Beyond the Income Gradient: Life Expectancy Gaps between LMICs and the United States

Natalia Serna   Iván Mejía   Paul Rodríguez-Lesmes   Eran Bendavid   Grant Miller

Abstract

This paper examines whether income is as strong a predictor of life expectancy in low- and middle-income countries (LMICs) as it is in the United States. We focus on Colombia, a country with unusually detailed small-area census data. On average, life expectancy in Colombia is only modestly lower than in the U.S., by about 2 to 5 years across all income deciles. However, at the bottom of the life expectancy distribution individuals in the first percentile in Colombia live nearly 25 years less than their U.S. counterparts. These extremely low life expectancies are not limited to the poorest income groups, suggesting income alone is a weak predictor of survival in LMICs. These findings point to the importance of broader social determinants in shaping health outcomes, and underscore the challenges to achieving parity in life expectancy in low- and middle-income countries.

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

1 Introduction

Within high-income countries, a robust body of research has documented a strong relationship between economic inequality and disparities in life expectancy. For example, in the U.S., the difference in expected age at death among those 40 years old between the highest and lowest one percent of the income distribution is 15 years for men and 10 years for women (Chetty et al., 2016). Similarly, in Norway the corresponding gaps are 13.8 years for men and 8.4 years for women (Kinge et al., 2019). The gradient in life expectancy by income has drawn considerable attention in public health and policy debates, pointing to critical social determinants of health.

However, the question of whether 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 Latin America, the wealthiest decile earns 12 times more than the poorest decile compared to 10 times in the U.S. (IDB, 2024; Guzman and Kollar, 2024). And in countries like Brazil and Chile, earnings at the top decile of the income distribution are 30 times greater than those of the bottom 50 percent—a disparity that is three times larger than in the U.S. (Chancel et al., 2022). Yet despite these stark income gaps in LMICs, there is limited evidence on whether they correspond to equally steep disparities in life expectancy as in the U.S.

This paper examines whether economic inequality in LMICs is associated with life expectancy in a manner comparable to that observed in the U.S. 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 start by documenting that life expectancy is higher in the U.S. than in Colombia, particularly towards the lower end of the distribution where Americans live, strikingly, 25 years longer than Colombians.

We proceed to investigate whether differences between the two countries can be explained by economic inequality. Surprisingly, we find that gaps in average life expectancy between top and bottom income deciles are lower in Colombia (5 years) than in the U.S. (9 years). However, in the lower percentiles of the life expectancy at birth distribution, we find evidence of extremely low values in Colombia—unlike in the U.S.—ranging from 20 to 40 years across income deciles, and not attributable to measurement error.

Conditional on the bottom income decile, our estimate for the gap in the first percentile of life expectancy at birth between the U.S. and Colombia is 25 years. Even conditional on the top income decile, the difference between the two countries remains remarkably at 30 years. Overall, the percent of the variation in life expectancy at birth that is explained by income, measured as the R-squared of a linear regression, equals 36 percent in the U.S. compared to only 7 percent in Colombia. This suggests the severe disadvantage in a country like Colombia is not easily explained by income or socioeconomic status.

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

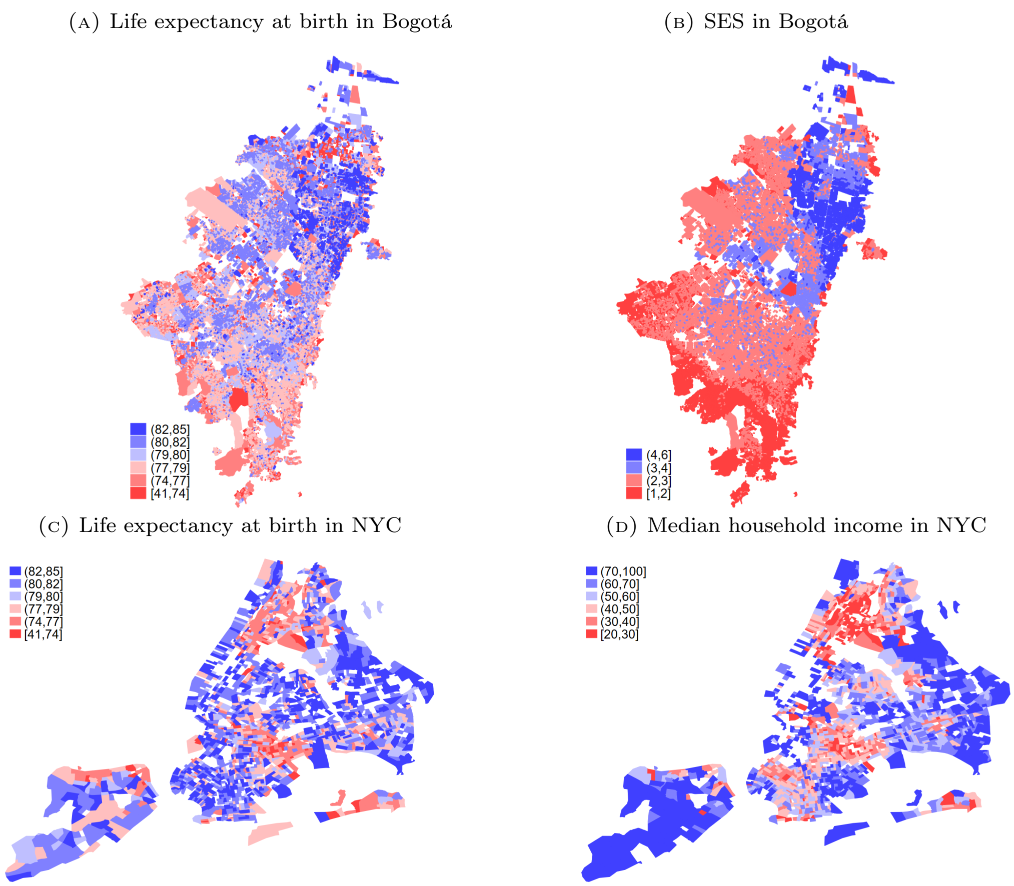
[1] We carefully evaluated the trade-off between using vital statistics records and census data to obtain the number of deaths. While vital statistics are based on death certificates, they may underreport deaths among disadvantaged groups (Burgard and Chen, 2014). In contrast, the census provides death counts disaggregated by whether a death certificate was issued. In Appendix B, we compare municipality-level life expectancy at birth using census data. We find that excluding deaths without a death certificate leads to a systematic overestimation of life expectancy—which is especially problematic when characterizing the tails of the distribution. Uncertified deaths are more common in municipalities with a history of forced migration. Thus, we center our analysis in the census data.

2 Results

We use detailed census tract data from the 2018 Colombian population census and the 2011 Colombian National Quality of Life Survey to investigate patterns of mortality and age-specific life expectancy. In the analysis, we cluster census tracts by geographical distance within a state to address missing population counts by age group. We stratify outcomes by a measure of socioeconomic status following the Colombian government's "estrato" classification and by household income (see Materials and Methods). For ease of exposition, we refer to these census tract clusters as simply "census tracts." We compare our estimates of life expectancy at birth by socioeconomic status in Colombia against corresponding estimates by median household income in the U.S. Life expectancy data for the U.S. comes from the 2015 National Center for Health Statistics and income data from 5-year American Community Survey estimates.

Figure 1 explores the geographic distribution in life expectancy at birth and socioeconomic status/income for the largest cities in each country, Bogotá and New York City. The geographic patterns in life expectancy closely align with patterns of socioeconomic status within city: in the left panels, blue areas representing census tracts with life expectancy at birth over 79 years correspond to those with traditionally high levels of wealth as seen in the right panels. In Bogotá, the maps also show that census tracts with extremely low life expectancies depicted in dark red are located toward the peripheries of the city. Appendix Figures 8 to 11 present maps for other cities in each country.

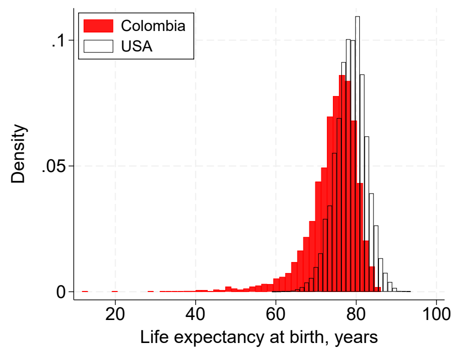
FIGURE 1: Geographic Distribution of Life Expectancy at Birth and SES in Colombia and the U.S.



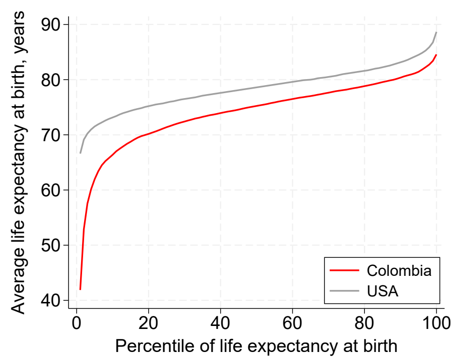
Note: Figure presents the geographic distribution of life expectancy at birth and socioeconomic status in Bogotá in Panels A and B, respectively. Panels C and D present the geographic distributions of life expectancy at birth and median household income in New York City. Darker blues represent higher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

Figure 2: Distribution of Life Expectancy at Birth in Colombia and the U.S.

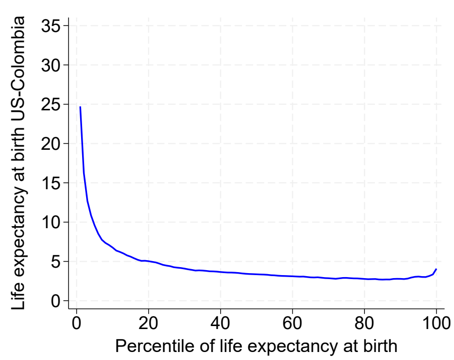
(a) Distribution of life expectancy at birth in Colombia and the U.S.



(b) Average life expectancy at birth by percentile



(c) Gap in life expectancy at birth between U.S. and Colombia



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 average life expectancy at birth by percentile of the distribution in Colombia (red) and the U.S. (gray). Panel C presents the gap in life expectancy at birth between the U.S. and Colombia for census tracts at each percentile of the distribution of life expectancy at birth.

Motivated by the patterns within cities suggesting a strong correlation between life expectancy at birth and measures of economic position, we next examine gaps in life expectancy at birth between the U.S. and Colombia and the extent to which these gaps can be explained by economic position. Figure 2, Panel A presents the distribution of life expectancy at birth across census tracts in Colombia (red) and in the U.S. (white). We see not only that the distribution of life expectancy in Colombia is shifted to the left relative to the U.S. but also that it has a long left tail, depicting census tracts where life expectancy at birth in 2018 was less than 60 years. Panel B shows the average life expectancy at birth by percentile of the distribution in each country. Overall, average life expectancy at birth is much flatter in the U.S. than in Colombia, and the within-country gap between top and bottom percentiles of the life expectancy distribution equals 20 years in the U.S. and 44 years in Colombia.

Panel C plots the difference in average life expectancy at birth between the two countries by percentile of the distribution. We find that the gap between countries is decreasing across percentiles. For example, among census tracts in the first percentile, the gap in average life expectancy at birth equals 25 years. This gap decreases rapidly to about 5 years among census tracts in the 20th percentile of the distribution, and further to 3 years for census tracts in the 99th percentile. In Appendix Figure 7 we conduct robustness checks on the number of census tracts per cluster finding that our estimates for the gap between the U.S. and Colombia is stable.

One challenge with quantifying the extent to which socioeconomic status explains the life expectancy gap between the two countries is that there is no corresponding measure of socioeconomic status for U.S. census tracts. To pursue this analysis, we leverage the fact that income is reported for census tracts in Bogotá to determine whether (1) disparities in life expectancy at birth by income closely match those by socioeconomic status in this city, (2) life expectancy patterns in Bogotá by socioeconomic status align with those in the entire country, and (3) a comparison of life expectancy gaps by socioeconomic status in Colombia against those observed in the U.S. by household income is possible.[2]

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.

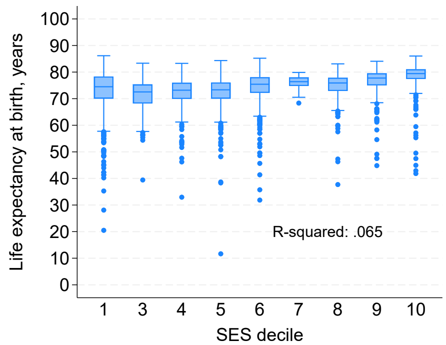
Given the similar patterns of mortality and life expectancy across socioeconomic status and income, and the fact that outcomes in Bogotá closely mirror national trends, we proceed to compare life expectancy in Colombia by socioeconomic status against the U.S. by household income. For ease of exposition, we refer to both measures as "income" hereafter. Figure 3, Panels A and B present the distribution of life expectancy at birth in each country conditional on income decile. The center line in each box represents the median of the life expectancy distribution, while the bottom and top of each box represent the 25th and 75th percentiles, respectively. The bottom whisker corresponds to the 25th percentile of the life expectancy distribution minus 1.5 times the interquartile range, and the upper whisker is the 75th percentile plus 1.5 times the interquartile range. Finally, each dot represents census tracts with life expectancies beyond the range captured by the whiskers.

Panel A shows that median life expectancy at birth in Colombia is weakly increasing with income and this is driven mainly by the increase in life expectancy among census tracts within the bottom whisker and the 25th percentile. Conditional on census tracts in the first percentile of the life expectancy distribution, the gap between top and bottom income deciles is 4 years, a gradient that remains mostly constant up to the 60th percentile of the life expectancy distribution. Importantly, even after conditioning on income, we find that the distribution of life expectancy at birth in Colombia continues to have a long left tail, suggestive of strong disparities in health along the distribution of wealth.

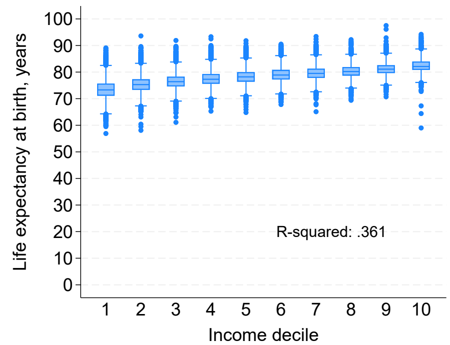
[2] Bogotá represents about 20% of the Colombian population and 10% of all census tracts in the country.

Figure 3: Mortality and Life Expectancy by Socioeconomic Status in Colombia and the U.S.

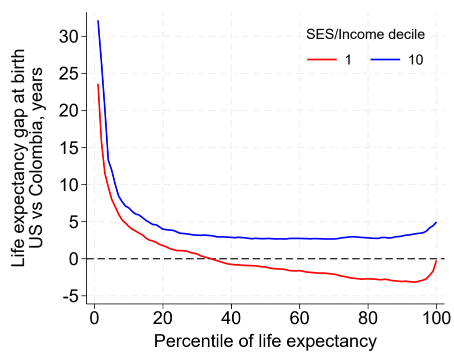
(a) Life expectancy distribution by SES in Colombia



(b) Life expectancy distribution by income in the U.S.



(c) Life expectancy gap at birth between the U.S. and Colombia by income



Note: Panels A and B present the distribution of life expectancy at birth conditional on each decile of socioeconomic stratum in Colombia and conditional on each decile of household income in the U.S., 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 stratum is a 0-6 categorization of residential units with the purpose of charging differential utilities fees. 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. Panel D presents the R-squared of a linear regression of life expectancy at birth on median per capita income in the U.S. and on socioeconomic status in Colombia. An observation in these regressions is a census tract.

The patterns in Colombia are substantially different from those in the U.S. reported in Panel B. First, while median life expectancy at birth increases with household income in the U.S., these gains are not just explained by improvements in life expectancy among census tracts below the bottom whisker, but also by those at the top of the distribution. Second, the distribution of life expectancy at birth within income decile is much more homogeneous in the U.S. than in Colombia. For instance, we do not observe a similar left tail in the U.S. after controlling for income. Third, conditional on the first percentile of the life expectancy distribution, the difference between top and bottom income deciles is 11 years in the U.S. This difference falls to 5 years at the top of the life expectancy distribution—a gradient that is much steeper than in Colombia.

Panels A and B report the proportion of the variation in life expectancy at birth that is explained by income in each country. These proportions correspond to the R-squared of a linear regression of life expectancy at birth on income, estimated separately for each country. Consistent with the idea that income is a stronger determinant of life expectancy in the U.S., we find that income explains 36% of the variation in the U.S., compared to just 7% in Colombia, nearly a 5-fold difference.

In Panel C, we directly plot the difference in life expectancy at birth between U.S. and Colombian census tracts in each percentile of the distribution, separately by top and bottom income deciles. The key takeaway is that gaps between the two countries conditional on the first percentile of the life expectancy distribution are strikingly large across all income deciles. For example, Americans live 30 years longer in the top income decile and 25 years longer in the bottom income decile. These gaps decrease substantially beyond the first percentile of the life expectancy distribution and remain mostly constant after the 40th percentile.

We also see that the difference between the two countries is significantly smaller towards the lower end of the income distribution across all percentiles of the life expectancy distribution (the red line is always below the blue line). For example, conditional on the 10th percentile of the life expectancy distribution, the difference between the U.S. and Colombia is 5 years in the lowest income decile and 7 years in the highest income decile. Remarkably, the gap reverses in the lowest income decile for census tracts above the 40th percentile of the life expectancy distribution. For instance, Colombians in census tracts at the 99th percentile live 3 years longer than Americans.

In Figure 4 we delve into additional sociodemographic characteristics that may explain patterns in life expectancy in each country—in particular the extremely low life expectancies observed across the socioeconomic distribution in Colombia.[3] We consider the fraction of the population who identifies as Black, the fraction of the population who during the last year migrated across municipalities in Colombia or across counties and states in the U.S., and the homicide rate. The blue plots present the first percentile of life expectancy at birth and the red plots present the maximum SES/income by deciles of these sociodemographic characteristics.

We find different gradients between the two countries for each characteristic. For example, while the first percentile of the life expectancy distribution is mostly flat with respect to the fraction of Blacks in Colombia, it declines substantially in the U.S. In fact, U.S. 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 U.S. than in Colombia. Migration rates are more strongly correlated with the first percentile of life expectancy at birth in Colombia than in the U.S. And, in both countries, higher homicide rates are associated with lower life expectancies. For example, in Colombia, the gap in the first percentile of life expectancy at birth between census tracts in the top and bottom deciles of the homicide rate is nearly 20 years.

However, not all of these sociodemographic characteristics explain the low life expectancies at high levels of socioeconomic status in each country. For instance, the red plots show a steep negative gradient between the homicide rate and the maximum socioeconomic status/income in each country, suggesting violence is a factor that affects survival only in poorer areas of Colombia and the U.S. Instead, the middle panels show nearly zero correlation between the maximum socioeconomic status and the migration rate in Colombia, while the correlation in the U.S. is strongly positive. This indicates that perhaps forced migration can explain some of the low life expectancies within Colombia across the distribution of socioeconomic status.

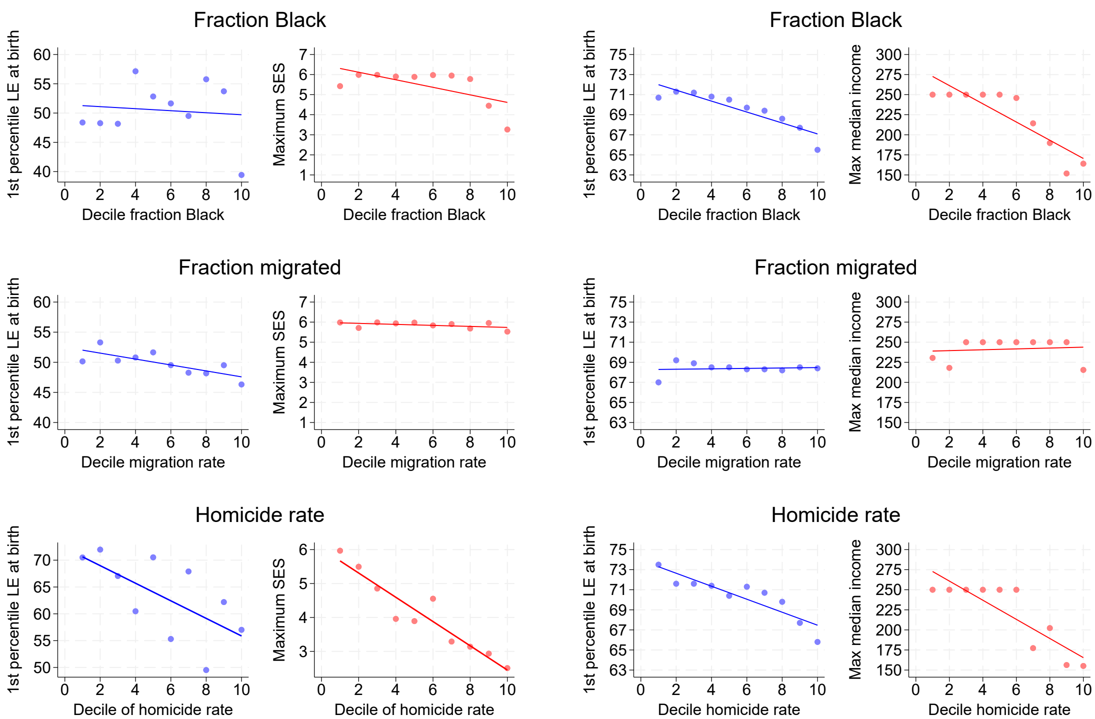
3 Discussion

Our results provide evidence to the hypothesis that disparities in life expectancy within LMICs are not easily explained by income as they are in high-income countries, despite presenting higher levels of 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 against the U.S., our analysis identifies 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. The fact that this disparity persists across the income distribution, suggests that some potential culprits like violent deaths, which are generally concentrated among the poor, are not primarily responsible. The longer left tail and the higher variance in life expectancy at birth in Colombia, suggest that health inequalities are driven by more than just economic status

[3]For this analysis in Colombia, we use sociodemographic characteristics reported by census tract from the 2018 census and geo-located homicide records for the city of Medellín. In the U.S., we use sociodemographic characteristics by census tract from the 2011-2015 American Community Survey and homicide records from the Center of Disease Control and Prevention accessed through CDC, 2025.

Figure 4: Distribution of Life Expectancy at Birth by Deciles of Demographic Characteristics in Colombia and the U.S.

(a) Gradients for Colombia (b) Gradients for the U.S.



Note: The blue plots present the first percentile of 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 maximum 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.

and are perhaps explained in part by forced migration. While economic position remains a strong determinant of health, our results point to the need for addressing broader structural and social factors to improve survival outcomes within LMICs.

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.

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, 629â652.

Guzman, G. and M. Kollar (2024): "Income in the United States: 2023, Current Population Reports," *United States Census Bureau.*

IDB (2024): "The Complexities of Inequality in Latin America and the Caribbean," *International Development Bank, Fact Sheets.*

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.

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

4 Materials and Methods

4.1 Data for the United States

For the U.S., we obtain 2010 to 2015 estimates of life expectancy at birth by census tracts from the National Center for Health Statistics at the Centers for Disease Control and Prevention (CDC). These estimates were constructed by the U.S. 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 and 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 homicide rates at the census tract level from the CDC.

4.2 Data for Colombia

We use detailed census tract (*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 tracts 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 tracts in total).

For each census tract, 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 stratum (*estrato*), which we term "socioeconomic status" throughout the paper. 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 reporting income by census tracts in this city. To perform analyses of life expectancy by income, we link census tracts in Bogotá from the 2018 population census with census tracts 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 tract 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 tract population reported in the 2018 census.

4.2.1 Clustering of census tracts

Colombian census tracts 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 tracts 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 tracts after excluding 101,080 tracts that lacked socioeconomic status data. 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 tracts per cluster were initially unknown, we approximated k by dividing the total number of census tracts 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 tract and the geographic distance between tracts—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 tracts in Bogotá we additionally implemented the clustering analysis modifying the distance function in the assignment step by incorporating census tract average household income and geographic distance between census tracts. 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 tracts throughout the country, 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 available at Gonzaga and Schmertmann (2016).

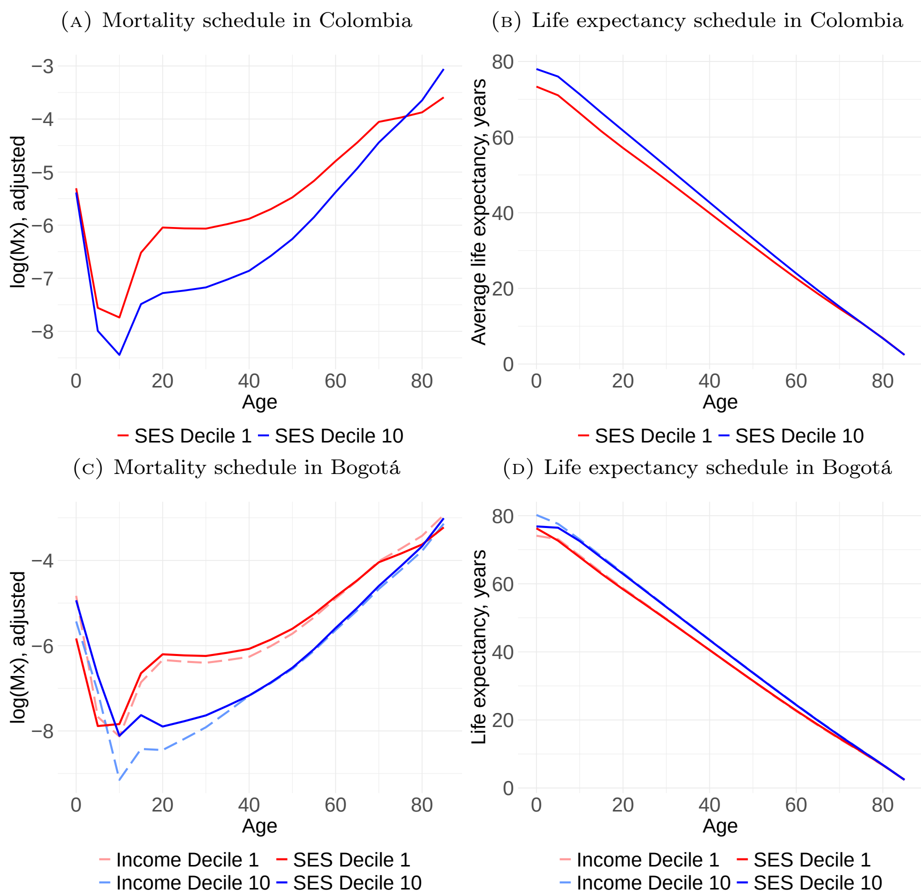
We adopted an approach that considers nested geographic areas to enhance the accuracy of the method. At each level, we used mortality patterns as standard schedules and applied the TOPALS algorithm at each stage to fit mortality schedules for smaller geographic areas, capturing variations in mortality 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 Clustering Methodology

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



(A) Mortality schedule in Colombia

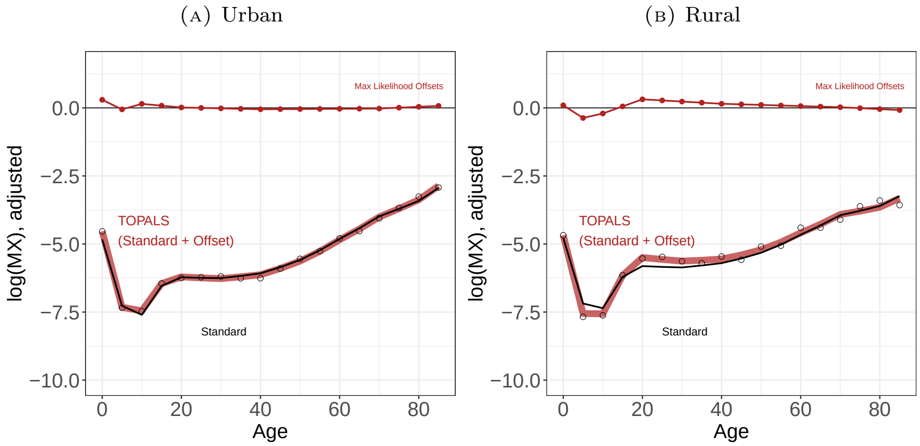
(B) Life expectancy schedule in Colombia

(C) Mortality schedule in Bogotá

(D) Life expectancy schedule in Bogotá

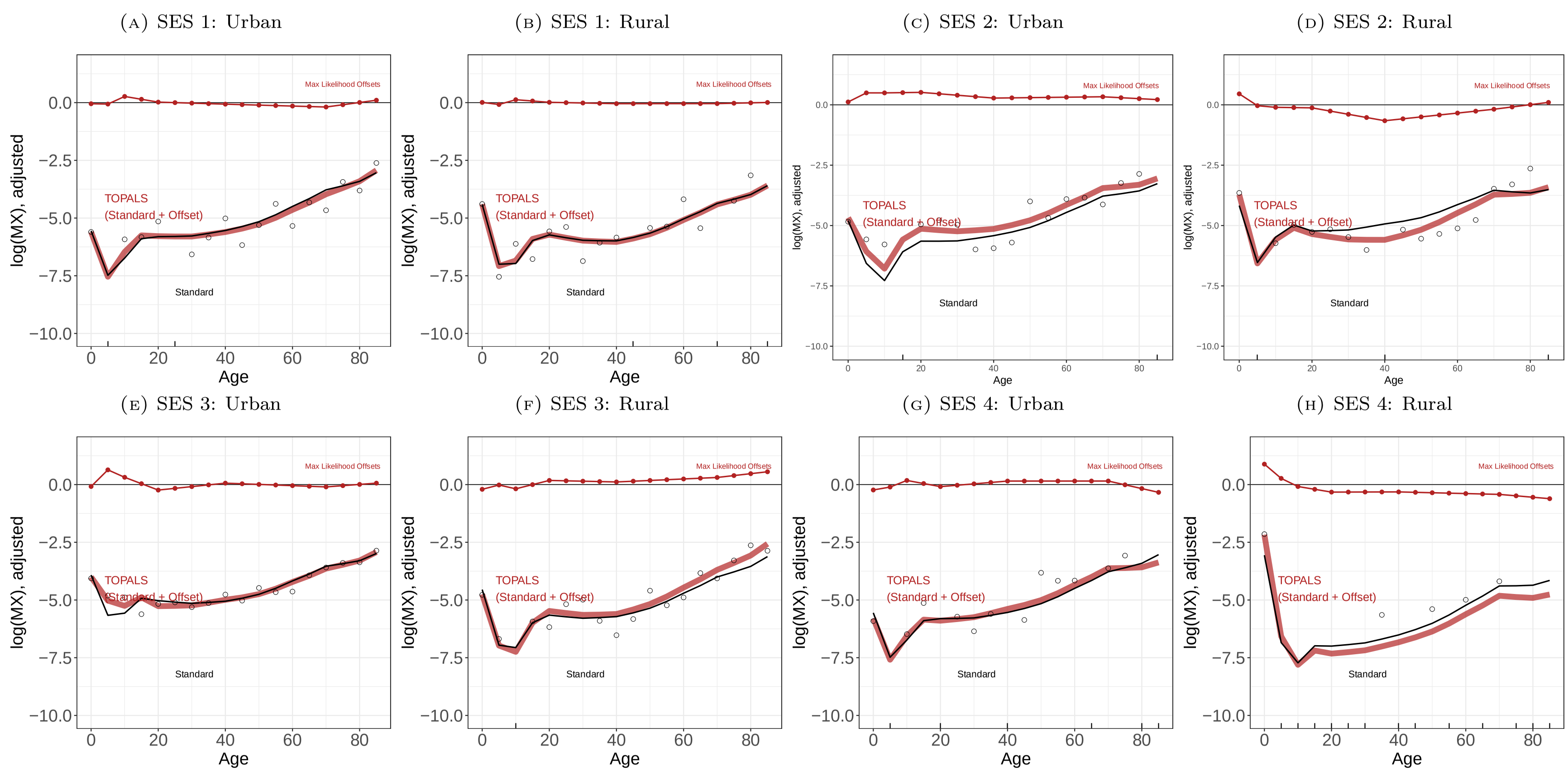
Note: Panel A shows the average log mortality rate at age 0 and by 5-year age groups thereafter across census blocks in the top (blue) and bottom (red) deciles of socioeconomic stratum in Colombia. Panel B presents the implied average life expectancy at birth and by 5-year age groups thereafter across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

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



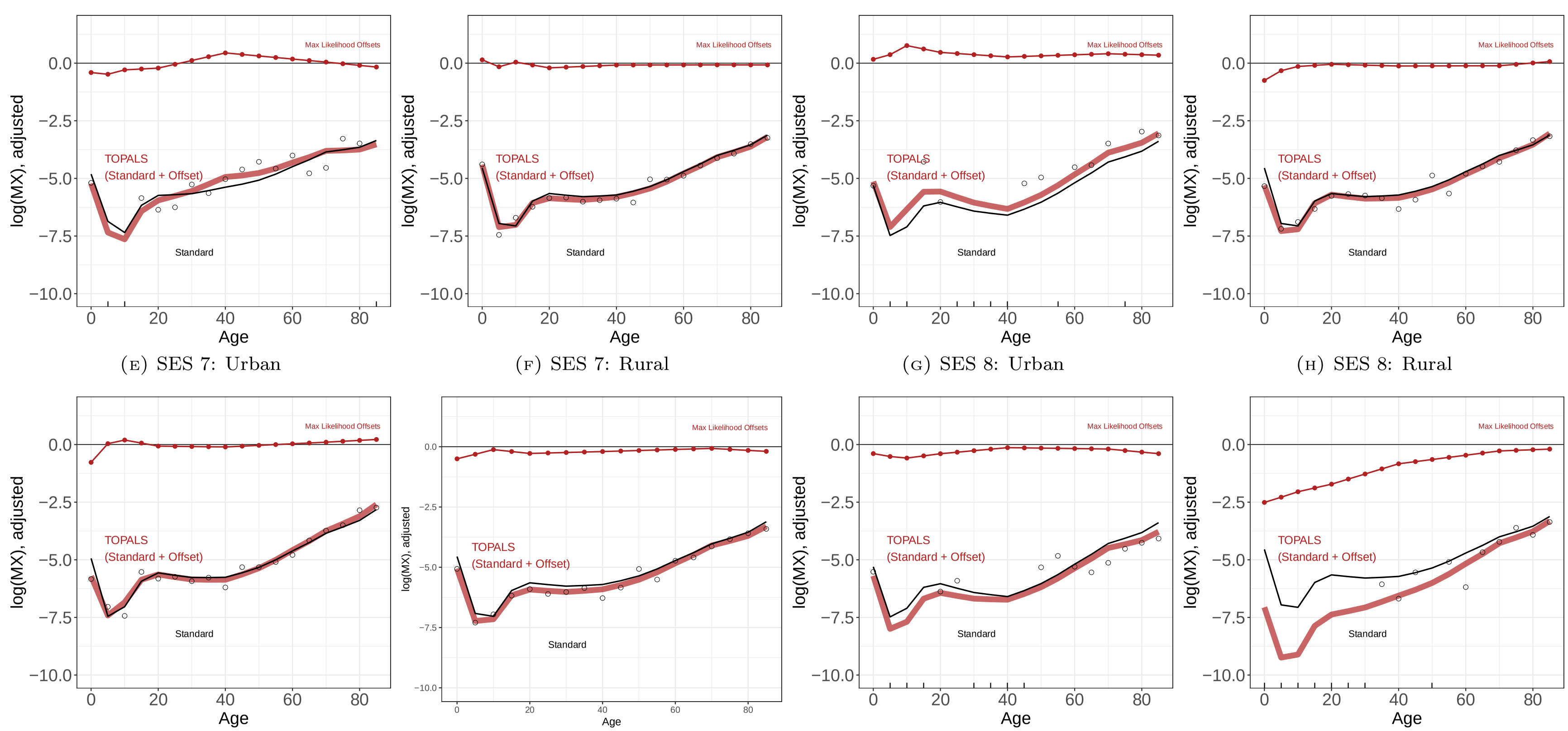
Note: Black lines in Panels A and B represent the standard mortality schedule for Colombia in 2018, derived from the WPP 2024 projections. The corresponding red lines show the fitted mortality schedules by age for urban (Panel A) and rural (Panel B) regions in the state of Antioquia, obtained using the 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 Stratum 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 (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 Stratum in Urban and Rural Colombia (Deciles 5-8)



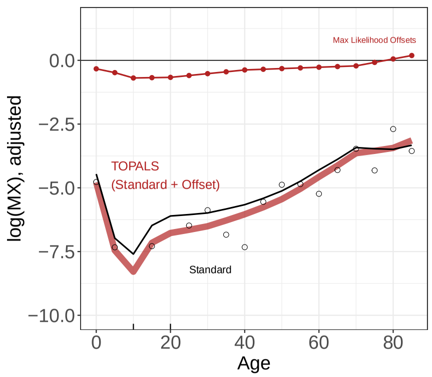
(A) SES 5: Urban (B) SES 5: Rural (C) SES 6: Urban (D) SES 6: Rural

(E) SES 7: Urban (F) SES 7: Rural (G) SES 8: Urban (H) SES 8: Rural

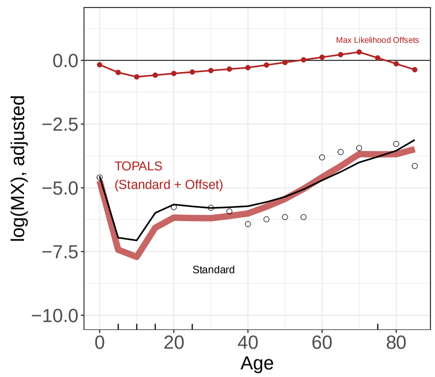
Note: Each panel displays the fitted mortality schedule for urban and rural areas, stratified into deciles of socioeconomic status (deciles 5 to 8). The standard schedules (black lines) are derived from the previous estimation step, fitted for the overall urban and rural areas, as shown in Appendix Figure 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 Stratum in Urban and Rural Colombia (Deciles 9-10)

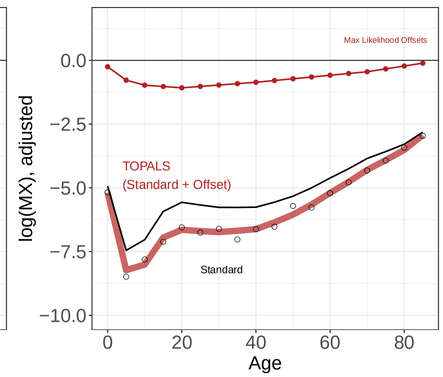
(a) SES 9: Urban



(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 (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 B Vital Statistics vs. Census Data

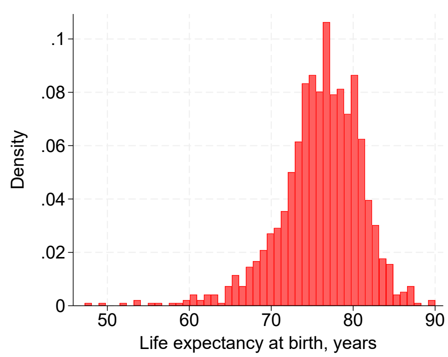
We obtain publicly available 2018 census data at the municipality level from the National Administrative Department of Statistics (DANE). These data include the population by 5-year age groups for 2018, and the number of deaths by 5-year age groups for 2017. The data also report the number of deaths disaggregated by whether a death certificate was issued. Of the 244,461 total reported deaths, only 183,389 have an associated death certificate.

Appendix Figure 6 shows the distribution of life expectancy at birth across municipalities. 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.

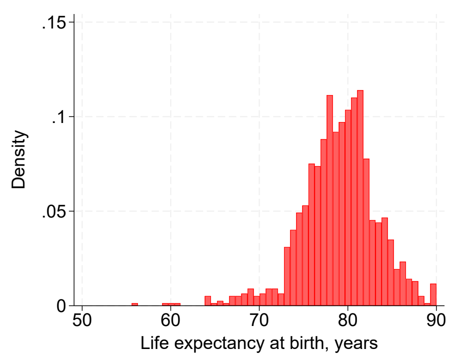
Panel C plots the geographic distribution of the fraction of total deaths that are uncertified. These types of deaths tend to be more common in municipalities with a history of forced migration such as those in Chocó, Llanos Orientales, and the Amazon.

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

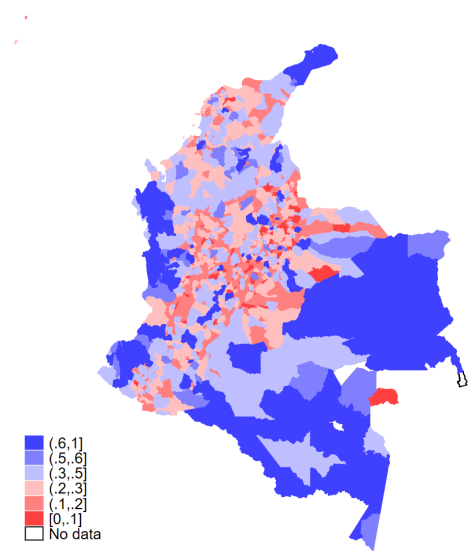
(a) All deaths



(b) Only deaths with death certificates



(c) Geographic distribution of uncertified deaths

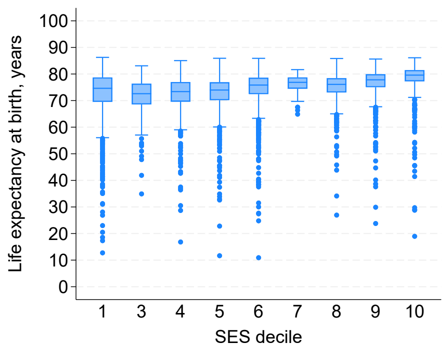


Note: Figure presents the distribution of life expectancy at birth using municipality level census data from the 2018 Colombian census. Panel A presents the distribution considering all reported deaths. Panel B presents the distribution considering only deaths with a death certificate. Panel C presents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

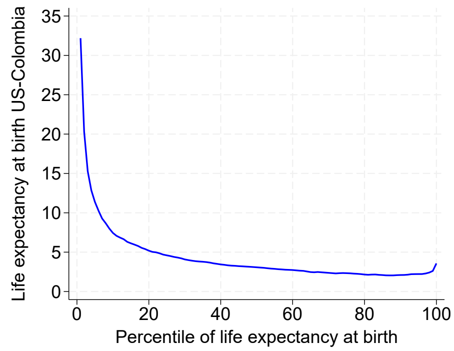
Appendix C Additional Results

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

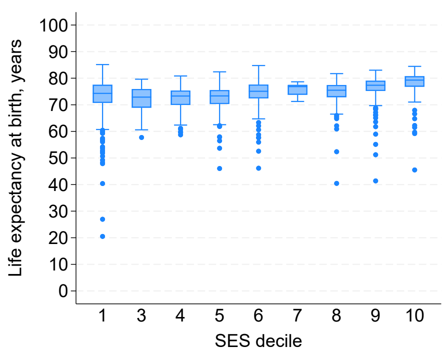
(a) Life expectancy distribution: 50-tract clusters



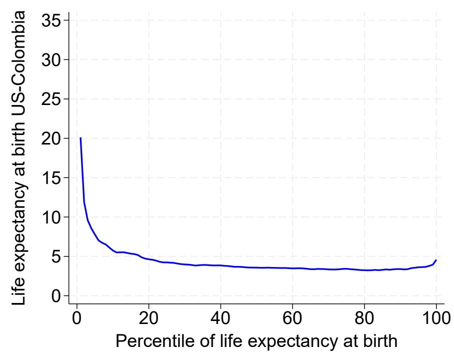
(b) U.S. vs Colombia (50-tract)



(c) Life expectancy distribution: 200-tract clusters



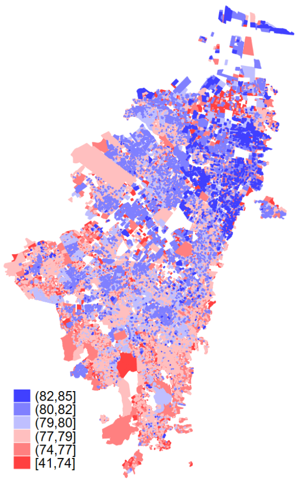
(d) U.S. vs Colombia (200-tract)



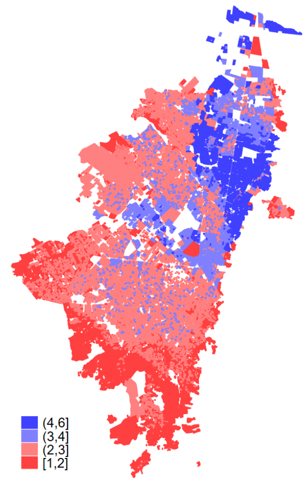
Note: Panels A and C present the distribution of life expectancy at birth conditional on each decile of socioeconomic status in Colombia. An observation is a cluster of census tracts. Clustering is performed using K-means with a minimum number of tracts per cluster equal to 50 in Panel A and to 200 in 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 stratum 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 borth between the U.S. and Colombia by percentiles of the life expectancy distribution. Panel B uses 50-tract clusters in Colombia and Panel D uses 200-tract clusters.

Appendix Figure 8: Geographic Distribution of Life Expectancy at Birth and SES in Colombia - Part 1

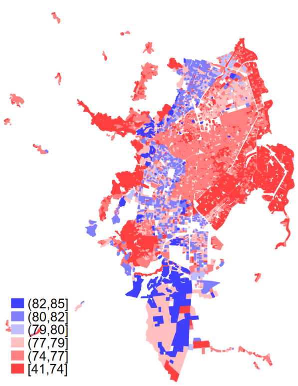
(a) LE Bogotá



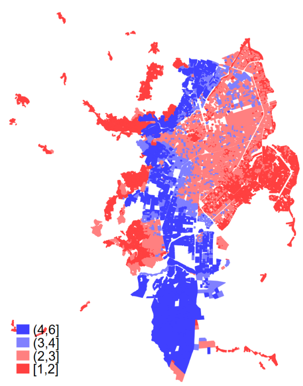
(b) SES Bogotá



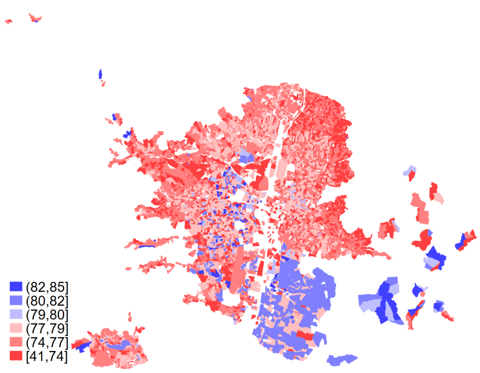
(c) LE Cali



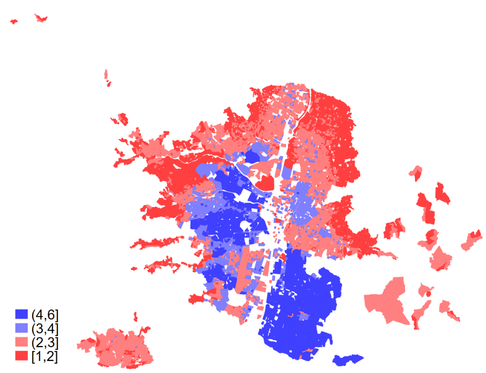
(d) SES Cali



(e) LE Medellín



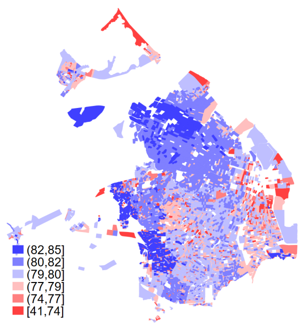
(f) SES Medellín



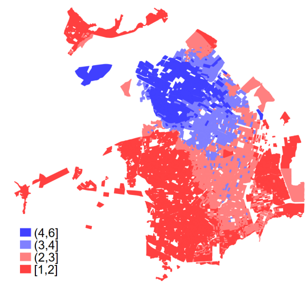
Note: Figure presents the distribution of life expectancy at birth and socioeconomic status by census tracts within the cities of Bogotá, Cali, and Medellín. Darker blues denote higher values and darker reds denote lower values.

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

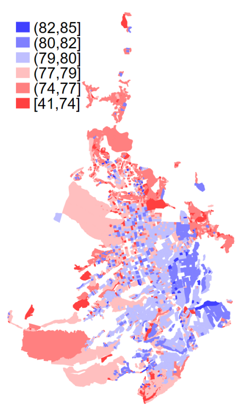
(a) LE Barranquilla



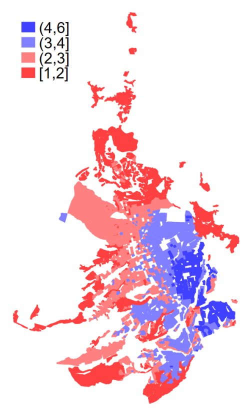
(b) SES Barranquilla



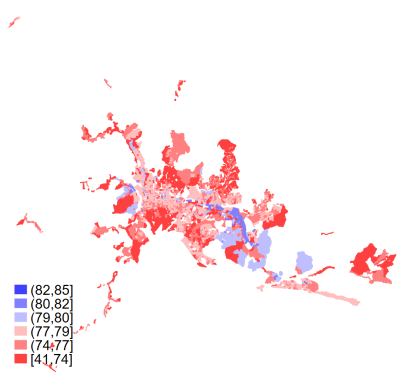
(c) LE Bucaramanga



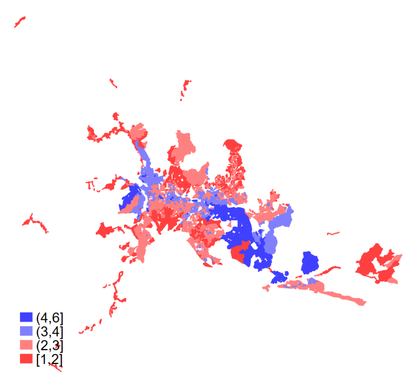
(d) SES Bucaramanga



(e) LE Manizales



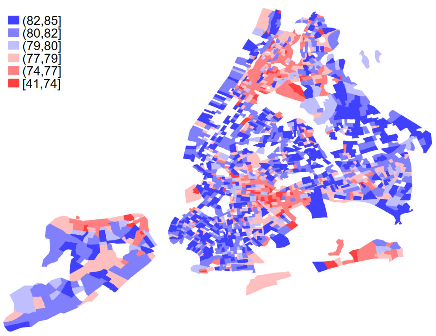
(f) SES Manizales



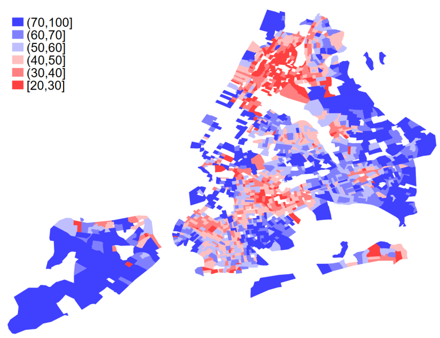
Note: Figure presents the distribution of life expectancy at birth and socioeconomic status by census tracts within the cities of Barranquilla, Bucaramanga, and Manizales. Darker blues denote higher values and darker reds denote lower values.

Appendix Figure 10: Geographic Distribution of Life Expectancy at Birth and SES in the U.S. - Part 1

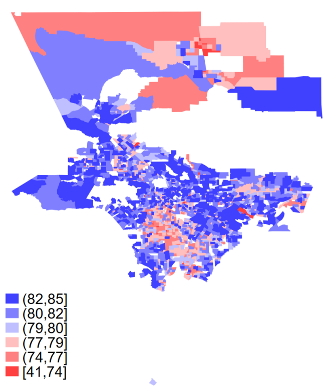
(a) Life expectancy NYC



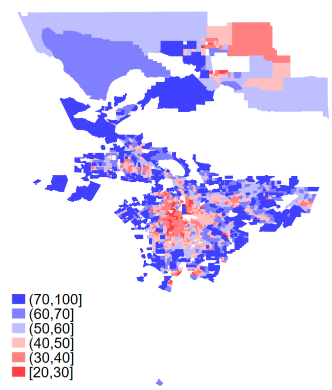
(b) Median household income NYC



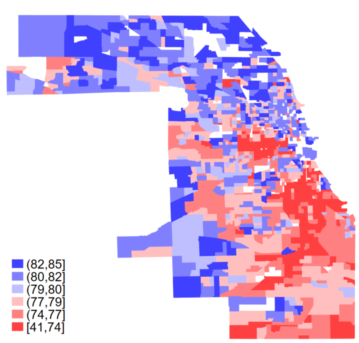
(c) Life expectancy Los Angeles, CA



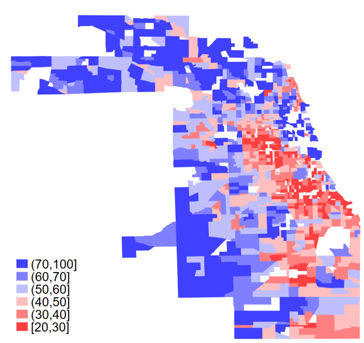
(d) Median household income, Los Angeles, CA



(e) Life expectancy Chicago, IL



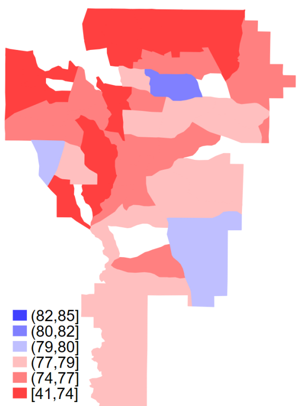
(f) Median household income, Chicago, IL



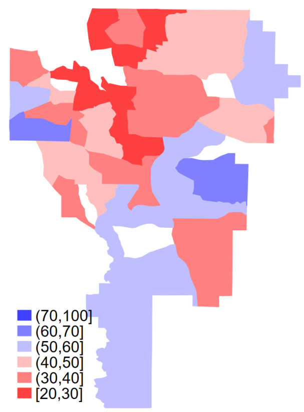
Note: Figure presents the distribution of life expectancy at birth and median household income by census tracts within the cities of New York City, Los Angeles, and Chicago. Darker blues denote higher values and darker reds denote lower values.

Appendix Figure 11: Geographic Distribution of Life Expectancy at Birth and SES in the U.S. - Part 1

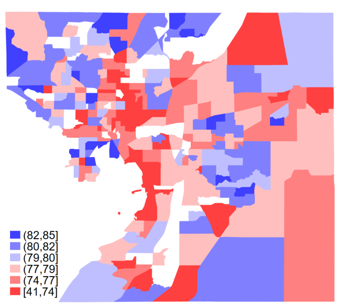
(a) Life expectancy Atlanta, GA



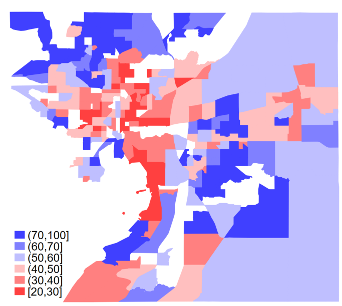
(b) Median household income Atlanta, GA



(c) Life expectancy Miami, FL



(d) Median household income, Miami, FL



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