Long-Term Association Between Inequality and Mortality

Sepideh Modrek, PhD
smodrek@stanford.edu

Abstract

The hypothesis that economic inequality adversely affects health outcomes has been extensively debated in the economics, demography, and public health literature in recent decades. This study evaluates the relationship between economic inequality and mortality in the context of a middle-income country, Costa Rica, whose social structure and history confer the unique benefit of being less susceptible to common sources of confounding. Approximately 16,000 individuals aged 30 or more were selected from the 1984 Census and linked to the Costa Rican National Death Registry until Dec. 31, 2007. Gompertz models were used to estimate the relative risk of mortality for various indicators inequality, while controlling for area and individual-level confounders. The results were somewhat mixed across the two measures of economic inequality, area-level income inequality and relative deprivation, but in the preferred specification there was some evidence that more relatively deprived individuals exhibited lower survival rates over the following 19 years.

This research was supported by the National Institute on Aging training grant 5-T32-AG00246-13. All errors are my own. I am grateful to Dr. William Dow and Dr. Luis Rosero-Bixby for suggestions and technical guidance.

Background:

An enormous amount of literature documents the relationship between measures of socio-economic status (SES) (correlates of income/wealth) and mortality, as well as other health outcomes (N. E. Adler et al., 1994). This relationship reflects the absolute income hypothesis, which posits that income and SES correlates dictate the amount of resources available to buy better nutrition and healthcare treatments, and may affect the time available for health producing activities.[1] A more contentious hypothesis argues that an individual's relative rank or the dispersion of ranks within a community also directly affects health. This latter hypothesis, known as the relative income hypothesis, is the focus of this study.

While there are several formulations of this relative income hypothesis, supporters of the hypothesis generally contend that holding constant an individual's financial resources, greater economic inequality leads to poorer health (Richard G. Wilkinson, 1996). This hypothesis posits that people who feel more economically disadvantaged than their peers may experience chronic stress, which then leads to illness (Andrew Baum et al., 1999). Three general groups of pathways have been proposed to explain associations between health and income/wealth distribution: psychosocial, resource allocation, and neomaterial.

The psychosocial pathway has two main components. The first component posits of the possibility that the mere evidence of social hierarchy causes stress (Richard G. Wilkinson, 1996), which activates a neuroendocrine response that ultimately increases susceptibility to disease (B. McEwen, 2004). This view of the relationship between inequality and health is based on the direct extensions of primate studies to humans. For animals, the adaptiveness of the acute stress response is thought to be pathogenic when repeatedly and unnecessarily activated (Robert M. Sapolsky, 2004). Studies show that when incorrectly activated, the acute-stress response leads to higher circulating levels of glucocorticoids, which can lead to insulin (in)sensitivity, elevated blood pressures, atherosclerotic plaque, repressed reproduction and inhibited long-term immune response. The second component of the psychosocial mechanism posits that economic inequality weakens social ties within a community, which are used to buffer against stress and disease. The weakening of social ties is often conceptualized to affect health by limiting social capital. Social capital, defined as the level of interpersonal trust between citizens, norms of reciprocity, and vibrancy of civic associations within a community, is thought to be important because of its potential to help individuals within a community buffer against stress and consequent disease (I. Kawachi et al., 1997).

Moving beyond these psychosocial pathways, others argue that inequality affects health through resource allocation. Some argue that economic inequality is in fact a proxy for greater poverty in an area, and that this poverty or lack of resources is what impacts health (A. Deaton, 2003). Others posit that the greater the inequality within a community, the more infrastructural and political interests might diverge. For example, a family at the 95th income percentile may pay substantially more taxes than a family at the 50th percentile, but does not receive a correspondingly higher benefit from public services, such as public health services. If these two families live in the same community, this

[1] Economists also argue that individuals with better health earn higher incomes, and thus, the relationship between income and health is likely to work in both directions.

income gap creates pressure from economic elites to reduce local public services, including those related to health. In contrast, in more equal areas, economic interests may be more aligned and investments in human capital infrastructure represent a more common interest (Paul Krugman, 1996). These types of mechanism could similarly explain how the lack of public resources in relatively unequal areas causes poor health outcomes.

Finally, the neomaterial view of the inequality and health relationship contends that the underlying contextual process that generates inequality is also what causes poor health outcomes. Under this interpretation, the effect of income inequality on health reflects a combination and accumulation of negative exposures, lack of individual resources and systematic underinvestment in health and social infrastructure. Proponents of this model see income distribution as a result of historical, cultural, and political-economic processes (such as slavery and discrimination in the United States). Some of these processes are thought to influence both the private resources available to individuals and shape the nature of public infrastructure for provision of education, health services, transportation, environmental controls, or quality housing (J. Lynch et al., 2004). The processes that generate the income inequality and the resulting maldistribution of resources or stress are thought to be the cause of poor health outcomes. This interpretation views these relationships as spurious correlation.

Previous Literature:

There still remains much controversy over results that support the various versions of the relative income hypothesis as many have argued that support for this hypothesis has been based on flawed analyses. Indeed numerous methodological shortcomings, such as the lack of a natural reference group over which to define communities, the cross-country ecological design of most studies, insufficient controls for correlates of inequality (e.g. race and poverty), and the subjective nature of the dimensions of health considered have plagued studies that support this hypothesis (A. Deaton, 2003). Moreover, most of the studies in this literature are based in wealthy nations, especially the United States, where local taxation laws, race, or other historical processes may confound the results.

This study aims to advance the income inequality hypotheses by examining it in the context of Costa Rica. For decades observers have noted that Costa Rica has remarkably high life expectancy (higher than the United States) despite its limited resources. [2] While many hypotheses have been suggested to help explain this outlier, one recurring untested notion is that Costa Rica is an unusually equitable society in certain dimensions, and that this has contributed to its good health. In fact, supporters of the income inequality hypothesis often point to Costa Rica as evidence for their views. Unlike the United States or Brazil where inequality is highly correlated with ethnic composition and historical discrimination, the Costa Rican setting allows us to assess the relative income hypothesis with less concern for these types of confounders. Moreover, Costa Rica's universal health insurance, highly centralized political system, and public health infrastructure make it an unlikely place to find an association if it is indeed due to local political or structural pathway as has been suggested in other settings (cite). Thus,

[2] Costa Rica's per capita income is roughly one-fifth that of the United States'.

if we do find an association here, where we might not expect to, the results are even more compelling.

This study contributes to the ongoing debate in several ways. By examining how and if inequality affects mortality on a prospective cohort with over 19 years follow-up, this paper looks at the long-term impact of inequality, which has rarely been addressed. With a few exceptions, most studies in this literature examine the income inequality hypothesis in developed countries. Yet, it may be important to look at developing countries that have recently passed through the demographic transition to understand how inequality affects behavior and chronic disease. This study focuses on a middle-income country, Costa Rica, with a strong social safety net, centralized government, and recent increase in inequality due to growing incomes at the top end (see Figure 1). Finally, following Eibner and Evans this study compares results from different measures of socio-economic inequality using measures of income inequality and wealth relative deprivation (C. E. Eibner and W. N. Evans, 2005).

Data Collection and Measures:

This analysis takes advantage of a new data source that links 1984 census records to mortality data in the Costa Rican Vital Statistics registries until December 31, 2007. These data have the distinct advantage of allowing us to follow a nationally representative prospective cohort of approximately 16,000 Costa Ricans adults for over 20 years.

This census-mortality dataset was collected in three phases under the supervision of the Centro Centroamericano de Población (CCP). First, a random stratified sample of approximately 20,000 adults aged 30+ was selected from the 1984 census. Next, the original census files were consulted in order to retrieve the name of the selected individuals, and then names were linked with the Civil Registry to obtain these individuals' "*cedula*" (a unique national identification number similar to a social security number). These matches are performed with an algorithm that employs alphabetic "triads" and observable demographic characteristics to link names from the census questionnaires with the Civil Registry names. Finally, the *cedula* is used to link these records with the Vital Statistics National Death Registry through 2007. For details see Rosero-Bixby, L. and D. Antich 2009.

This matching method linked 18,258 of the census names to *cedulas*. Only those observations with a high probability of a high-quality match are used in this study leading to a sample size of 16,316.[3] Table1 summarizes the data creation process. In this sample of high-quality matches, there were 3,749 deaths between January 1, 1989 and December

---

[3] Since census names could match multiple names in the Civil Registry (e.g. Maria Lopez), an automated process was used to rate the likelihood that the link indeed identified the same person across the two datasets. In this study, I include individuals that either had the best match or the second best match. A match was considered to be a best match if its weight was over 80, the similarity of the year of birth in was no more than + / - 2 years, and the next closest match was at least 10 points or more lower. The second best match were cases where the weight exceeded 70 and had a difference of 15 points or more to the next closest match. Details of matching can be found in Spanish in Luis Rosero-Bixby & Antich, 2009 **Rosero-Bixby, Luis and Daniel Antich.** 2009. "Descripción De Una Muestra De Adultos Del Censo De 1984 Enlazada Con El Registro Civil Para El Estudio De Determinantes De La Mortalidad De Adultos Costarricenses," In. San Jose, Costa Rica: Centro Centroamericano de Población de la Universidad de Costa Rica..

31, 2007. Of these deaths, 3,445 could be linked to the Vital Statistics Records, which further indicates the cause of death information.

To ensure the quality of this data and further assess the potential for the under-registration of deaths, the CCP carried out several consistency checks. First, of the 151 persons aged 85+ in 1984, in the data none were deemed alive by December 2007 and only 3 were lost to follow-up. For those aged 75-84 in 1984, only 16 were alive by December 2007, and several of these individuals were confirmed to be alive and were actually contacted.

Individual-Level Covariates

Throughout all the models demographic characteristics including gender, age, age squared, age cubed, insurance status in 1984, and whether the individual was in a consensual union in 1984 are included as controls. Other controls include education, wealth, and area of residence, which are described below in detail.

<u>Education</u>

Since the sample consists of those aged 30 years or older, most individuals have already completed their education. Consequently, the education variable accounts for the highest level of education completed. It has four categories: (1) No formal education or some primary education; (2) completed primary or completed primary with some secondary education; (3) completed secondary education or completed secondary education with some college education; and (4) received a college or graduate degree.

<u>Wealth</u>

Since there is no income measure in the 1984 census, a wealth measure is included to control for individual-level resources. The wealth measure is a summary index based on owning certain assets (telephone, hot water heater, refrigerator, radio, television, and car) or having certain amenities in the household (access to electricity, piped in water, sewage, and non-dirt floors) in 1984. This measure ranges from 0 to 10. It should also be noted that there were clear regional disparities in this wealth measure. In the province of San José, 29% of the population had the maximum possible value for the wealth measure, while in Limón province only 3% had the maximum possible. Figure 2 presents the box plot of the created wealth index by Province.

<u>Place of Residence</u>

Figure 2 shows that four provinces have consistently wealthier residents. These four provinces, San José, Alajuela, Cartago, and Heredia, also contain parts of the San José Metro area, which reflects the higher wealth in the capital and surrounding central valley. A control variable for living in the San José Metro area is included in models without area fixed effects (discussed below). This may be an important proxy for SES in the models without fixed effects. Moreover, a control variable for whether or not one lives in a rural or urban area is also added to the models. Urbanicity varies within cantons, and as such, it is included as a control even in models with fixed effects.

Inequality Measures

Two measure of socio-economic inequality will be assesses in this study; canton level Gini Coefficient based on income and individual-level relative deprivation based on wealth. In areas with greater income inequality, we expect that there is greater wealth deprivation, thus these two measures are related. However, they do not capture inequality the exact same way. First the under lying data come from two distinct constructs transient income and asset based wealth. Seconds these measures are constructed from distinct data sources. Third, while income inequality varies at the canton level, measures of relative deprivation vary at the individual-level. This final difference is especially important because there may be many unobserved factors associated with income inequality at the area-level, but not strongly related to an individual's deprivation.

The measure of income inequality was based on three years of the Costa Rican Household Surveys for Multiple Purposes, 1989-1991. From this survey total household income was scaled based on a household measure to adjustments for household size and the inequality variable was constructed from households with salaried workers within cantons. These same data were used to construct other canton variables including unemployment, in-migration, and mean household income (see dissertation for complete description of measure (Sepideh Modrek, 2009).

Measures of relative deprivation were calculated from the 1984 census. The relative deprivation measure was calculated over the canton of residence, an individual's age (based on two age groups 30-44, 45-59, 60+), and gender. Following Eibner and Evans, relative deprivation (RD) was defined as the sum of the differences in the wealth index between person i and all others who have more wealth than person i in their canton and age group, j. This measure was calculated as follows:

$$RD_i=1/N^*\sum(y_j-y_i)\ \forall\ y_j> y_i \quad \text{or} \quad RD_i=[E(y|y > y_i)- y_i]^*\ Pr(y> y_i)$$

This measure can be interpreted as the expected difference in wealth between person i and all other individuals with greater wealth than that individual in his/her canton, age, and gender group.

Table 2 presents the summary statistics for the individual-level measures, while Table 3 presents the summary statistics for the area-level measures.

Outcomes

The outcome of interest is time to death. Survival between January 1, 1989, and December 31, 2007 is examined for a potential 19-year follow up.[4] Using these data we find that median age of survival was 80.96 for men (Interquartile range: 71.44-88.58) and 84.6 for women (Interquartile range: 75.80-90.85).

In the presented analyses the follow-up period began on January 1, 1989, because 1989 was the first year for which consistent inequality measures are available, even though the follow-up in the initial data began on September 9, 1984. Examining the data from 1989 onward, 4,085 deaths were recorded, and 3,725 of them had an associated

---

[4] In general, Costa Rica's Vital Statistics Registry is ranked as complete by WHO and considered of high quality similar to that of the United States, Japan etc. **Mahapatra, Prasanta; Kenji Shibuya; Alan D. Lopez; Francesca Coullare; Francis C. Notzon; Chalapati Rao and Simon Szreter.** 2007. "Civil Registration Systems and Vital Statistics: Successes and Missed Opportunities." *The Lancet*, 370(9599), pp. 1653-63.

cause of death. Heart disease mortality was examined separately because studies suggest that it is highly related to socio-economic factors (Nancy E. Adler and Joan M. Ostrove, 1999). Figure 3 and 4 present Kaplan-Meier estimate of the survival function for these two outcomes by gender. Separate survival curves are drawn for three levels of relative deprivation. These descriptive graphs show a slight difference in survival rate by these categories, but we examine these data in a multivariate context.

Analysis Methods:

This study employs maximum likelihood survival regressions with a Gompertz hazard distribution. The Gompertz distribution has been widely used to study mortality, and these models are known to fit these data well (L. Rosero-Bixby and W. Dow, 2009). The key assumption in the Gompertz model is that the underlying rate of mortality is monotonically increasing with time, or in our case, age.

The Gompertz hazard equations is

$$H(t)=\lambda e^{\gamma t}$$

where γ is fixed for all individuals and represents the external "force" of mortality, while λ is estimated based on a log-linear equation. The λ is a function of the covariates, $\lambda=\exp(\boldsymbol{\beta}\mathbf{X})$. The first two sets of models will examine the income inequality hypothesis in its various formulations.

In the first group of models, I_{ct} is the income inequality in canton *c* at time *t*, A_{ct} is a vector of the canton-level controls at time *t*, $\mathbf{Z}_i$ is a vector of individual controls measured in 1984, and I_{ct} *$LW_{\mathbf{i}}$ is an indicator variable for having low wealth interacted with the canton Gini coefficient. The estimate for β_1 is meant to capture the association between income inequality and survival risk. The interaction is included to examine if income inequality has a different impact on the poor.

Group 1

$$\boldsymbol{\beta}\mathbf{X} = \beta_0 + \beta_1 (I_c) + \beta_2 (\mathbf{A}_c) + \boldsymbol{\beta}_{\mathbf{3Z}} (\mathbf{Z}_i) + \boldsymbol{\varepsilon}$$

$$\boldsymbol{\beta}\mathbf{X} = \beta_0 + \beta_1 (I_{ct}) + \beta_2 (\mathbf{A}_{ct}) + \boldsymbol{\beta}_{\mathbf{3Z}} (\mathbf{Z}_i) + \boldsymbol{\beta}_{\mathbf{4}}\, I_c * LW_{\mathbf{i}} + \boldsymbol{\varepsilon}$$

A second group of models will assess the relative deprivation version of the relative income hypotheses. These models will use wealth as the underlying metric to create the relative deprivation measure. Though not directly comparable to the models above, they serve as complementary analyses. Since RD varies at the individual level, we can also include canton-level fixed-effects to account for time invariatant canton characteristics that are corrolated with relative deprivation and mortalitiy risk in this second set of models.

Group 2

$$\boldsymbol{\beta}\mathbf{X} = \beta_0 + \boldsymbol{\beta}_{\mathbf{3Z}} (\mathbf{Z}_i) + \beta_6 RD_i + \boldsymbol{\varepsilon}$$

$$\boldsymbol{\beta}\mathbf{X} = \beta_0 + \boldsymbol{\beta}_{\mathbf{3Z}} (\mathbf{Z}_i) + \beta_6 RD_i + \boldsymbol{\beta}_{\mathbf{7}}\mathbf{y}_{\mathbf{c}} + \boldsymbol{\varepsilon}$$

Here $\mathbf{Z}_i$ is a vector of individual controls measured in 1984 (gender, age, age squared, age cubed, insurance status, consensual union, wealth, education, and place of residence);

RD_i is the individual's relative deprivation as compared to those living within their canton and in their age group; and $\mathbf{y_c}$ is a vector of area-level fixed effects. Here the canton fixed effects will capture common unobserved factors shared among those living in the same canton, such as access to health services that may be related to both the variation in deprivation and mortality.[5] The RD variable varies at the individual-level, so area-level controls are not considered (but will be considered in sensitivity analysis).

The start time used in the presented analyses is January 1, 1989. This was necessary so that we could compare similar samples across the two measures of socio-economic inequality. All models are weighted by the probability that the observation had a successful match and are clustered at the canton level.

Results:

Income Inequality Hypotheses

Table 4 presents the results for all-cause mortality using the Gompertz models described above. The first three models (columns 1-3) present the relationship between income inequality and survival time for all causes of death, whereas the last three models (columns 4-6) present the results for heart disease mortality. Columns 1 and 4 present the basic model without including the canton-level characteristics controls. Columns 2 and 5 include the area-level control variables. Finally columns 3 and 6 include an interaction term between income inequality and having low wealth.

The results from these analyses do not support the income inequality hypothesis. These models show that those residing in areas with greater income inequality actually have *longer* survival. The unscaled coefficients in Table 4 are quite large because they represent a one-unit change in inequality, effectively going from perfect equality to perfect inequality; the scaled coefficient imply that a one standard deviation change in inequality is associated with a 6% lower hazard (HR=0.94, CI 0.91-0.97) which is a modest effect size. The magnitude on the interaction of income inequality and having low wealth, further suggests that those with lower wealth have longer survival in more unequal area.[6] Moreover, area-level mean income is associated with lower survival time. This results counters initial expectations and contradicts results from the individual wealth measure where those with greater individual wealth have longer survival.

In contrast to the area-level measures, the results for individual-level charactersitics are as expected. Women have a 30% lower hazard of mortality. Likewise, those who are partnered in 1984 have a survival advantage (HR= 0.81, CI 0.78-0.85), as do the wealthy. Each additional household ameneity owned, our measure of wealth, confers a 3% or 4% lower mortality hazard (HR= 0.97, CI 0.96-0.98). Those who complete college have a 24% lower mortality hazard (HR= 0.74, CI .68-.80).[7] Those who reside in a rural area also have 12% lower hazard of mortality. Finally, those without insurance in 1984 also live longer providing evidence that there was adverse selection into the public insurance program at that time.

---

[5] Eibner and Evans (2005) also include these types of fixed effects in their models.
[6] Though the coefficients are not significant.
[7] However, this population accounts for only 10 percent of the sample and these tend to be concentrated in the younger generations.

The analyses for cardio-vascular disease mortality exhibit the same puzzling results with regard to income inequality. Here again the estimated coefficient suggests that residents in areas with higher income inequality have better survival. A one standard deviation change in income inequality is associated 11% better survival (HR=0.89, CI 0.84-.093]. Although not significant, the magnitude on the coefficient of the interaction of income inequality and having low wealth, suggests that those with lower wealth living in more unequal areas have better survival. Finally, area-level mean income is associated with lower survival time.

In contrast to the area-level measures, the results for the individual-level characteristics associated with heart disease mortality were consistent with previous literature. Women have a lower hazard of cardiovascular mortality, as do those with college education. The results also show that those with no insurance in 1984 have reduced heart disease mortality, again pointing to adverse selection into the public insurance program.

The results for the canton-level variables are contrary to expectations and give rise to concerns of omitted variables at the area-level that maybe correlated with both income inequality and mortality. Hence, we examine an alternate measure of inequality, relative deprivation, to assess the relative income hypothesis.

Relative Deprivation Hypothesis

Table 5 presents the results from these analyses. Columns 1-2 present the results for all-cause mortality and columns 3-4 present the results for heart disease mortality. Each additional unit of wealth deprivation, i.e. having one fewer item than those who are wealthier, is associated with a 11% higher hazard of death (HR=1.11, CI 1.06-1.15), even when we include canton-fixed effects. Also noteworthy, the RD measures eliminate the significance of the protective effects of having high wealth. This suggests the RD maybe capturing some of effects of wealth as those that are more wealth are by definition less deprived and the correlation between the RD measure and wealth is quite high (correlation coefficient=-0.82, p-val=0.000). From a theoretical perspective we are interested in both parameter estimates, so we should include both in the model, but if we choose to drop the wealth variable the estimated effect of RD is reduced (HR=1.063, CI 1.04-1.08). The other coefficients on the other individual-level covariates remain largely unchanged in these models.

For heart-disease mortality, the results indicate that each additional unit of deprivation is associated with a 17% increase in the risk of death from heart disease (HR=1.17 CI 1.1-1.23), but this is coefficient is not significant with the inclusion of canton fixed effects. Again there is a big change in the wealth effect. When we exclude the wealth variable and examine the parameter estimate on the RD variable, the magnitude of the estimated coefficient becomes essentially zero and the standard errors increase (HR= 1.021 CI=0.989-1.05); other coefficients remain largely unchanged in this exercise.

Finally, we examine if the RD measures correlate to poor health behavior or other markers of health as was shown by Eibner and Evans (C. E. Eibner and W. N. Evans, 2005). While we cannot do these analyses with the census-linked mortality data, other data with health behaviors are available. Using an on-going longitudinal study of a nationally representative sample of adult Costa Ricans aged 60+ and the 2000 census to construct an RD measure, we examine if RD was related to self-reported health, smoking, and alcohol consumption. We find that higher RD is related to both smoking behavior and self-reported poor health (see Table 6). While only suggestive, these independent findings strengthen the results from the RD analysis presented above.

Sensitivity Analyses:

In order to ensure that these results were not too reliant on any specification or assumptions, several additional analyses were done.

First, the parametric Gompertz models were compared with Cox Proportional Hazard models, which have less stringent parametric assumptions. The coefficients across these two sets of models are very similar (not shown). We chose to use the parametric Gompertz models because they are often the models of choice for mortality studies. Next several parametric models are compared in terms of their fit to the data—in case other models were better fits. Comparing the exponential, Weibull, Gompertz, Lognormal, Log-Logistic, and Gamma models using the Akaike's information criterion (Hirotugu Akaike, 1981) showed that the Gompertz had the lowest AIC score for both all-cause mortality and cardiovascular disease mortality; confirming that the Gompertz model is the best parametric survival model for these data.

Second, we test we see if our results vary by sample inclusion and the timing criteria. We use Jan 1, 1989 as the beginning of the analysis because that is the earliest year for which we have a consistent measure of income inequality using a large underlying sample (See dissertation for details). Nevertheless, we also examine the inequality effect of another year, 1987, and find that the results are similar. Moreover, for the RD measure we again examine deaths after 1989 because we want to have a comparable sample across measures of inequality. However when we reanalyze the data using different start dates for the study (and hence different samples), the results do not change appreciably. Comparing the results from three time points as the starting point of the study, September 9, 1985, January 1, 1987, January 1, 1989, we find that the results are remarkably similar.

Third, we examine multiple versions of the RD measure. In line with Eibner and Evans, we create different measures to account for potentially different relationships based on reference group. Since we lack detailed information on individual's networks, we use observable demographic characteristics to construct these measures. One RD measure was based on just canton, one was based on sex & canton, and another one age, sex and canton. The final variant is the one presented in the results above. These three measures were highly correlated (correlation coefficients ranging from 0.90-.98) and the results were very stable across all three measures. Note that the RD measures are constructed based on the complete 1984 census, so the measures take into account the entire population, not just those selected for the linkage.

Fourth, looking carefully at the Kaplan-Meier survival curves in Figure 3 and 4, indicate that RD may be have a different impact on mortality hazard by gender. The low RD group (blue) seems to have slightly better survival for women and slightly worse survival for men. In order to test this observation, we include a gender and RD interaction in our models. We find that RD affects do not vary by gender in a multivariate context.

Finally, to account for the potential for interdependence for observations within canton, we look at different ways of clustering for canton interdependence effects for income inequality measure. Regardless of method, either shared frailty models or clustering the standard errors, the standard errors were very similar. We choose to use clustering as it allows us to use weights whereas the share frailty will not in STATA 11.

Discussion and Limitations:

How do we reconcile the results from the two measures of socio-economic inequality? One reason for the difference in results may be related to the cross-sectional design of the study. Since we cannot include controls for area-level fixed effects in the models with income inequality, the extent that income inequality is related to other area-level phenomena, such as economic growth, make these analyses particularly vulnerable to omitted variable bias. In Costa Rica, the growing divergence in incomes has been most concentrated in wealthy and economically active areas (Sepideh Modrek, 2009). Although area-level income inequality and area-level mean wealth deprivation are positively correlated (correlation coefficient= 0.2646; p-val= 0.0177), the results indicate opposite effect on health, which may be due to area-level confounders.

Another reason for the differences in results may be that since relative deprivation varies at the individual level, it is a better measure of socio-economic heterogeneity especially for the poor. In an effort to understand if income inequality affected the poor disproportionately, an interaction term was included in the income inequality models above. But, this interaction term may be too crude a measure, and thus, the RD measure may be capturing inequality affects for the poor better.

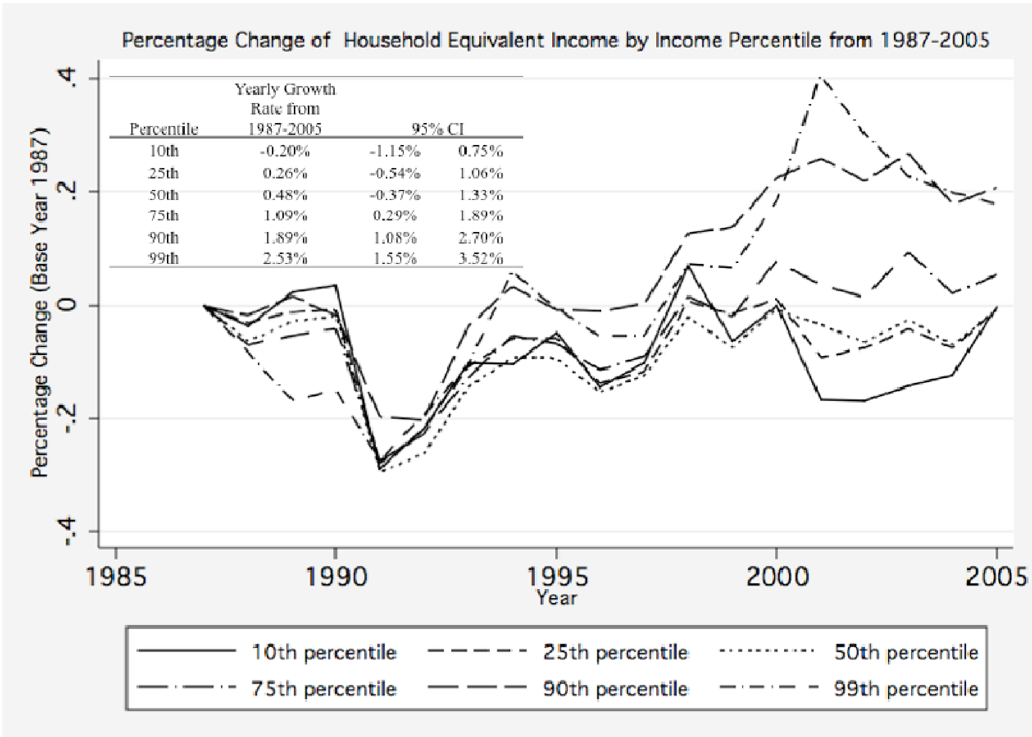
Yet another reason for the difference in results, it may be that the relative deprivation measure is picking up other non-linear effects of wealth on health. As noted above, when the RD measure is included in the models, an individual wealth is not longer significantly related to their mortality hazard (although coefficients do not change much). Thus, it may be that RD capturing non-linear wealth affect. Also, the magnitudes of the effect sizes are generally small (in either direction) and only marginally significant, regardless of which results/measure we decide to take seriously. It may be that the relative socio-economic status is not a particularly important hazard to health in this setting. Nonetheless, given that the results from the RD measure are in line with the theoretical model, results on behavior and perceived health, and other studies, they may warrant further investigation even though the results indicate a relatively weak relationship.

Beyond the results, there are several other methodological weaknesses that should be noted. First using canton as a reference group maybe problematic if individuals do not compare themselves to those living nearby, but rather compare themselves to the nation as a whole. This is an inherent problem in much of this literature that requires further theoretical and empirical clarification. Second, measures of income inequality are based on household surveys done on a 1% sample and thus for some cantons, inequality was measured using less that 250 observations. This may lead to some measurement error of the construct. Third, creating relative deprivation measure from measures of wealth that have only 10 levels may significantly underestimate the level of heterogeneity of real socio-economic assets (real net wealth including savings and debt) within cantons. Finally, it maybe that unobservable characteristics such as one's discount rate or level of risk-aversion may be related to both their socio-economic status and their health behaviors.

While this study had several methodological limitations, there are also a few noteworthy advantages. First and foremost, this study has an exceptionally long-term follow-up period. There are few studies with as many observations, followed for so long, on a nationally representative sample. Second, Costa Rica's institutional setting and homogenous population makes us less worried about political channels that may be related to both inequality and health. That may explain why we find such small effect compared to Eibner and Evens who look at RD within the US where political channels and race relations are likely to be important.

Although the results of this study provide mixed support for the association between inequality and long-term mortality, there seems to be some evidence that RD may be related to mortality through health behaviors that remains to be explored. We have only begun to understand Costa Rica's exceptional health outcomes and their relations to social hierarchy; and this topic merits continued scrutiny in future work.

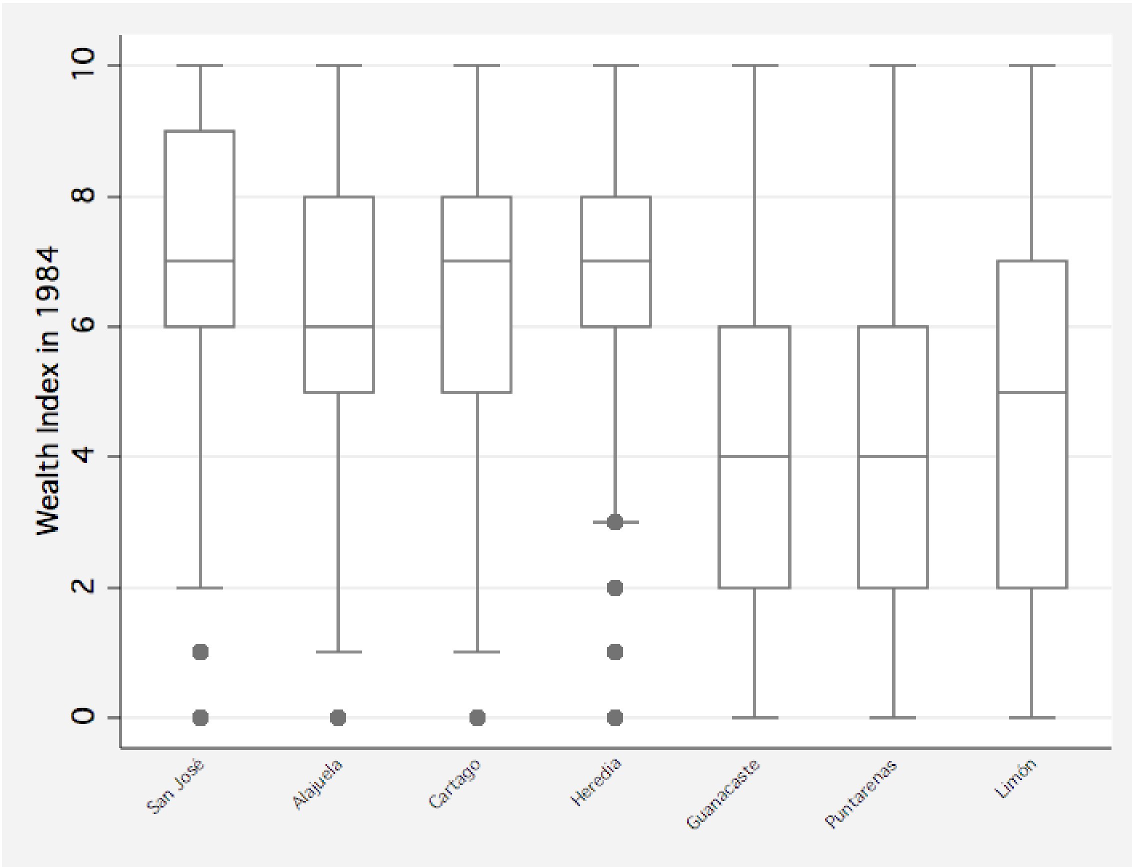
Figure 1: Percentage change in Household Equivalent Income from 1987 to 2005 by Income Percentile



Percentile	Yearly Growth Rate from 1987-2005	95% CI	
10th	-0.20%	-1.15%	0.75%
25th	0.26%	-0.54%	1.06%
50th	0.48%	-0.37%	1.33%
75th	1.09%	0.29%	1.89%
90th	1.89%	1.08%	2.70%
99th	2.53%	1.55%	3.52%

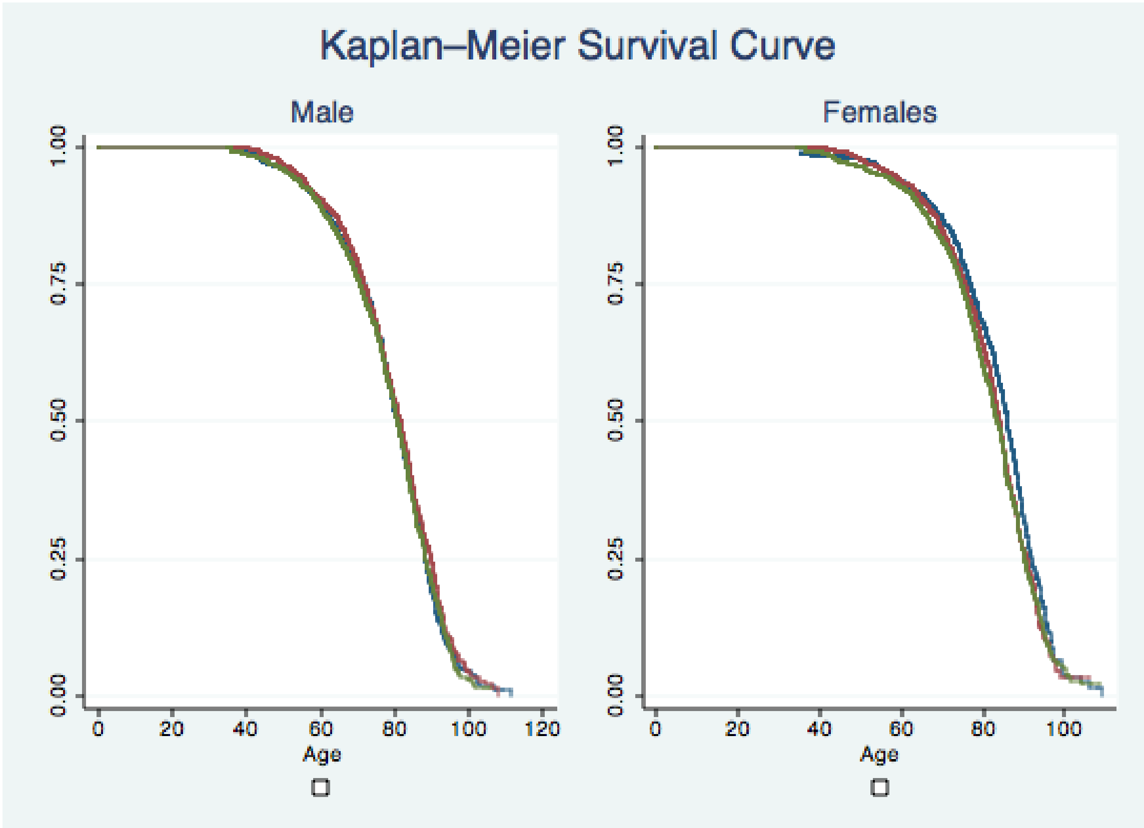
Source: Author's calculation based on Costa Rican Household Survey for Multiple Purposes and Costa Rican CPI deflators found at Banco Central De Costa Rica (http://indicadoreseconomicos.bccr.fi.cr). All colones adjusted to 2006 colones.

Figure 2: Wealth Index Box Plot by Province



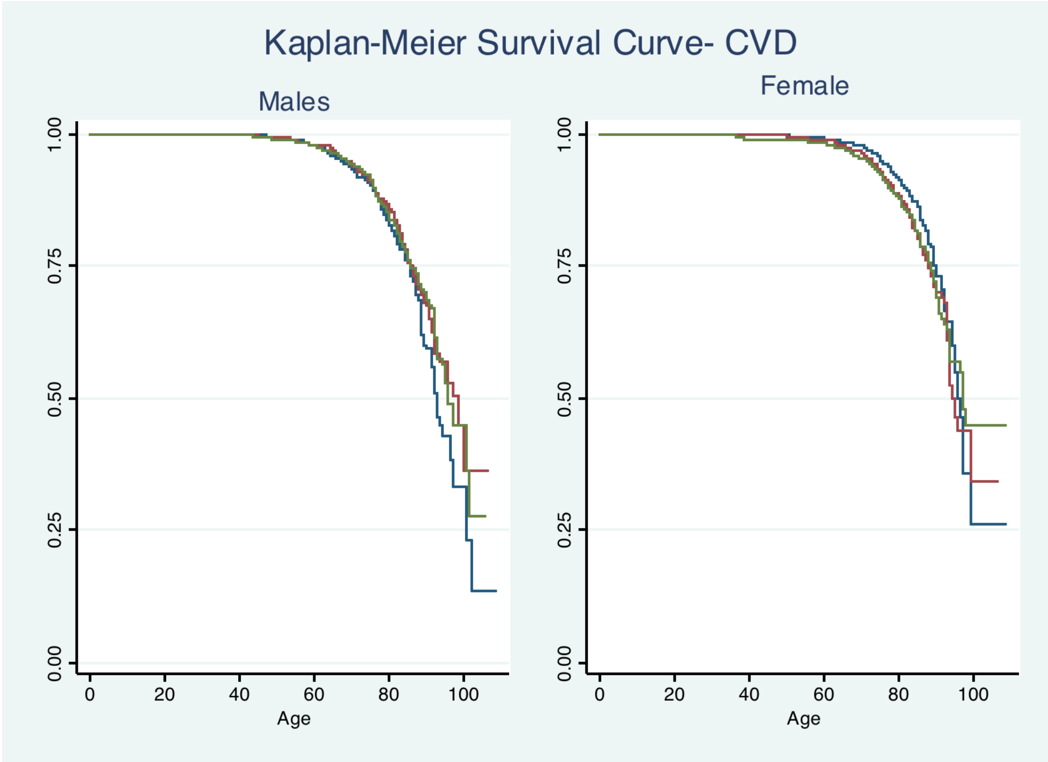
Note: Wealth index calculated for Costa Ricans over age 30. Author's calculations based on 1984 census-10% micro sample from IPUMS

Figure 3: Kaplan-Meier Survival Curves for All-Cause Mortality



Note: Relative Deprivation (RD) measure was broken down into tertiles. Blue line indicates low RD, red line indicates mid-level RD, and green line indicates high RD.

Figure 4: Kaplan-Meier Survival Curves for Cardio Vascular/Heart Disease Mortality



Note: Relative Deprivation (RD) measure was broken down into tertiles. Blue line indicates low RD, red line indicates mid-level RD, and green line indicates high RD.

Table1: Census Mortality Linkage Sample Creation

	N	Using 09Sep1984 as start date	Using 01Jan1987 as start date	**Used in Analysis** Using 01Jan1989 as start date
*Stratified Random Sample **				
Met selection criteria in 1984 census	21161			
*Names Obtained From Census Ballots**				
Names Found	19954			
*Names Linked to Cedula Number**				
All Matches	18258			
Good Matches	16315			
Live in Regular HH & All Covariates	15721	15698	15473	15276
Deaths				
Dead by December 31, 2007		4168	3943	3746
Cause of Death Known		3818	3445	3445

*Details found in Rosero-Bixby, L. and D. Antich (2009). Descripción de una muestra de adultos del censo de 1984 enlazada con el Registro Civil para el estudio de determinantes de la mortalidad de adultos costarricenses. San Jose, Costa Rica, Centro Centroamericano de Población de la Universidad de Costa Rica.

Table 2: Means of Individual-Level Variables

	Weighted Mean	Linearized SD
Age	46.43	0.24
Gender		
Female (%)	49.97%	0.83%
Education		
No Formal Education Completed (%)	50.59%	3.25%
Completed Primary (%)	32.25%	1.56%
Completed Secondary (%)	6.53%	0.83%
Completed College (%)	10.62%	1.40%
Geographical Area		
Live in San Jose Great Metro Area in 1984 (%)	55.11%	8.46%
Live in Rural Area (%)	49.63%	7.04%
Wealth Group		
Wealth Index [Range 0-10]	6.31	0.27
Low Wealth (<=5) (%)	33.05%	4.49%
Insurance Status		
No insurance in 1984 (%)	22.74%	0.93%
In Union		
Union (%)	75.85%	1.26%
Relative Deprivation Measure		
Relative Deprivation [Range 0-7.9] By Canton, Age, Gender	0.9858	0.0480
Relative Deprivation [Range 0-7.9] By Canton	0.9857	0.0480
Observations	15276	15276

Table 3: Area Level Covariates

Variable	Population Weighted Mean	Std. Dev.	Obs
Gini Coefficient *	0.354	0.049	80
(Mean HHEMI)/1000 in 2006 Colones*	189.10	54.30	80
Unemployment Rate	0.046	0.018	80
In-Migration	0.042	0.023	80

Weighted by canton population

* Includes only salaried workers

Source: Author's calculation based on Costa Rican Household Survey for Multiple Purposes and Costa Rican CPI deflators found at Banco Central De Costa Rica (http://indicadoreseconomicos.bccr.fi.cr). All colones adjusted to 2006 colones.

Table 4: Gompertz Models of Mortality Risk and Income Inequality

	ALL-CAUSE			Heart Disease		
Canton Level Area Covariates from July Household Surveys- Time varying						
Gini of Monthly Household Equivalent Salary	-0.8609	-1.05109	-0.92838	-1.98772	-1.95616	-1.65622
	[0.48286]*	[0.48698]**	[0.50940]*	[0.75035]***	[0.82767]**	[0.80125]**
Mean of the Monthly Household Equivalent Salary in 2006/ 10,000	0.00841	0.01343	0.01325	0.02553	0.02389	0.02347
	[0.00554]	[0.00635]**	[0.00634]**	[0.00866]***	[0.01187]**	[0.01172]**
Interaction of Wealth & Area Level Gini						
Gini*Low Wealth			-0.19986			-0.49228
			[0.18695]			[0.34348]
Percent Unemployed		0.24842	0.21789		-0.88979	-0.92379
		[0.71320]	[0.71971]		[1.68237]	[1.66258]
Percent In-Migration		-1.65366	-1.65974		0.24711	0.21599
		[0.67260]**	[0.67234]**		[1.78315]	[1.77795]
Individual Level Covariates from 1984 Census						
Female	-0.35506	-0.35603	-0.35452	-0.36605	-0.36661	-0.3622
	[0.03999]***	[0.03987]***	[0.03981]***	[0.07251]***	[0.07293]***	[0.07322]***
Education Catagories (None/some Primary omitted)						
Completed Primary	-0.08336	-0.08349	-0.08309	-0.17323	-0.17297	-0.17108
	[0.04439]*	[0.04444]*	[0.04440]*	[0.09930]*	[0.09922]*	[0.09974]*
Completed Secondary	-0.01373	-0.01356	-0.00779	0.02356	0.02225	0.03894
	[0.06376]	[0.06367]	[0.06312]	[0.14958]	[0.14997]	[0.14803]
Completed College	-0.30234	-0.30349	-0.29614	-0.56278	-0.56443	-0.54427
	[0.07761]***	[0.07814]***	[0.07920]***	[0.13327]***	[0.13318]***	[0.13290]***
Geographical Area						
Live in San Jose Metro Area	0.02487	0.02229	0.0227	-0.1052	-0.10052	-0.09965
	[0.06043]	[0.05774]	[0.05743]	[0.12144]	[0.12717]	[0.12646]
Live in Rural Area	-0.12816	-0.12904	-0.1257	-0.10238	-0.10408	-0.09687
	[0.05002]**	[0.04835]***	[0.04914]**	[0.08888]	[0.09063]	[0.09142]
Wealth (Range 0-10)						
Wealth	-0.02695	-0.02728	-0.03902	0.00563	0.00596	-0.02287
	[0.01209]**	[0.01197]**	[0.01573]**	[0.01831]	[0.01824]	[0.02485]
Insurance Status						
No insurance in 1984	-0.0887	-0.08818	-0.08759	-0.22492	-0.22596	-0.22331
	[0.03281]***	[0.03268]***	[0.03258]***	[0.08327]***	[0.08341]***	[0.08344]***
In Union						
Union	-0.20979	-0.20857	-0.20888	-0.11112	-0.11152	-0.1126
	[0.04582]***	[0.04595]***	[0.04592]***	[0.08777]	[0.08799]	[0.08767]
Constant	-6.58095	-6.5476	-6.46116	-11.14971	-11.11607	-10.9565
	[0.82572]***	[0.82565]***	[0.82085]***	[1.55125]***	[1.54326]***	[1.54968]***
Canton Fixed Effects	No	No	No	No	No	No
Log-Likelihood	-651.99	-650.48	-649.87	-1265.02	-1264.89	-1264.01
Gamma	0.08	0.08	0.08	0.08	0.08	0.08
Observations	15276	15276	15276	14975	14975	14975
* significant at 10%; ** significant at 5%; *** significant at 1%						
Robust clustered standard errors in brackets						
Controls for age, age squared and age cubed included						

Table 5: Gompertz Models of Mortality Risk and Relative Deprivation

	All-Cause		Heart Disease	
Individual Level Covariates from 1984 Census	0.09361	0.10761	0.16081	0.1635
Relative Deprivation Measure	[0.02974]***	[0.04188]**	[0.05889]***	[0.09957]
	-0.36104	-0.36678	-0.37246	-0.36906
Female	[0.03974]***	[0.03913]***	[0.07112]***	[0.07362]***
Education Catagories (None/some Primary omitted)	-0.08878	-0.06867	-0.1809	-0.17081
Completed Primary	[0.04410]**	[0.04547]	[0.09810]*	[0.10134]*
	-0.02452	0.01865	0.02536	0.03161
Completed Secondary	[0.06000]	[0.06408]	[0.16342]	[0.15615]
	-0.31759	-0.26654	-0.56924	-0.55548
Completed College	[0.07399]***	[0.07340]***	[0.14497]***	[0.13398]***
Geographical Area	0.02152		-0.02856	
Live in San Jose Metro Area	[0.05229]		[0.10730]	
	-0.10767	-0.10926	-0.07047	-0.12628
Live in Rural Area	[0.05022]**	[0.06232]*	[0.09456]	[0.11360]
Wealth Group(Middle Wealth omitted)	0.0218	0.03012	0.09233	0.08784
Wealth	[0.01916]	[0.02611]	[0.03604]**	[0.06046]
Insurance Status	-0.09428	-0.08184	-0.23639	-0.22816
No insurance in 1984	[0.03281]***	[0.03548]**	[0.08344]***	[0.08605]***
In Union	-0.20635	-0.21283	-0.1067	-0.09285
Union	[0.04513]***	[0.04589]***	[0.08766]	[0.09180]
	-6.99165	-7.17487	-11.90832	-12.14068
Constant	[0.83545]***	[0.84669]***	[1.54726]***	[1.60038]***
Canton Fixed Effects	No	Yes	No	Yes
Log-Likelihood	-650.73	-604.41	-1267.22	-1226.07
Gamma	0.08	0.08	0.08	0.08
Observations	15276	15276	14975	14975
Robust clustered standard errors in brackets				
* significant at 10%; ** significant at 5%; *** significant at 1%				
Controls for age, age squared and age cubed included				

Table 6: Logistic Regression Models of Perceived Health and Smoking

		Means [Weighted Std. Errors]	Coefficient on Relative Deprivation in 2000 [Robust Clustered Standard Errors]	
	N			
Poor Health	1905	0.4467	0.1558	0.2787
		[0.0214]	[0.1197]	[0.1339]**
Smoking	1905	0.4303	0.2416	0.1934
		[0.0139]	[0.1114]**	[0.1116]*
Canton FE			No	Yes

Logit model
* significant at 10%; ** significant at 5%; *** significant at 10%
Controls of age, age squared, male, male*age, married, education (3 categories), wealth (3 categories), childhood wealth, and urban/rural.

Note: Based on the 2005 Costa Rican Study on Longevity and Healthy Aging project, a study of healthy aging in Costa Rican aged 60+ in 2000 census. The RD measures were constructed based on wealth measures (ownership of assets and other household amenities) from the 100% sample of the 2000 census. The reference group was constructed by age, sex and canton of residence in 2000.

Adler, N. E.; T. Boyce; M. A. Chesney; S. Cohen and S. Folkman. 1994. "Socioeconomic Status and Health: The Challenge of the Gradient." *American Psychologist*, 49(1), pp. 15-24.
Adler, Nancy E. and Joan M. Ostrove. 1999. "Socioeconomic Status and Health: What We Know and What We Don't." *Annals of the New York Academy of Sciences*, 896(1), pp. 3-15.
Akaike, Hirotugu. 1981. "Likelihood of a Model and Information Criteria." *Journal of Econometrics*, 16(1), pp. 3-14.
Baum, Andrew; J. P. Garofalo and Ann Marie Yali. 1999. "Socioeconomic Status and Chronic Stress: Does Stress Account for Ses Effects on Health?" *Annals of the New York Academy of Sciences*, 896(1), pp. 131-44.
Chong, A. and M. Gradstein. 2006. "Inequality and Institutions." *Review of Economics and Statistics*, 89(3), pp. 454-65.
Deaton, A. 2003. "Health, Inequality, and Economic Development." *Journal of Economic Literature*, 41, pp. 113–58.
Eibner, C. E. and W. N. Evans. 2005. "Relative Deprivation, Poor Health Habits and Mortality." *Journal of Human Resources*, 40(3), pp. 591–620.
Kawachi, I.; B. P. Kennedy; K. Lochner and D. Prothrow-Stith. 1997. "Social Capital, Income Inequality, and Mortality." *Am J Public Health*, 87(9), pp. 1491-98.
Krugman, Paul. 1996. "The Spiral of Inequality," In *Mother Jones*, 44-49. San Francisco
Lynch, J.; G. D. Smith; S. A. M. Harper and M. Hillemeier. 2004. "Is Income Inequality a Determinant of Population Health? Part 2. Us National and Regional Trends in Income Inequality and Age-and Cause-Specific Mortality." *The Milbank Quarterly*, 82(2), pp. 355-400.
Mahapatra, Prasanta; Kenji Shibuya; Alan D. Lopez; Francesca Coullare; Francis C. Notzon; Chalapati Rao and Simon Szreter. 2007. "Civil Registration Systems and Vital Statistics: Successes and Missed Opportunities." *The Lancet*, 370(9599), pp. 1653-63.
McEwen, B. 2004. "Protective and Damaging Effects of the Mediators of Stress and Adaptation: Allostasis and Allostatic Load." *Allostasis, Homeostasis, and the Costs of Physiological Adaptation*.
Modrek, Sepideh. 2009. "Income Inequality and Mortality: The Costa Rican Case," In *Health Services and Policy Analysis*, 122. Berkeley: University of California, Berkeley.
Putnam, Robert D. 2000. *Bowling Alone: The Collapse and Revival of American Community*. New York, NY Simon & Schuster.
Rosero-Bixby, L. and W. Dow. 2009. "Ses Gradients in Adult Mortality in a Middle Income Country: The Costa Rican Longitudinal Mortality Study 1984-2007," In. Berkeley: UC Berkeley.
Rosero-Bixby, Luis and Daniel Antich. 2009. "Descripción De Una Muestra De Adultos Del Censo De 1984 Enlazada Con El Registro Civil Para El Estudio De Determinantes De La Mortalidad De Adultos Costarricenses," In. San Jose, Costa Rica: Centro Centroamericano de Población de la Universidad de Costa Rica.
Sapolsky, Robert M. 2004. "Social Statud and Health in Humans and Other Animals " *Annual Review of Anthropology*, 33(1), pp. 393-418.

Wilkinson, Richard G. 1996. *Unhealthy Societies: The Afflictions of Inequality*. London: Routledge.