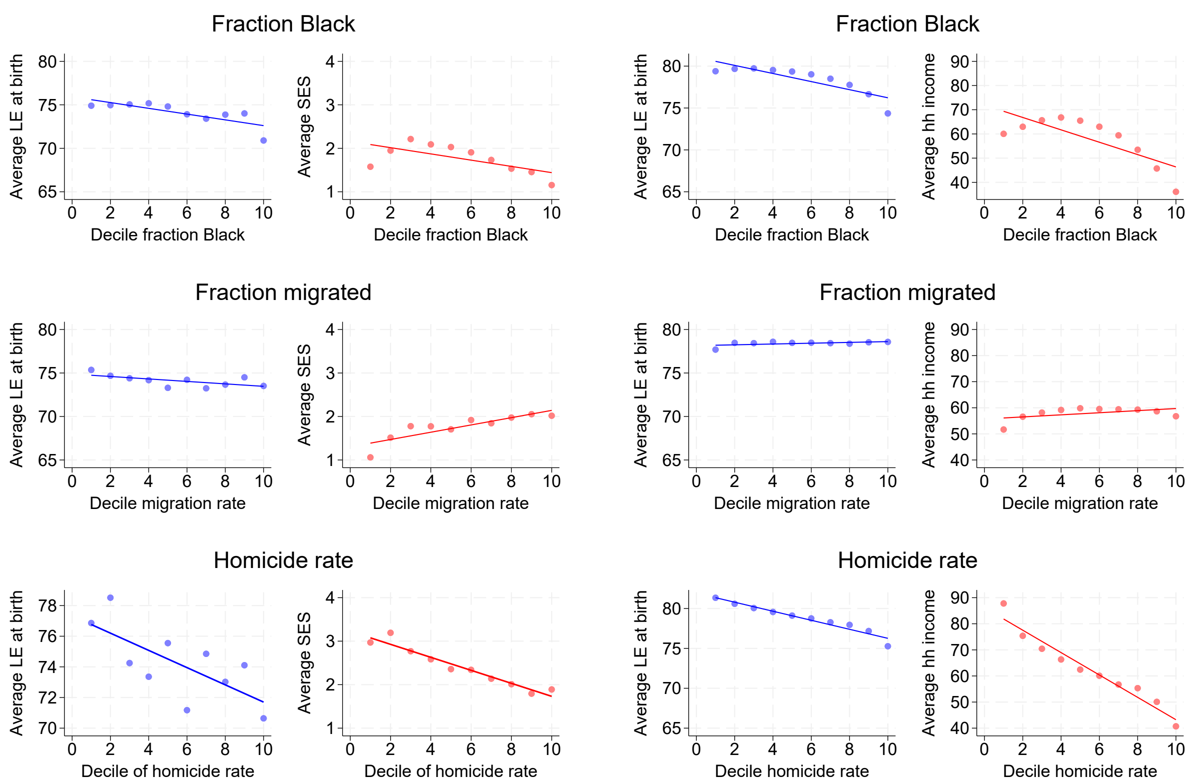
In Appendix Figure 12, the blue plots present the average life expectancy at birth and the red plots present the average socioeconomic status/household income by deciles of each sociodemographic characteristic. We find that while average life expectancy at birth is mostly flat with respect to the fraction of Blacks in Colombia, it declines substantially with this characteristic in the US. US census tracts in the first decile of the fraction of Blacks live 7 years longer than those in the last decile. This suggests race is a stronger determinant of survival in the US than in Colombia. Average life expectancy at birth is mostly flat with respect to the migration rate in the US but has a negative slope in Colombia. Finally, homicide rates are associated with much steeper declines in average life expectancy at birth in Colombia compared to the US. For example, in Colombia, the gap in average life expectancy at birth between the first and last deciles of the homicide rate is nearly 7 years.

We also observe similar gradients between the two countries with respect to the average socioeconomic status/household income. For instance, the red plots show a steeper negative gradient between average household income and the fraction of Blacks in the US compared to Colombia. The difference in average household income between bottom and top deciles of the fraction of Blacks in the US is more than 20 thousand dollars. The homicide rate is also strongly negatively associated with socioeconomic status/household income in each country, with a gap between bottom and top deciles of the homicide rate equal to almost 50 thousand dollars in the US. Instead, the middle panels show nearly zero correlation between household income and the migration rate in the US compared to a steep positive gradient in Colombia.

Appendix Figure 12: 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 average 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 US) in the middle panel, and the homicide rate in the bottom panel. The red plots present the average socioeconomic status (in Colombia) or median household income (in the US) 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 US. Homicide rates in Colombia correspond to the city of Medellín for which census tract-level data are available.