



California Center for Population Research
University of California - Los Angeles

Income Inequality and Population Health in Geographic Perspective

Michael E. Shin

CCPR-028-06

December 2006

*California Center for Population Research
On-Line Working Paper Series*

Income Inequality and Population Health in Geographic Perspective

Michael E. Shin
Department of Geography
University of California, Los Angeles
1255 Bunche Hall
Box 951524
Los Angeles, California 90025-1524
t. 310.825.3727
f. 310.206.5976
shinm@geog.ucla.edu

Abstract

A considerable amount of research indicates that relative income is related to health, or more specifically, that greater disparities in income are related to poorer health. Referred to as the relative income hypothesis, the causal mechanisms, social implications and general nature of this relationship continue to stimulate interesting and important research across several disciplines. One dimension of the association between income inequality and health that has yet to be explored or considered fully is the geographical. In an attempt to elucidate and evaluate how geography is implicated with the relative income hypothesis, the role and significance of geographic spillovers, regional effects and diffusion are theorized and explored. This research uses a geographic framework, spatial analytic techniques, and current and imputed data to revisit and re-examine the linkage between relative income and health at the global scale. Results indicate that the association between income inequality and health is somewhat tempered when examined within an explicitly spatial framework, and that regional and geographic effects need to be considered and treated carefully in cross-sectional, international studies of health.

Keywords: income inequality, life expectancy, geographic spillovers, diffusion, spatial analysis

1. Introduction

Income inequality is one of the predominant features of global society. Moreover, income inequalities between and within countries appear to be increasing (e.g., see Jones, 1997; Milanovic, 1999; Prichitt, 1997; Wade, 2004). One of the more perplexing social costs related to income inequality is its impact upon health, or more specifically, the finding that high levels of income inequality are associated with lower life expectancy, and higher infant and childhood mortality rates throughout the world (Rodgers, 1979; Waldmann, 1992; Wilkinson, 1996). The relationship between income inequality and health, referred to as the 'relative income hypothesis', remains robust even when absolute levels of income are controlled and when making comparisons between and within countries (Ben-Slomo, White, & Marmot, 1996; Kennedy, Kawachi, & Prothrow-Stith, 1996).

One dimension of the relative income hypothesis that has not been thoroughly explored or examined is the geographical (but see Gatrell, 1997; MacIntyre, MacIver, & Sooman, 1993). This is a notable oversight because the geographies of health (e.g., life expectancy, infant mortality, access to doctors and drugs, etc.) are neither evenly nor randomly distributed around the world, within countries or even across cities. By exploring the relative income hypothesis in geographic perspective, the relationship between income inequality and health is arguably clarified.

Health, in its many and competing conceptualizations, is not only a function of individual attributes, but is also related to where one lives. For example, the combinations and conjunctures of tropical environments, which tend to harbor and promote infectious disease, economic underdevelopment, authoritarian regimes and civil and inter-state conflict seriously challenge the population health of many countries in sub-Saharan Africa (e.g., see Collier, Elliot, Hegre, Hoeffler, Reynal-Querol, & Sambanis, 2003). Conversely, Western

Europe is home to economically developed and democratic countries that enjoy elevated and relatively stable standards of living and health. Such social, economic and health challenges and privileges are not necessarily contained or circumscribed by international borders; rather, such patterns and the processes that generate them are often regional in nature and extent.

This paper uses a geographic framework to examine the relationships between income inequality and health. First, the theoretical bases of the relative income hypothesis and the importance of diffusion and geography are discussed. Second, issues about missing data and international comparisons are presented and followed by exploratory spatial analyses of income inequality and health. The spatial analyses inform the final part of the paper that evaluates the linkage between income inequality and health using spatial econometric techniques. In addition to elucidating the relations between wealth, inequality and health, this paper calls for a more comprehensive treatment of geography when examining health from a social science perspective.

2. Theoretical considerations

Over the course of the last century, levels of economic prosperity and life expectancy increased dramatically and concurrently around the world. Increases in life expectancy can partially be attributed to the reduction of death rates among the young, or the fact that more people reach older ages now than in any other period in history. Though increases in absolute income are undoubtedly linked to better health, especially in developing countries, it is interesting to note that the beneficial returns on health appear to diminish after a per capita income of approximately \$5,000 is achieved (see Figure 1). In fact, after this \$5,000 threshold is surpassed, the relationship between life expectancy and wealth is unclear.

Figure 1 here: Wealth and health.

For example, life expectancy in the United States, the richest country in the world in terms of per capita income at the turn of the new century, falls short of that for several countries such as Japan and Italy, and infant mortality rates in America are higher than those in the Czech Republic (World Bank, 2003). It may appear that the limits to human longevity have been reached in many developed countries, but the rate of increase in life expectancy (i.e., 2-3 years per decade) remained remarkably stable over the course of the twentieth century (Wilkinson, 1996).

Country to country variations in life expectancy and infant mortality are frequently understood in terms of the 'epidemiological transition model' (see Omran, 1971, 1983; Wilkinson, 1994). This model documents a given society's pathway to the reduction of death rates, particularly those deaths associated with communicable diseases. Demographic characteristics of countries that have yet to complete the epidemiological transition include: high levels of mortality primarily due to infectious diseases, high levels of fertility and a predominantly young population which is especially vulnerable to infectious disease. Once the detrimental effects of communicable disease can be overcome by a society, mortality and fertility rates gradually decrease and stabilize, the population ages and non-communicable diseases such as heart disease, stroke, obesity and cancer become more common as causes of death (Wilkinson, 1996). The former phase of the epidemiological transition is also referred to as the 'infectious disease model', and the latter the 'chronic disease model'. Note that the geographic and temporal manifestations of the epidemiological transition are not necessarily identical in different places nor are they consistent over time (Curtis, 2004; Jones & Moon, 1992).

The epidemiological transition arguably marks an important stage in the process of economic development, and in addition to the advantageous effects of increases in absolute income are the effects that the distribution of income may have upon health. To illustrate this point, the theoretical relationship (f) between income and life expectancy is diagrammed in Figure 2.

Figure 2 here. The relationship between income, income inequality and health. The concave shape of the curve provides a general description of the how health (i.e., life expectancy) tends to diminish with subsequent increases in income. It is clear from Figure 2 that increases in absolute income benefit health, especially for those with lower incomes, but given the curvilinear nature of this relationship, life expectancy also depends upon the overall distribution of income. Consider the dispersion of income between x_1 and x_4 , with a mean income of \bar{x} and a mean life expectancy of \bar{y}_1 . Reducing this dispersion by increasing x_1 to x_2 , and decreasing x_4 to x_3 , while holding mean income constant, increases mean life expectancy from \bar{y}_1 to \bar{y}_2 .

There is evidence that supports the relative income hypothesis when making international comparisons (e.g., Le Grand, 1987; van Doorslaer, Wagstaff, Bleichrodt, Calonge, Gerdtham, & Gerfin, 1997; Wilkinson, 1996), comparisons between less developed countries (e.g., Flegg, 1982; Rodgers, 1979), and comparisons between developed economies (e.g., Duleep, 1995; Wennemo, 1993). Precisely how income inequality leads to poorer health continues to be a subject of debate. One explanation is based on the argument that inequality leads to problematic variations in access to health care, education and social services. Another explanation is based upon the argument that income polarization strains the beneficial components of social capital (e.g., cooperative interpersonal ties, reciprocal

social networks, etc.), which in turn adversely affects population health (Kawachi & Kennedy, 2002; Kawachi, Kennedy, Lochner, & Prothrow-Stith, 1997).

The relationship between income inequality and health has also been subject to much scrutiny (e.g., Gravelle, 2002; Judge, 1995). Some skeptics view the relationship to be a statistical artifact arising from the ecological fallacy. Since income inequality is a societal attribute and not an individual one, inferring health outcomes from aggregate data is inherently problematic (e.g., Gravelle, 1998). Despite such concerns and reservations, the potential health, social, and policy implications of the linkages between income, income inequality and health warrant further study.

This examination of the relative income hypothesis builds upon prior research by re-examining the relationship between income inequality by formally evaluating the role of geography and by using current and imputed data. The three questions that motivate this research are: (1) Using current and complete data, is there a significant relationship between income inequality and health?; (2) Is the relationship between income inequality and health manifest differently in different regions of the world?; and, (3) To what extent are regional effects influencing the relationship between income inequality and health. By framing geographically the linkage between income inequality and health, important insights into the relative income hypothesis are obtained.

2.1 Health mapping: Spatial dependencies and heterogeneities

Global patterns of health, such as those illustrated in Figure 3, are well recognized and familiar. When measured in terms of life expectancy and infant mortality rates, poor health persists in Sub-Sahara Africa and parts of South Asia, moderate levels of health exist in the Middle East and Latin America, and better health is found across North America and

Europe. What is clear from Figure 3 is that global patterns in health, among other things, are not random nor are they evenly distributed. When considering intraregional (e.g., Europe, Sub-Saharan Africa, etc.) health patterns, variations tend to be relatively small. In other words, similar life expectancies and infant mortality rates are concentrated or clustered geographically. Such geographic clustering of a phenomenon is formally referred to as spatial autocorrelation or spatial dependence. There are also notable inter-regional differences in health, and these geographic variations are referred to as spatial heterogeneity.

Figure 3 here. World life expectancies and infant mortality rates.

The presence of spatial autocorrelation is suggestive of diffusive processes and geographic spillovers, and spatial heterogeneity indicates that such processes and spillovers may operate and occur differently in different places. Diffusion generally refers to how a phenomenon spreads over space and time. Commonly associated with the spread of disease, diffusion is frequently viewed to be a function of proximity. For instance, it is often through direct or immediate contact with an infected agent that a disease such as HIV/AIDS, severe acute respiratory syndrome (SARS) or tuberculosis spreads. The term spatial contagion is frequently used to describe the process of diffusion at the local scale. When things spread from one location to another without affecting places in between, it is referred to as relocation diffusion. SARS provides a recent example of both types of diffusion; the disease was transmitted between individuals via spatial contagion within places like Hong Kong and Singapore, but relocation diffusion describes how the disease ‘traveled’ from Southeast Asia to Europe and Canada while bypassing most of Africa (Shin, 2004).

In addition to infectious diseases, several other social phenomena spread via diffusion and to a certain degree depend upon geographic proximity. Technological innovation, levels of international trade and economic growth (e.g., Frankel, 1997; Shin,

2002), democracy and democratization (e.g., O'Loughlin, Ward, Cohen, Brown, Reilly, Gleditsch et al., 1998), and the outbreak of international conflict (e.g., Siverson & Starr, 1991; Vasquez, 1995) not only diffuse, but are arguably among the most relevant social processes with regard to population health and economic development.

For example, in an examination of civil wars, Collier et al. (2003, 33-49) illustrate the numerous repercussions or spillovers that such conflicts create for neighboring states. Among the economic spillovers are: the accommodation of refugees, many of whom are poor and have few alternative destinations; the increased cost of transporting goods due to safety concerns; the illegal trafficking of goods and narcotics to fund war efforts; decreased levels of foreign investment because of regional insecurity; and, pressures to increase military expenditures. Note that military spending often comes at the expense of social spending, it increases the likelihood of a regional arms race (i.e., continued and greater military expenditures) and in some cases it increases the risk of international conflict, all of which can have lasting negative health consequences.

The most significant social spillovers linked to civil war are those related to health. Conflict and war displaces large numbers of people and their large-scale movement, often from cities and towns to unfamiliar rural locations and desolate border areas can put enormous pressure to provide services that are scarce or unavailable (e.g., shelter, medical, access to clean water, etc.). Moreover, in areas where infectious disease and illness is endemic, forced migrations increase the likelihood of contracting and spreading sickness to non-infected populations and unaffected areas. Those displaced by war are also more prone to contracting HIV/AIDS because of increased poverty, disrupted familial and social networks and increased levels of sexual violence. Ghobarah et al. (2003) indicate that HIV/AIDS is the single most significant spillover of civil war for neighboring countries,

especially for the young. Finally, civil war itself is often a precursor to violent international conflict (e.g., the Balkans, the African Great Lakes region), which can perpetuate the spillovers identified above, and that can negatively impact the population health of an entire region.

Geographic spillovers, however, can also be positive. For example, regional institutions such as the European Union not only promote trade and economic growth, but can also provide the financial and organizational means to improve health and to introduce health improvement policies in member states. Similarly, development strategies such as those sponsored by the United Nations Development Program, are often targeted toward improving regional health through education, immunization and medical assistance. Political agreements and arrangements such as treaties and accords that minimize and prevent conflict can also have positive and beneficial impacts upon regional population health. Regardless of the nature of such spillovers, geographic relationships exist and matter, therefore, geography needs to be acknowledged and evaluated whenever possible.

Frequently used population health statistics, such as those mapped in Figure 3, arguably reflect and capture many of the spillovers described above. In many respects, especially in this era of globalization and unprecedented economic, cultural, social and political interdependence, this is to be expected. Yet in all international examinations of the relationship between income inequality and health, countries are treated as independent units of analysis as probability theory necessitates. This is potentially problematic because the dependent variable (i.e., life expectancy, infant mortality rates) is clearly not independent. The consequences of disregarding such spatial relationships within the context of the relative income hypothesis are thus explored using spatial analysis and spatial regression techniques.

3. Data, Missing Data & Spatial Methodologies

A. Data and Missing Data. Life expectancy, and somewhat paradoxically, mortality rates prove to be concise and reliable measures of population health in light of the difficulties in determining, defining and gauging the concept of health. Several studies find that a robust statistical relationship exists between income inequality and a variety of population health statistics (e.g., Rodgers, 1979; Waldmann, 1992). In this study, 2001 life expectancy data are used to measure health and are taken from the 2003 World Development Indicators compiled by the World Bank. Life expectancy at birth is defined as the number of years a newborn would live if prevailing patterns of mortality remained constant throughout its life.

Income inequality data in the form of the Gini index are also drawn from the World Bank's 2003 World Development Indicators. Derived from the Lorenz curve that portrays graphically the cumulative shares of income at successive levels, the value of the Gini index falls between zero and one, the former indicating perfect equality and the latter perfect inequality. Though other measures of inequality exist, and are evaluated with respect to the relative income hypothesis elsewhere (see Kawachi & Kennedy, 1997), the Gini index is the statistic used most frequently in income inequality and health research and is correlated strongly with other such measures.

The use of the Gini index, however, presents several comparability issues (Deininger & Squire, 1996; Milanovic, 1999; Wade, 2004). Inequality measures such as the Gini index are often based upon household surveys that are difficult to adjust and to standardize over time and between countries. In order to provide the most complete set of inequality measures, the World Bank reports Gini index values based on both income and consumption, the former tending to capture greater levels of inequality. Additionally, variations in survey years, household size and levels of income make strict country-to-

country comparisons difficult. Despite these issues, recent efforts to update and to improve comparability make these data, at the very least, as valid and as reliable as those used in previous studies of the relative income hypothesis. Finally, to control for the effects of absolute income levels, 2000 real GDP per capita, in constant 1996 dollars, from the Penn World Tables (v.6.1) are used (Heston, Summers, & Aten, 2002).

Though several data sets are available for making international comparisons and analyses, missing data are common. The problem of missing data is frequently compounded when combining data from different sources, historical periods and geographical regions. When a data item or value is missing it is common to omit the entire observation from the analysis. This practice is referred to as listwise deletion. The omission of an entire observation, perhaps due to a single missing variable, is not only a loss of information, but it can result in severe selection bias and biased conclusions (i.e., over- and under-estimated statistical significance) (King, Honaker, Joseph, & Scheve, 2001).

To combat the problem of missing data this analysis uses a multiple imputation algorithm to predict missing values (see King et al. (2001) for details). The basic strategy behind this technique is to learn something about the missing data (i.e., predict missing values) by using and analyzing the traits of the observed data. Assuming the data are jointly multivariate normal, missing values are imputed linearly. This assumption is robust and holds up well when compared to more complex alternatives (Schafer & Olsen, 1998). For each missing item in a data set, five separate values are imputed, thus creating five complete data sets. Within each of these five data sets, the observed values are identical, but the imputed values are different and reflect uncertainty levels or what is known about the missing data. Determining a statistic of interest (e.g., variance, mean, a regression

coefficient, etc.), involves calculating the statistic using each of the five complete data sets and effectively averaging the five results.

After reviewing and comparing existing data for completeness and geographical coverage, the five final imputed data sets each contain 153 observations. For comparison, Rodgers' (1979) analysis of the relationship between income inequality and health uses a total of 53 observations, and more recent examinations of the relative income hypotheses tend to use small samples (i.e., < 20) that focus solely on developed countries (Judge, Mulligan, & Benzeval, 1998). Table 1 provides descriptive summaries for the key observed variables and for corresponding variables from the five imputed data sets.

Table 1 here. Descriptive summaries of observed and imputed data sets.

Though there are differences within and between each data set, it appears that the imputed values are both reasonable and realistic when compared to the original observed values. The degree of missingness ranged from a minimum of three values for 2001 life expectancy to a maximum of 34 missing values for real GDP. Note that the GDP values are logged in order to meet the linearity assumption of the imputation model. Additional variables (i.e., 1980 GDP, 1990 GDP, 1980 life expectancy, 1980 and 2000 infant mortality rates, and upper- and lower-decile income data) are also included in the imputation models to improve predictions of missing values (see Honaker, Joseph, King, Scheve, & Singh, 2003). The subsequent regression and spatial analyses are based upon these five completed data sets.

B. Methodology and results. Most examinations of the relationship between income inequality and health use ordinary least squares (OLS) regression, and follow the general specification,

$$y = \alpha + \beta X + \varepsilon \quad (1)$$

where the dependent variable y is life expectancy, α is the intercept term, β contains the parameter estimates for X , the matrix of independent variables, and ε is the error term distributed normally and independently (i.e., $\varepsilon_i \sim \text{NID}(0, \sigma^2)$). Table 2 provides two sets of estimates for comparison. The first five columns contain parameter estimates taken directly from Rodgers (1979), and the second five columns are from replications using the current data and the multiple imputation algorithm for missing values. Results from all of the replications are strikingly similar to those in Rodgers (1979), with all models returning high R^2 values. To capture the diminishing relationship between absolute income and health, it is specified in reciprocal form. Across all equations, measures of absolute and relative income are consistently significant and negatively associated with life expectancy, except for the reciprocal of squared-income when included. While the magnitude of absolute income estimates and intercepts reflect clearly the overall increase in levels of world income and life expectancy over the last twenty-five years, the degree of the association between relative income and health appears to have remained relatively stable.

Table 2 here. OLS estimates of the relative income hypothesis.

Regression diagnostics (not reported) indicate that the error terms from all replications in Table 2 exhibit a significant amount of heteroscedasticity, which is often the by-product of non-random geographic effects such as spatial dependence. In regression analysis, if the dependent variable exhibits spatial dependence, parameter estimates will be biased and inferences based upon the parameter estimates will thus be incorrect (Anselin, 1995b). Similarly, if the error term is contaminated by spatial dependence, thus violating the assumptions of independence and often normality, parameter estimates will be inefficient. This inefficiency can result in misleading assessments of t and F statistics, and evaluations of a model's fit using R^2 will also be incorrect.

To evaluate the degree of spatial dependence across the imputed datasets, the local Moran's I statistic is calculated for each variable of interest. Expressed formally in matrix notation:

$$I = \frac{y'Wy}{y'y} \quad (2)$$

where y is the standardized variable of interest and W is a row-standardized spatial weights matrix that summarizes the geographic relations between observations, the local Moran's I statistic provides a general indication of whether or not a variable is spatially clustered. Values of Moran's I fall between -1 (i.e., perfect negative spatial autocorrelation), which indicates a chessboard pattern of dissimilarity to $+1$ (i.e., perfect positive spatial autocorrelation), which indicates that the value of one record can effectively be predicted by averaging the values found in neighboring observations. The value of Moran's I is based upon the assumption of randomness, or that each value is equally likely to occur at all locations, and involves calculating a standard deviate (Anselin, 1995a).

The geographic relations summarized by the spatial weights matrix W are central to measuring correctly levels of spatial autocorrelation. First, a geographical information system (GIS) is used to fill the spatial weights matrix with ones and zeros, the former indicating contiguity or adjacency between countries and the latter denoting separation. Adjacency, or a shared international border on the 2003 world political map, is used to define proximity in this analysis. Distance thresholds can also be used to create such matrices (see Gleditsch & Ward, 2001), but adjacency provides a relevant and succinct profile of international geographic relations and the opportunities for diffusion and geographic spillovers. The weights matrix is then row-standardized (i.e., each element of the matrix is divided by its row sum) which corresponds to spatial averaging. Multiplying W and

y returns what is called the spatial lag of y (i.e., Wj), and is the weighted average of adjacent values for each observation.

Since Moran's I depends upon the values of proximate observations as captured by the spatial weights matrix W , it and other spatial statistics are particularly sensitive to listwise deletion and the problems associated with missing data. When data are incomplete or when cases are omitted, which is frequently the case in international cross-sectional studies, spatial statistics will be biased. By using the multiple imputation method to provide estimates for missing data, more reliable and accurate measures of spatial autocorrelation are obtained. Table 3 provides local Moran's I statistics for all variables used in this and subsequent analyses.

Table 3 here. Moran's I statistics.

All variables of interest exhibit high levels of positive spatial autocorrelation and all are statistically significant (i.e., $p < 0.01$). Life expectancy, the dependent variable in this analysis, contains the highest degree of geographic clustering followed respectively by the log of per capita income, the reciprocal of per capita income and the Gini index of income inequality. In addition to detecting the presence of spatial dependence across a dataset, the individual components that constitute the local Moran's I statistic can be plotted and mapped. Figure 4 provides the Moran scatterplot for 2001 life expectancy. Note that the data used in the scatterplot are from the second imputation. Each quadrant in the scatterplot contains a different type of spatial association: high values surrounded by high values are found in the upper-right quadrant, low surrounded by low are in the lower-left quadrant; high surrounded by low are in the lower-right; and, low surrounded by high are in the upper-left quadrant. The slope of the regression line passing through the cloud of data points is equivalent to Moran's I.

Figure 4 here. Moran's scatterplot for 2001 life expectancy.

The clustering of similar values is clearly portrayed by the strong positive relationship between life expectancy and the spatial lag of life expectancy. However, countries with high life expectancy surrounded by low life expectancy (lower-right quadrant) include: Algeria, China, Thailand, Uzbekistan, Libya, Tajikistan, Vietnam, South Korea, Turkmenistan, the Dominican Republic and Iran; and countries with low life expectancy surrounded by high life expectancy include: Yemen, Nepal, Bolivia, Iraq, Afghanistan, North Korea, Haiti, Papua New Guinea, Kazakhstan and Guyana. Of particular interest is the cluster of low values surrounded by similarly low values in the lower-left quadrant. Unsurprisingly, the countries of Sub-Saharan Africa comprise this cluster of points. This pattern is reinforced when significant Moran's I values of positive spatial dependence are mapped in Figure 5. Western Europe emerges as a cluster of high life expectancy while Sub-Saharan Africa is characterized by low life expectancy.

Figure 5 here. Spatial clustering of high and low 2001 life expectancies.

Though the above patterns are familiar, the purpose of these spatial analyses is to evaluate formally patterns of wealth, inequality and health with statistical measures of spatial dependence. Of particular interest are the levels of spatial dependence associated with life expectancy, the dependent variable in this and several other analyses of the relationship between income inequality and health. The results reinforce and support the two key arguments that: (1) diffusive processes and geographical spillovers create detectable intra-regional similarities, and (2) such processes arguably work differently around the world thus creating inter-regional variations. Failure to recognize and to control for local similarities and regional differences is likely to result in biased results and incorrect conclusions.

Though many studies of the relative income hypothesis concentrate solely on economically

developed countries (e.g., OECD countries), spatial autocorrelation may still contaminate the results. Recognizing the geographic structure and characteristics of the data, the previously estimated models of the relationship between income inequality and health are re-specified and re-estimated.

Due to the notable regional differences in life expectancy and the level of geographic clustering, a spatial regimes/spatial error approach to modeling the relative income hypothesis is taken. Formally expressed:

$$y = \alpha_1 + \beta_1 X + \varepsilon \quad \text{for } d = 0 \quad (3)$$

$$y = \alpha_2 + \beta_2 X + \varepsilon \quad \text{for } d = 1 \quad (4)$$

$$\varepsilon = \lambda W\varepsilon + u \quad (5)$$

this specification determines whether or not the relative income hypothesis operates similarly or differently in different regions of the world, and also introduces a spatial constraint for the error term. Guided by the previous spatial analyses and for illustrative purposes, the data are divided into two subsets, Sub-Sahara Africa (i.e., $d = 1$) and the Rest of World (i.e., $d = 0$), and estimates are obtained for each geographic domain. Though the regionalization scheme is simple, it is an improvement over the tendency to disregard completely geographic effects when examining the relationship between income inequality and health. Though some studies put countries into categories such as developed and underdeveloped, the spatial relationships between countries is seldom captured. To control for such spatial relations (i.e., spatial dependence) in the residuals, the spatial lag of the error term, $W\varepsilon$, is also incorporated into the regression equation, λ is the corresponding parameter estimate and u is the error term normally and independently distributed. Table 4 provides the maximum likelihood estimates and relevant diagnostics for the best-fitting models in Table 2, controlling for the geographic effects described above.

Table 4 here. Spatial regime/spatial error models, maximum likelihood estimation.

Notwithstanding its functional form, the reciprocal of absolute income is consistently significant across equations and regions. The Gini index is negatively related to life expectancy and is consistently significant only in Sub-Saharan Africa. Note that its magnitude in Sub-Saharan Africa remains relatively stable across the three estimations. Across the remaining countries (i.e., Rest of World), the relative income measure is also negatively related to health but only significant (at the $p < 0.05$ level) in the second regression equation. The size and significance of λ , the spatial error parameter, confirm that spatial effects are indeed present within the error term and need to be controlled. When compared to the OLS estimates in Table 2, the inclusion of the spatial constraint on the error term appears to temper the relationship between income inequality and health by about one-quarter to over one-half.

The regression diagnostics reported for the spatial regime/spatial error models also return some interesting results. First, the Chow tests for structural instability indicate that there is indeed a significant difference between the two regions, and that this difference is in large part determined by differences in levels of absolute income (and are also reflected in the intercept values, which approximate current life expectancy values for each region in the second and third set of equations). Second, though a casual examination of parameter estimates suggest that the relationship between income inequality and health operates differently in Sub-Saharan Africa than elsewhere, Chow tests on the stability of parameters across regions indicate otherwise, or that there is no statistically significant difference between the income inequality estimates across the two subsets. Fourth, the insignificant LaGrange Multiplier test for spatial autocorrelation indicates that the inclusion of the spatial constraint on the error term successfully controlled spatial dependence within the residuals

that contaminated previous results. Finally, the likelihood ratios indicate that the third and final spatial regime/spatial error regression returns the best fit using updated and imputed data.

The results from these analyses perhaps present more questions than answers with regard to the relationship between income inequality and health. Despite the fact that the relative income hypothesis appears to be rather robust in Sub-Saharan Africa in this study, most analyses of this association focus upon economically developed and advanced economies (Judge, Mulligan, & Benzeval, 1998). Though increases in absolute income are probably more important to health in less developed countries than are reductions in income inequalities, the societal impact of such disparities may be a significant piece of the population health and development puzzle. Hence, there is a need to examine further the linkage between income inequality and health outside of the developed world.

These analyses also underscore the importance and utility of a geographic perspective and methodology. Though population health is usually measured and understood on a country-by-country basis, regional circumstances play a very important role with respect to the opportunities for better or poorer health, and need to be considered and explored further. The methodological pitfalls and theoretical implications of ignoring intra-regional similarities and inter-regional differences are too significant to disregard. Similarly, and perhaps not stressed enough given the cross-sectional nature of this analysis, is the fact that geographic regions are not static but are dynamic and fluid entities that change over time. Recognizing the possibility and occurrence of social, political, economic and cultural spillovers within and between such regions warrants a temporal and historical perspective as well. Situating future examinations of the relative income hypothesis within a geographic

framework will ultimately help to clarify and extend our understanding of the linkages between wealth, inequality and health.

4. Conclusions

These analyses both support and challenge previous work on the relative income hypothesis, and highlight the utility and necessity of a geographic framework when examining the relative income hypothesis. Few studies acknowledge the geographic dimensions of income inequality and health or the problems associated with missing data. There are two important implications of this oversight. First, in methodological terms, the practice of listwise deletion and the failure to test and to control for spatial effects, which are present in most cross-sectional data sets, can lead to incorrect model specification, biased estimates and incorrect inferences. The exploratory spatial techniques confirm that levels of income, income inequality and health in one country are strongly correlated to levels found in surrounding countries. Second, disregarding geographic processes and spillovers is analogous to assuming that the world consists of autarkic, isolated and solitary countries. A geographic approach helps to clarify and identify regional obstacles and pathways to better health. For example, though mortality rates are most frequently reported by country, regional environmental, political, economic and social factors arguably play an important role in elevating or depressing such rates.

The finding that income inequality may very well be linked to health provides a considerable amount of impetus to both academics and policy-makers to try to understand better the social consequences of inequality. However, as noted by Sen (1992: 88), “the evaluation of inequality has to take note of both the plurality of spaces in which inequality can be assessed, and the diversity of individuals”. By situating questions and analyses about

inequality and health within geographic context, ranging from the global to the local, our understanding of the relative income hypothesis will ultimately be expanded.

5 References

- Anselin, L. (1995a). Local indicators of spatial association - LISA. *Geographical Analysis*, 27, 93-115.
- Anselin, L. (1995b). *SpaceStat, version 1.80 user's guide*. Morgantown, WV: Regional Research Institute
- Bank, W. (2003). *World Development Indicators 2003* Washington, D.C.: The World Bank
- Ben-Slomo, Y., White, I., & Marmot, M. (1996). Does the variation in the socioeconomic characteristics of an area affect mortality. *British Medical Journal*, 312, 1013-1014.
- Collier, P., Elliot, L., Hegre, H., Hoeffler, A., Reynal-Querol, M., & Sambanis, N. (2003). *Breaking the Conflict Trap* Washington, D.C.: The World Bank
- Curtis, S. (2004). *Health and Inequality* London: Sage
- Deininger, K., & Squire, L. (1996). A new data set measuring income inequality. *World Bank Economic Review*, 10, 565-591.
- Duleep, H. (1995). Mortality and income inequality among economically developed countries. *Social Security Bulletin*, 58, 35-50.
- Flegg, A. (1982). Inequality of income, illiteracy and medical care as determinants of infant mortality in underdeveloped countries. *Population Studies*, 36, 441-458.
- Frankel, J. (1997). *Regional Trading Blocs in the world economic system* Washington, D.C.: Institute for International Economics
- Gatrell, A.C. (1997). Structures of Geographical and Social Space and Their Consequences for Human Health. *Geografiska Annaler, Series B, Human Geography*, 79, 141-154.
- Ghobarah, H., Huth, P., & Russett, B. (2003). Civil Wars Kill and Maim People - Long After the Shooting Stops. *American Political Science Review*, 97(2).

- Gleditsch, K.S., & Ward, M.D. (2001). Measuring Space: A Minimum Distance Database and Applications to International Studies. *Journal of Peace Research*, 38, 739-758.
- Gravelle, H. (1998). How much of the relation between population mortality and unequal distribution of income is a statistical artefact? *British Medical Journal*, 316, 382-385.
- Gravelle, H. (2002). Income, income inequality and health: what can we learn from aggregate data. *Social Science and Medicine*, 54, 577-589.
- Heston, A., Summers, R., & Aten, B. (2002). Penn World Table Version 6.1: Center for International Comparisons at the University of Pennsylvania (CICUP).
- Honaker, J., Joseph, A., King, G., Scheve, K., & Singh, N. (2003). *Amelia: A Program for Missing Data* Cambridge: Department of Government, Harvard University
- Jones, C. (1997). On the Evolution of World Income Distribution. *Journal of Economic Perspectives*, 11, 19-36.
- Jones, K., & Moon, G. (1992). Medical Geography: global perspective. *Progress in Human Geography*, 16, 563-572.
- Judge, K. (1995). Income distribution and life expectancy: a critical appraisal. *British Medical Journal*, 311, 1282-1285.
- Judge, K., Mulligan, J., & Benzeval, M. (1998). Income Inequality and Population Health. *Social Science and Medicine*, 46, 567-579.
- Kawachi, I., & Kennedy, B. (1997). The relationship of income inequality to mortality: Does the choice of indicator matter? *Social Science and Medicine*, 45, 1121-1127.
- Kawachi, I., Kennedy, B., Lochner, K., & Prothrow-Stith, D. (1997). Social capital, income inequality and mortality. *American Journal of Public Health*, 87, 1491-1498.
- Kawachi, I., & Kennedy, B. (2002). *The Health of Nations: Why Inequality Is Harmful to Your Health* New York: The New Press

- Kennedy, B., Kawachi, I., & Prothrow-Stith, D. (1996). Income distribution and mortality: cross-sectional ecological study of the Robin Hood Index in the United States. *British Medical Journal*, 312, 1004-1007.
- King, G., Honaker, J., Joseph, A., & Scheve, K. (2001). Analyzing Incomplete Political Science Data: An Alternative Algorithm for Multiple Imputation. *American Political Science Review*, 95(1), 49-69.
- Le Grand, J. (1987). Inequalities in health. Some international comparisons. *European Economic Review*, 31, 182-191.
- MacIntyre, S., MacIver, S., & Sooman, A. (1993). Area, class and health: should we be focusing on places or people? *Journal of Social Policy*, 46, 26-34.
- Milanovic, B. (1999). True world income distribution, 1988 and 1993: First calculation based on household surveys alone., *World Bank, Development Research Group*. Washington, D.C.
- O'Loughlin, J.V., Ward, M.D., Cohen, J.S., Brown, D.S., Reilly, D., Gleditsch, K.S., & Shin, M.E. (1998). The diffusion of democracy, 1946-1994. *Annals of the Association of American Geographers*, 88, 545-574.
- Omran, A. (1971). The epidemiological transition: a theory of the epidemiology of population change. *Milbank Quarterly*, 64, 355-391.
- Omran, A. (1983). The epidemiological transition theory: a preliminary update. *Journal of Tropical Pediatrics*, 29, 305-316.
- Prichitt, L. (1997). Divergence, Big Time. *Journal of Economic Perspectives*, 11, 3-17.
- Rodgers, G. (1979). Income and Inequality as Determinants of Mortality: An International Cross-Section Analysis. *Population Studies*, 33, 343-351.
- Schafer, J., & Olsen, M. (1998). Multiple Imputation for Multivariate Missing-Data Problems: A Data Analyst's Perspective. *Multivariate Behavioral Research*, 33(4), 545-571.

- Shin, M.E. (2002). Measuring economic globalization: Spatial hierarchies and market topologies. *Environment and Planning A*, 34, 417-428.
- Shin, M.E. (2004). Globalization and Health. In E. Greenberg (Ed.), *Globalization and Its Outcomes* (pp. 193-208). New York: Guilford Press.
- Siverson, R., & Starr, H. (1991). *The Diffusion of War: A Study of Opportunity and Willingness*. Ann Arbor: University of Michigan Press
- van Doorslaer, E., Wagstaff, A., Bleichrodt, H., Calonge, S., Gerdtham, U., & Gerfin, M. (1997). Income-related inequalities in health: some international comparisons. *Journal of Health Economics*, 16, 93-122.
- Vasquez, J. (1995). Why Do Neighbors Fight? Proximity, Interaction, or Territoriality. *Journal of Peace Research*, 32(3), 277-293.
- Wade, R. (2004). Is Globalization Reducing Poverty and Inequality? *World Development*, 32(4), 567-589.
- Waldmann, R. (1992). Income distribution and infant mortality. *The Quarterly Journal of Economics*, 107, 1283-1302.
- Wennemo, I. (1993). Infant mortality, public policy and inequality - a comparison of 18 industrialised countries 1950-1985. *Sociology of Health and Illness*, 15, 429-446.
- Wilkinson, R. (1994). The epidemiological transition: from material scarcity to social disadvantage? *Daedalus*, 123, 61-77.
- Wilkinson, R. (1996). *Unhealthy Societies: The Afflictions of Inequality* London: Routledge

Table and Figure captions

Table 1. Descriptive summaries of observed and imputed data sets.

Table 2. OLS estimates of the relative income hypothesis.

Table 3. Moran's I statistics.

Table 4. Spatial regime/spatial error models, maximum likelihood estimation.

Figure 1. Wealth and health.

Figure 2. The relationship between income, income inequality and health.

Figure 3. World life expectancies and infant mortality rates.

Figure 4. Moran's scatterplot for 2001 life expectancy.

Figure 5. Spatial clustering of high and low 2001 life expectancies.

	<i>mean</i>	<i>variance</i>	<i>minimum</i>	<i>maximum</i>	<i>valid</i>	<i>missing</i>
	<i>Observed</i>					
<i>lifex</i>	64.09	163.44	37.00	81.00	150	3
<i>log GDP</i>	8.48	1.28	6.18	10.69	119	34
<i>Gini</i> ($\times 100$)	40.59	105.26	24.40	70.70	123	30
	<i>Imputation 1</i>					
<i>lifex</i>	64.19	162.83	37.00	81.00	153	0
<i>log GDP</i>	8.38	1.26	6.08	10.69	153	0
<i>Gini</i> ($\times 100$)	41.02	99.11	24.40	70.70	153	0
	<i>Imputation 2</i>					
<i>lifex</i>	64.22	163.76	37.00	87.00	153	0
<i>log GDP</i>	8.37	1.29	5.73	10.69	153	0
<i>Gini</i> ($\times 100$)	40.91	97.29	22.59	70.70	153	0
	<i>Imputation 3</i>					
<i>lifex</i>	64.25	166.02	37.00	92.00	153	0
<i>log GDP</i>	8.35	1.29	5.84	10.69	153	0
<i>Gini</i> ($\times 100$)	41.21	103.97	24.40	70.70	153	0
	<i>Imputation 4</i>					
<i>lifex</i>	64.07	163.49	37.00	81.00	153	0
<i>log GDP</i>	8.40	1.37	5.97	10.69	153	0
<i>Gini</i> ($\times 100$)	40.96	94.11	22.87	70.70	153	0
	<i>Imputation 5</i>					
<i>lifex</i>	64.22	163.71	37.00	87.00	153	0
<i>log GDP</i>	8.38	1.25	6.18	10.69	153	0
<i>Gini</i> ($\times 100$)	40.72	103.24	17.12	70.70	153	0

Table 1. Descriptive summaries of observed and imputed data sets.

	<i>Moran's I</i>	<i>z-score</i>
<i>Life expectancy</i>	0.772	13.09
<i>log Y</i>	0.622	10.41
$\frac{1}{Y}$	0.521	8.76
<i>Gini</i>	0.492	8.25

Table 3. Moran's I statistics.

	<i>Sub-Sabara Africa</i>	<i>Rest of World</i>	<i>Sub-Sabara Africa</i>	<i>Rest of World</i>	<i>Sub-Sabara Africa</i>	<i>Rest of World</i>
<i>constant</i>	77.1	123.7	57.2	79.6	61.1	80.5
$\frac{1}{Y}$			-3226.8* (1278.8)	-17,404 (1378.3)	-11,839 (3745.1)	-25,571.8 (3466.1)
$\frac{1}{Y^2}$					3,892,884* (1,622,141)	5,033,988* (2,321,429)
$\frac{1}{\log Y}$	-157.6 (49.4)	-439.1 (32.1)				
<i>Gini</i>	-19.68 (6.93)	-5.64 (4.81)	-18.55 (7.34)	-10.66* (5.05)	-18.79 (7.01)	-9.62 (4.83)
λ (residuals)	0.41 (0.09)		0.50 (0.08)		0.52 (0.08)	
			<i>Diagnostics</i>			
<i>Chow tests: overall</i>	149.9		187.1		131.8	
$\frac{1}{Y}$			59.5		19.2	
$\frac{1}{Y^2}$					8.1	
$\frac{1}{\log Y}$	23.7					
<i>Gini</i>	3.1		1.08		1.29	
<i>LaGrange Multiplier (spatial lag)</i>	0.34		1.05		0.91	
<i>Likelihood ratio</i>	-423.2		-430.6		-422.6	
<i>N</i>	153		153		153	

Standard errors are in parentheses; estimates in **boldface** are significant at the $p < 0.01$ level, asterisks (*) denote significance at the $p < 0.05$ level.

Table 4. Spatial regime/spatial error models, maximum likelihood estimation.

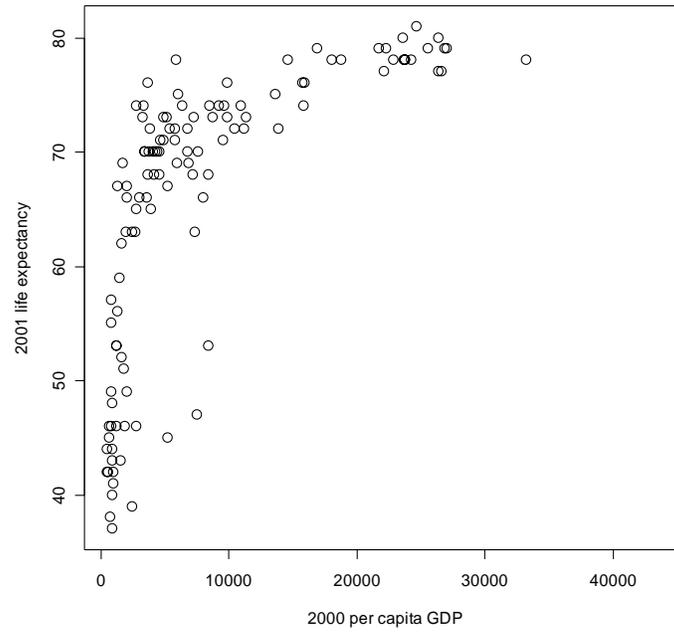


Figure 1. Wealth and health.

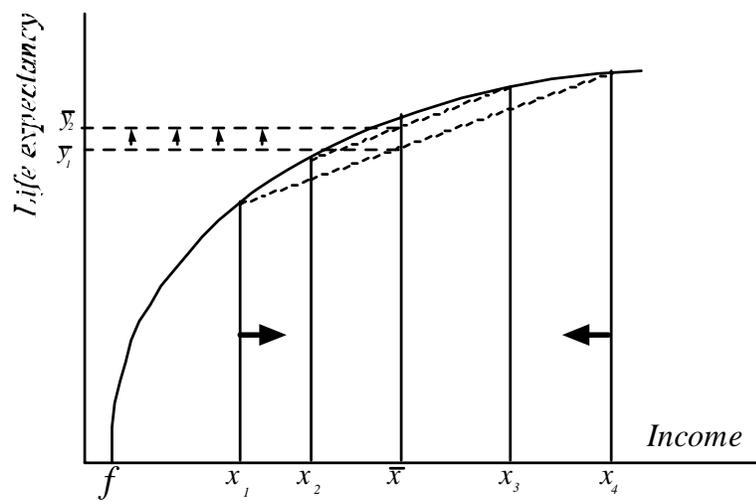


Figure 2. The relationship between income, income inequality and health.

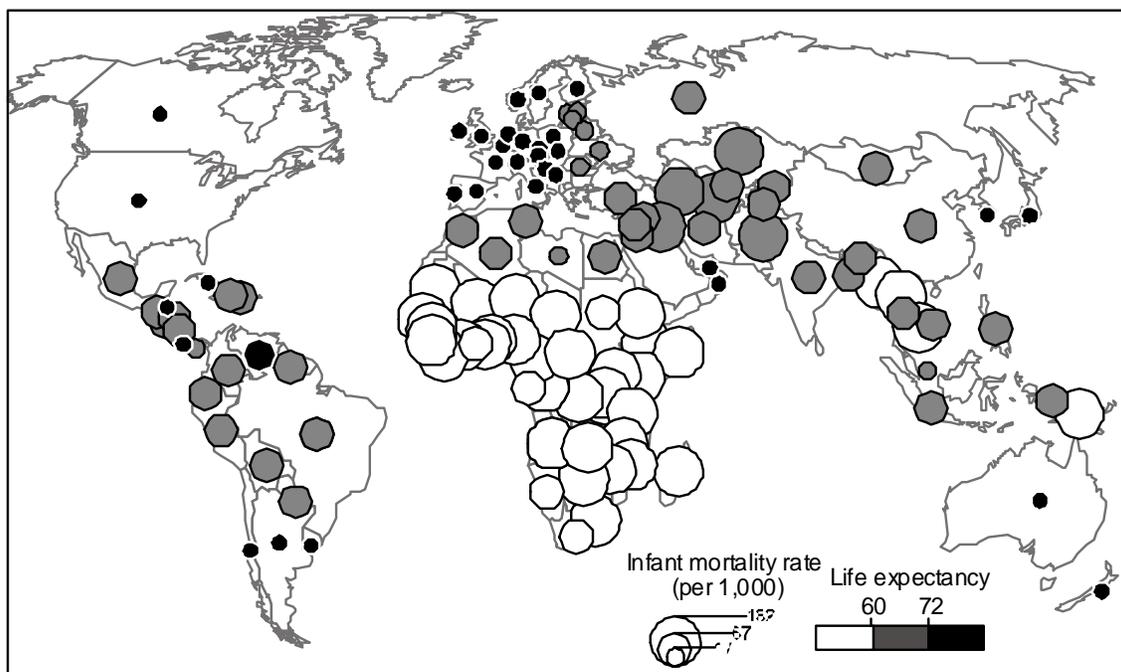


Figure 3. World life expectancies and infant mortality rates.

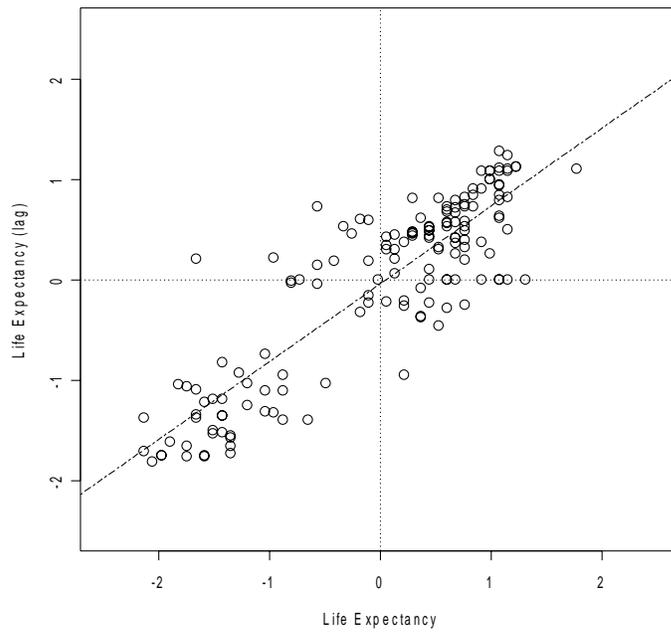


Figure 4. Moran scatterplot for 2001 life expectancy.

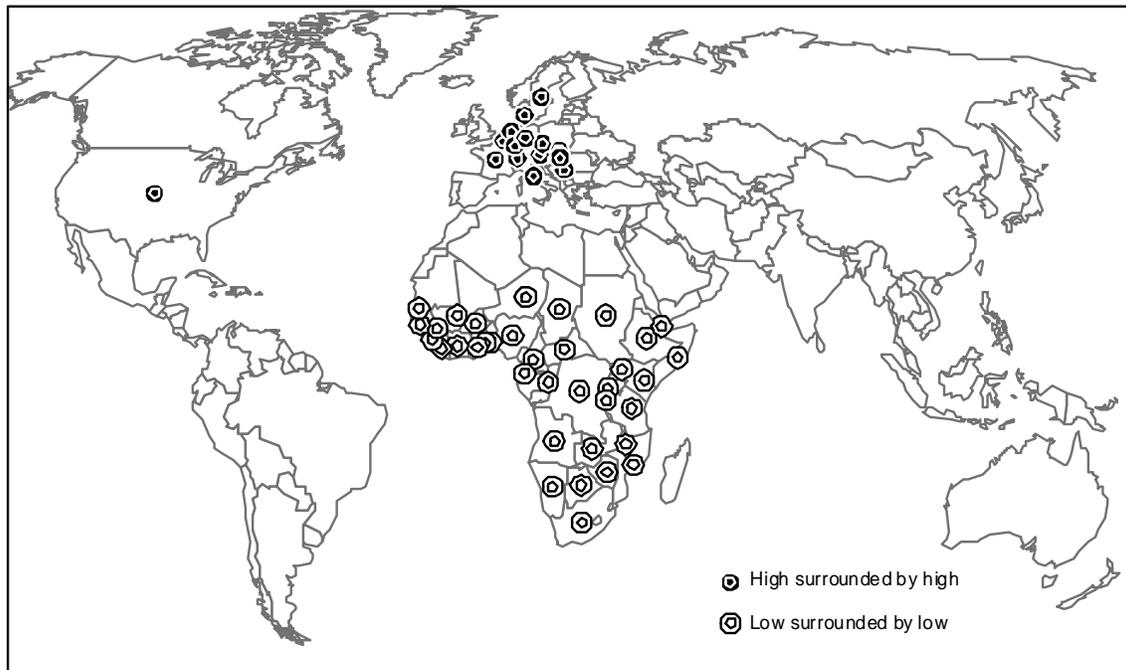


Figure 5. Spatial clustering of high and low 2001 life expectancies.