Urban Preston Curves in Black and White:
Is there a Causal Link?

Avik Chakrabarti *

July 2024

*Associate Professor of Economics, University of Wisconsin-Milwaukee (UWM).
Affiliated Faculty, Northwestern Mutual Data Science Institute (NMDSI).
Email: chakra@umich.edu

Urban Preston Curves in Black and White: Is there a Causal Link?

A vast and growing empirical literature aims at identifying plausible links between health and income. While the most compelling evidence upholding any such relationship has typically been borne by cross-country data, the relevant literature is extensive and controversial leaving policymakers confounded. We revisit the popular hypothesis, that the diseases of the poor are diseases of poverty, using granular data on African American and Caucasian populations from urban counties in the US. In doing so, we quantify a *causal* link between urban life expectancy and household income by employing a machine learning approach harnessing optimum explanatory power conditional on data. We observe that every percentage point increment in household income raises life expectancy by at least 0.10 percentage points: the point elasticity of life expectancy with respect to household income varies between 0.10 and 0.142 among the African American population and between 0.145 and 0.198 among the Caucasian population of urban counties of the US.

JEL Classification Code(s): R10, D91, F1, J17, I12, O15.

Keywords: Urban Life Expectancy, Income, Ethnicity, Preston Curve, Machine Learning, Instrumental Variables.

1. Introduction

The most notable relationship between health and income has typically been borne by cross-country data since Preston (1975) discovered a plausible association between life expectancy at birth and real per capita GDP from cross-country data. This association, popularly known as the "Preston curve" in the relevant literature that followed, has been replicated in a wide range of settings. At the dawn of the new millennium, Deaton (2003) drew renewed attention toward a Preston curve that fits well with data from 2000: the correlation between the log of GDP per capita and life expectancy is 0.84 when countries are weighted by population (and only slightly lower if they are not weighted). The relationship comes across as nearly linear: a doubling of GDP per capita is associated with an increase in life expectancy of roughly five years. The curvature of the relationship between income and health has often led to the suggestion that a policy of redistributing income from the rich to the poor will improve average health outcomes with the gains in health of those earning low incomes outweighing the losses of those earning high incomes. Extant evidence upholding the Preston curve, as Cutler et al. (2006) underscored, has often been construed to imply a causal link between income and health to claim, "If the diseases

of poor … are indeed 'diseases of poverty' in the sense that they will vanish if poverty is reduced, then direct health interventions may be less important than economic growth. Economic growth would be 'twice blessed'; it would increase material living standards directly and improve health as a bonus."

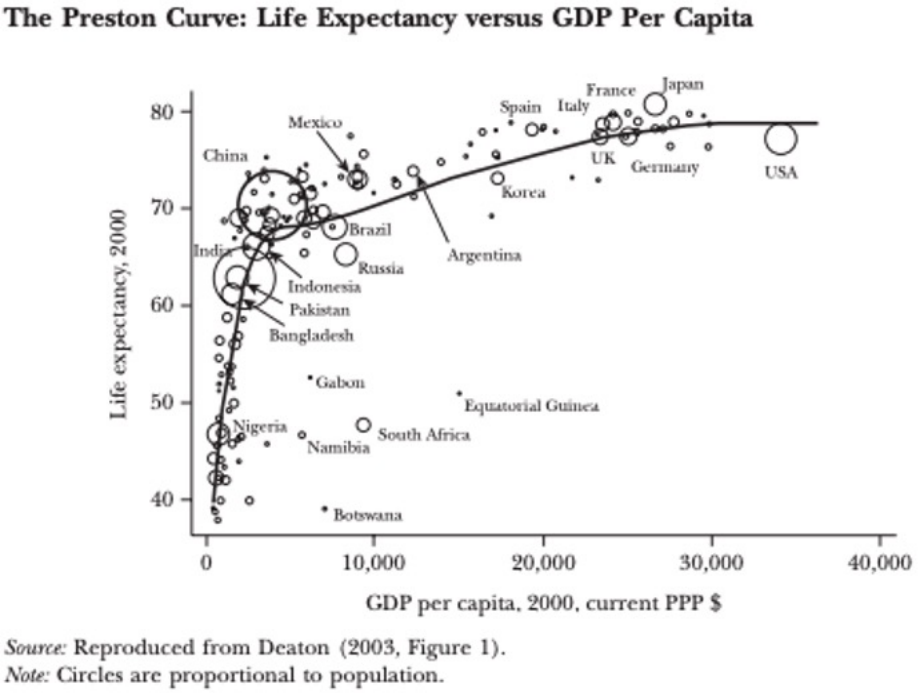


Figure 1. The Preston Curve

Any such policy prescription must rely on the relationship between income and health being *causal* rather than income merely acting as a proxy for some broader notion of Socio-Economic Status (SES). By the same token, there is no compelling reason to refute the possibility that a link may run in the reverse direction from improved health to higher incomes. Healthier workers are more productive, and longer lifespans create incentives to invest in schooling and save for retirement. The policy implications of health to income causality are non-trivial to the extent that they suggest that reducing inequalities in health can be a powerful instrument of economic development and poverty reduction. In addition,

of course, healthier children are likely to attend school more regularly, more easily absorb knowledge while in school, and increase their cognitive ability. Also, while most compelling evidence upholding any relationship between health and income has typically been borne by cross-country data, a fundamental problem with making international comparisons of wellbeing is that there is no standard metric for health. More importantly, any regression analysis presupposes that observations are drawn from a distinct population. As such, the validity of inferences drawn from any cross-country regression analysis can be questioned on the ground that countries may have very little in common that can merit including them in the same regression. Also, notwithstanding the credibility of the sources that allow access to relevant data, inferences drawn from any cross-country analysis must be interpreted with caution since the quality and reliability of data may be sensitive to reporting accuracies that can vary across countries. To tackle these challenges, we use granular data on African American and Caucasian residents populating urban counties in the US to test if the diseases of the poor are diseases of poverty. In the next section, we review the relevant literature on the Preston curve. In section 3, we describe the data and estimation used in this paper and present our findings. We conclude in the final section.

2. The Preston Curve

Preston (1975) was the earliest to investigate the links between life expectancy and per capita GDP, when he observed a positive association between the two and that successive increases in per capita income resulted in less than proportional increases in life expectancy in 1900s, 1930s, and 1960s. As de la Escosura (2023) recollects, the Preston curve shifted over time with higher longevity corresponding to the same per capita income level: this led to the conclusion that increases in per capita income only partially explained the rise in

longevity, which was attributable to other factors, including the diffusion of medical advances resulting from the 'empirical validation' of the germ theory of disease in the developing as well as developed world. Since Preston's landmark contribution, scientists have studied the extent and explanation of the increased dissociation between longevity and per capita income. Wilkinson (2007) argued that material assets, as much as income, affect health, and Bloom and Canning (2007) considered that per capita income captures SES, while Preston (2007) pointed to the existence of lags in the relation between income and mortality which may operate over the life cycle and through wealth-generating processes. Attempts have been made to explore the empirical association between life expectancy and per capita income beyond the time span considered by Preston. Kakwani (1993) confirmed that the association between life expectancy at birth and GDP per head shifted upwards between the 1970s and 1980s, meaning higher longevity for a given per capita income level. Mackenbach and Looman (2013) replicated the Preston curve for 1900–2008. Their findings confirm Preston's interpretation for the pre-1960 period and suggest that per capita income explained most of the life expectancy gains from 1960 onwards. They attributed such a discrepant behavior to the fact that, before 1960, the diffusion of medical knowledge to fight infectious disease, the leading cause of mortality reduction, was less dependent on income levels than in the late twentieth century, when the decline in mortality from cardiovascular disease, from which longer life expectancy resulted, had a significantly higher cost. From a large country sample over the last two centuries, Jetter et al. (2019) confirmed the existence of a non-linear relationship between per capita income and life expectancy and that economic growth has been the main driver of long-run gains in life expectancy while the impact of medical innovations was largely

restricted to higher income countries. Based on a series of replication exercises, de la Escosura (2023) reports his findings from having tested Preston's association between life expectancy at birth and real per capita income for a sample of 112 countries over 1870–2015. Instead of using the original values of life expectancy at birth and real per capita GDP, this paper estimated the statistical relationship between the Kakwani index of life expectancy and the natural logarithm of real per capita GDP using polynomial regressions at different cross-sections, corresponding to 1870, 1913, 1950, 1970, and 2015, that mark the phases of life expectancy's evolution. The predicted values for the health proxy (the Kakwani index of life expectancy) from the resulting relationship become convex, rather than concave, to the log of per capita income, suggesting that advances in medical knowledge lead to more than proportional gains in health as income rises. Moreover, the exercise confirms the existence of sustained outward shifts in the life expectancy-per capita income relationship over time, so the Preston curve does not break down: strong outward shifts between 1913 and 1950 and between 1970 and 2015 contrast with moderate ones during 1870–1913 and 1950–1970. This study also replicated Preston's attempt to 'factor out' the contribution of per capita income to life expectancy gains at each cross-section. Thus, in order to estimate the contribution of per capita income to the rise of the Kakwani index of life expectancy over 1870 and 1913, the author combined the parameters of their statistical relationship for the 1870 cross-section (the initial year) with the values of real per capita income (in logs) for 1870 to get the predicted level of the Kakwani index of life expectancy, and with the values of real per capita income (in logs) for 1913 to obtain the counterfactual level of the Kakwani index had there been no medical advancement during the period 1870–1913, and all changes resulted exclusively from a higher level of per capita

income. The author concluded that modern economic growth contributed above 60 per cent of the increase in the Kakwani index of life expectancy over 1870–1913 and, again, during the last period, 1970–2015. However, its contribution was substantially lower during the first of half of the twentieth century, when it only provided only a quarter of the Kakwani index gains while three-fourths of them accrued to advances in medical knowledge. Conversely, during the Golden Age (1950–1970) the increase in per capita income accounted for the entire improvement in the Kakwani index of life expectancy. In sum, over the past centuries and a half, economic growth and medical advances appear to have contributed alike to gains in life expectancy.

3. Data, Estimation, and Findings

With this backdrop, we revisit the popular hypothesis that the diseases of the poor are diseases of poverty using granular data on African American and Caucasian populations from urban US counties. All data used in this paper are publicly available from Population Level Analysis and Community Estimates (PLACES), American Community Survey (ACS), and National Center for Health Statistics (NCHS). According to the Centers for Disease Control and Prevention (CDC), PLACES is an expansion of the 500 Cities Project. Launched by the CDC, in partnership with the Robert Wood Johnson Foundation, the 500 Cities Project was expanded to provide data that spans up to 72,337 census tracts; 32,409 ZIP Code Tabulation Areas (ZCTA); 28,484 Census Designated Places (CDP); and 3,143 counties hosting more than 332 million people across the 50 states and District of Columbia of the United States of America. The ACS reports earnings for each household. Life expectancy is measured, based on mortality files from the NCHS, by the average number of years a person can expect to live.

Table 1. OLS: Households in Urban US Counties, 2021

Regression	1	2	3	4	5	6	7	8
log (Median Household Income)	0.1153717***	0.0898128***	0.0952471***	0.0797034***	0.0806064***	0.0495646***	0.0474594***	0.0491782***
log (Motor Vehicle Crash Death)		-0.0204174***	-0.023098***	-0.0226655***	-0.0232559***	-0.0229195***	-0.0106483***	-0.0107831***
log(% 65 and Older)			0.0370421***	0.0294189***	0.0297761***	0.0294017***	0.0474903***	0.0446377***
log(Preventable Hospital Stays)				-0.0245196***	-0.0242817***	-0.0217462***	-0.0198154***	-0.0183193***
log(Flu Vaccinations)					-0.0030446	-0.0052758**	-0.0087118***	-0.0063103***
log(Children in Poverty)						-0.0217646***	-0.016708***	-0.0105539***
log(Injury Deaths)							-0.0540868***	-0.0525491***
log(Low Birthweight)								-0.0182989***
Constant	1.337046***	1.482066***	1.412215***	1.580094***	1.580267***	1.750538***	1.820211***	1.812561***
Sample size	2438	2094	2094	2090	2090	2090	2090	2090
F-statistic	2009.36	1478.20	1184.48	1077.16	862.37	761.26	955.21	865
Adjusted R-squared	0.4518	0.5853	0.6291	0.6733	0.6734	0.6859	0.7618	0.7678

Notes: Dependent variable is log of Median Life Expectancy.

*** indicates that a regression coefficient is statistically significant at 1% level

** indicates that a regression coefficient is statistically significant at 5% level; and

* indicates that a regression coefficient is statistically significant at 10% level.

With life expectancy as the dependent variable and income as our variable of interest, we set out to estimate following specification:

$$\log(Median\ Life\ Expectancy_c) = \beta_0 + \beta_0 \log(Median\ Household\ Income_c) + \Gamma Z + \epsilon_c$$

where $Median\ Life\ Expectancy_c$ and $Median\ Household\ Income_c$ represent the life expectancy and income of the median household, respectively, in urban county c; vector Z lists controls with vector Γ collecting the corresponding parameters; and ϵ_c captures errors. Table 1 above reports our benchmark results from Ordinary Least Squares (OLS) estimation of this specification. Figure 2 below compares the life expectancy and income, for the median household from African American and Caucasian populations in urban US.

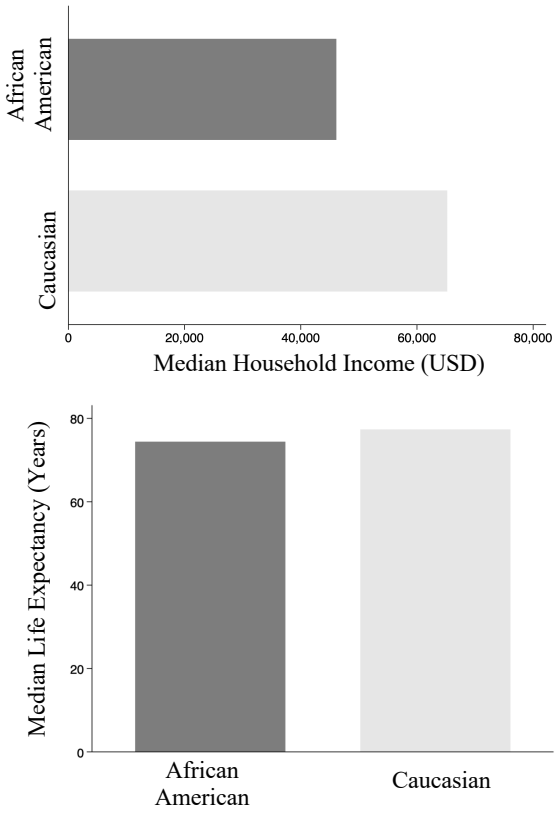


Figure 2. Median Household Income and Life Expectancy:
Caucasian and African American Populations in urban US (2021)

Figure 3 below plots Preston curves, using non-parametric locally weighted scatterplot smoothing, for African American and Caucasian households in urban US.

Urban Preston Curves in Black and White

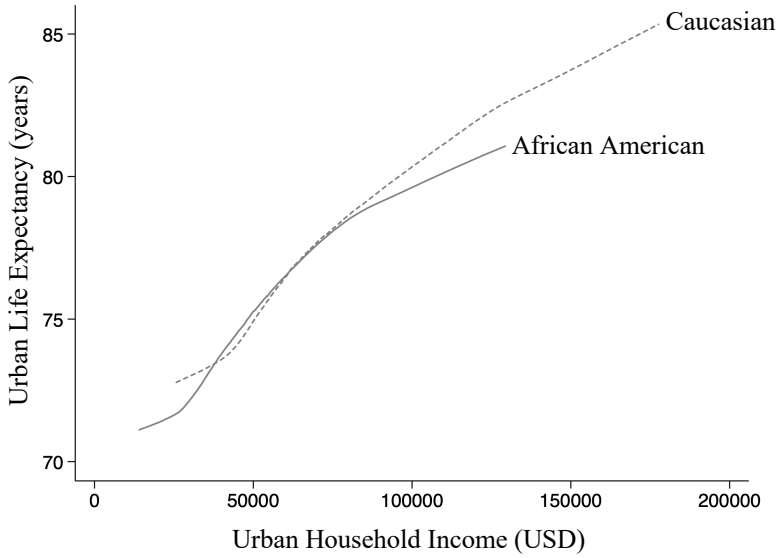


Figure 3. Urban Preston Curves: Caucasian and African American Populations in the US (2021)

Our OLS estimation is, of course, open to caveats. For instance, there may be unobserved heterogeneities across populations belonging to different ethnicities. Measurement errors in household income can introduce bias into OLS estimates. Household incomes in the ACS are likely to be reported with error and, that too for the previous twelve-month period rather than any calendar year. It is also possible that omitted variables introduce bias to the extent that they are correlated with the household income. Moreover, as members of healthier households may be more productive and longer lifespans create incentives to invest in schooling and save for retirement, the direction of causality between life expectancy and household income is ambiguous. To circumvent problems related to unobserved heterogeneity, we run separate regressions with observations on distinct African American and Caucasian households. For selection of controls and instrumental variables, we use a machine learning approach (Least Absolute Shrinkage and Selection Operator) harnessing optimum explanatory power conditional on data: standardized test scores are included as instruments correlated with household income but not with life expectancy to tease out any causal effect of the former on the latter.

Table 2. IV: African American Households in Urban US Counties, 2021

Regression	1	2	3	4	5	6	7	8
log (Median Household Income)	0.1117481***	0.1275893***	0.129521***	0.1266157***	0.126119***	0.1417887***	0.1013439***	0.1007487***
log (Motor Vehicle Crash Death)		0.0037047	0.0015547	0.0008932	-0.0002692	-0.0015158	0.0010569	0.0031352
log(% 65 and Older)			0.0043323	0.0045464**	0.0043521**	0.0026391	-0.0028714	-0.0016113
log(Preventable Hospital Stays)				-0.0103074**	-0.0106407**	-0.0127147***	-0.0100134**	-0.0086954**
log(Flu Vaccinations)					-0.0108817	-0.0096162	-0.0078142	-0.00562
log(Children in Poverty)						0.0221075*	0.0193927*	0.0205188**
log(Injury Deaths)							-0.050778***	-0.0508957***
log(Low Birthweight)								-0.0203293
Constant	1.353533***	1.272061***	1.260484***	1.312191***	1.334405***	1.238192***	1.516562***	1.528848***
Sample size	816	465	465	465	465	465	465	465
F-statistic	131.70	142.20	100.69	86.68	70.04	68.23	118.20	110
Adjusted R-squared	0.3050	0.4528	0.4509	0.4670	0.4697	0.4587	0.6523	0.6560

Notes: Dependent variable is log of Median Life Expectancy.

*** indicates that a regression coefficient is statistically significant at 1% level

** indicates that a regression coefficient is statistically significant at 5% level; and

* indicates that a regression coefficient is statistically significant at 10% level.

Table 3. IV: Caucasian Households in Urban US Counties, 2021

Regression	1	2	3	4	5	6	7	8
log (Median Household Income)	0.1539309***	0.1518008***	0.1507601***	0.1507869***	0.145384***	0.1983688***	0.187325***	0.1943877***
log (Motor Vehicle Crash Death)		-0.0048866	-0.0038396	0.0022915	0.0023076	0.0067919	0.0064828	0.0086667
log(% 65 and Older)			0.0173421***	0.0182285***	0.0172116***	0.0200068***	0.0190124***	0.0198004***
log(Preventable Hospital Stays)				-0.0212523***	-0.0223516***	-0.0250058***	-0.0253622***	-0.0242993***
log(Flu Vaccinations)					0.0082126	0.0159314*	0.01663**	0.0185787**
log(Children in Poverty)						0.0245833**	0.0232443**	0.0250662**
log(Injury Deaths)							-0.007287	-0.0053219
log(Low Birthweight)								-0.009436
Constant	1.144432***	1.157794***	1.131111***	1.195374***	1.213109***	0.9178167***	0.9890853***	0.9464056***
Sample size	1112	628	628	622	622	622	622	622
F-statistic	459.05	369.64	263.23	236.31	198.40	126.75	121.51	104
Adjusted R-squared	0.3720	0.4875	0.5202	0.5448	0.5652	0.4072	0.4528	0.4240

Notes: Dependent variable is log of Median Life Expectancy.

*** indicates that a regression coefficient is statistically significant at 1% level

** indicates that a regression coefficient is statistically significant at 5% level; and

* indicates that a regression coefficient is statistically significant at 10% level.

Tables 2 and 3 above summarize the results from Instrumental Variables (IV) estimation quantifying causal links between life expectancy and household income of African American and Caucasian households, respectively, in urban US. We infer that a significant robust positive association exists between income and life expectancy of the urban population in the US and, to the extent that our IV estimates capture causality, the effect of every percentage point increment in household income on life expectancy ranges from 0.10 to 0.142 among the African American population and from 0.145 to 0.198 among the Caucasian population in urban counties of the US.

4. Conclusion

Debates on policy implications of observed associations between income and health has attracted much attention and controversy. While mounting evidence on positive associations between income and health has given rise to the widespread view that increased income leads to increased health, some scientists hold a less definitive position due to the possibility that apparent empirical associations between income and health could be spurious and fail to capture any causal link. We hope to have contributed to this debate by testing if, indeed, the diseases of the poor are diseases of poverty. To the best of our knowledge, so far, our study is the first to employ tools of machine learning for analyzing granular data on African American and Caucasian populations from urban counties in the US and quantifying a causal link between urban life expectancy and household income. We infer that the causal effect of every percentage point increment in household income on life expectancy ranges from 0.10 to 0.198: these point elasticities fall between 0.10 and 0.142 for the African American population and between 0.145 and 0.198 for the Caucasian population from urban counties of the US.

References

Bloom, D.E. and Canning, D., 2000. The health and wealth of nations. Science, 287(5456), pp.1207-1209.

Cutler, D., Deaton, A. and Lleras-Muney, A., 2006. The determinants of mortality. Journal of economic perspectives, 20(3), pp.97-120.

Deaton, A., 2003. Health, inequality, and economic development. Journal of economic literature, 41(1), pp.113-158.

de la Escosura, L.P., 2023. Health, income, and the preston curve: A long view. Economics & Human Biology, 48, p.101212.

Jetter, M., Laudage, S. and Stadelmann, D., 2019. The intimate link between income levels and life expectancy: Global evidence from 213 years. Social Science Quarterly, 100(4), pp.1387-1403.

Kakwani, N., 1993. Performance in living standards: an international comparison. Journal of development economics, 41(2), pp.307-336.

Mackenbach, J.P. and Looman, C.W., 2013. Life expectancy and national income in Europe, 1900-2008: an update of Preston's analysis. International journal of epidemiology, 42(4), pp.1100-1110.

Preston, S.H., 1975. The changing relation between mortality and level of economic development. Population studies, 29(2), pp.231-248.

Preston, S.H., 2007. Response: On the changing relation between mortality and level of economic development. International journal of epidemiology, 36(3), pp.484-490.

Wilkinson, R.G., 2007. Commentary: The changing relation between mortality and income. International Journal of Epidemiology, 36(3), pp.492-494.