# Heart Mortality and Urbanization: The Role of Unobserved Predictors.

PRELIMINARY VERSION, PLEASE DO NOT CITE WITHOUT AUTHORS' PERMISSION.

Thomas Bassetti* Nikos Benos† Stelios Karagiannis‡

February 27, 2013

**Abstract**

This article explores rural-urban differences in cardiovascular mortality. Using U.S. county-level data and the Blinder-Oaxaca decomposition technique, we study how observed and unobserved risk factors affect the mortality rate in metropolitan vs. non-metropolitan areas. Results indicate that traditional risk factors are less abundant in metropolitan counties, in line with the existing literature. However, unobserved factors are less prevalent in rural areas. In other words, observed and unobserved variables have opposite effects on the mortality gap. Our findings remain intact when we distinguish between different types of heart diseases and examine men and women separately. Finally, unobserved determinants of mortality are notably significant in nonmetro counties adjacent to metro ones.

**JEL codes:** I10; E6; C15
**Key words:** Cardiovascular disease; Decomposition analysis; Urbanization.

**Acknowledgment**: For helpful comments, we thank Efrem Castelnuovo, Luca Corazzini, Blaise Melly, Guglielmo Weber, all seminar participants at the University of Padua and at the IMAEF 2012. The usual disclaimer applies.

*Corresponding author. University of Padova, Dept. of Economics "Marco Fanno" and Interuniversity Center for Growth and Economic Development (CICSE), Via del Santo 33, 35123 Padova (PD), Italy. E-mail: thomas.bassetti@unipd.it

†University of Ioannina, Dept. of Economics, University Campus, 45110, Ioannina, Greece. E-mail: nbenos@cc.uoi.gr.

‡Centre of Planning and Economic Research (KEPE), Amerikis 11, 10672, Athens, Greece. E-mail: stkarag@kepe.gr.

# 1 Introduction

More than 86 million, or 36.9% of adults in the United States were afflicted with some form of cardiovascular disease (CVD) in 2010. Projections show that prevalence will increase to 40.5% (116 million) by 2030. Specifically, heart diseases and stroke are respectively the first and fourth leading causes of death in the United States for 2011.[1]

All together, these diseases contribute to around 31% of deaths in the U.S. From an economic point of view, a recent report states that medical expenses devoted to CVD reached 2.6% of US GDP in 2008 (Heidenreich et al., 2011). Direct medical costs of CVD are projected to triple from $273 to $818 billion, in real terms, between 2010 and 2030. Indirect costs, associated with lost productivity, are also estimated to rise from $172 to $276 billion during the same period (see Table 1 below). Coronary heart disease alone costs the United States $108.9 billion each year. Real medical costs of coronary and heart failure are projected to rise by around 200% until 2030, and stroke is forecast to exhibit the strongest relative increase of 238%. In this framework, since chronic diseases have negative effects on mental and physical capabilities, there is a negative relationship between heart diseases and human capital accumulation. Therefore, a reduction in CVDs will cause an increase in both labor productivity and economic growth. By looking at high income countries, Suhrcke and Urban (2010) verify that CVD mortality play a negative role on subsequent five-year growth rate.[2]

[1] Source: Centers for Disease Control and Prevention (2012).

[2] Suhrcke and Rocco (2008) provide a detailed review of microeconomic literature on non-communicable diseases. Other interesting studies on the relationship between health and economic growth are: Levine and Renelt (1992), Sala-i-Martin et al. (2004), Lorentzen et al. (2008) and Swift (2011).

Table 1. Projected Costs of CVD in the United States (Billions 2008$), 2010–2030

| Year | All CVD* | Hypertension | CHD | HF | Stroke | Hyper. as RF** |
|---|---|---|---|---|---|---|
| Direct (Medical) | | | | | | |
| 2010 | 272.5 | 69.9 | 35.7 | 24.7 | 28.3 | 130.7 |
| 2015 | 358.0 | 91.4 | 46.8 | 32.4 | 38.0 | 170.4 |
| 2020 | 470.3 | 119.1 | 61.4 | 42.9 | 51.3 | 222.5 |
| 2025 | 621.6 | 155.0 | 81.1 | 57.5 | 70.0 | 293.6 |
| 2030 | 818.1 | 200.3 | 106.4 | 77.7 | 95.6 | 389.0 |
| % Change | 200 | 186 | 198 | 215 | 238 | 198 |
| Indirect (Lost Productivity) | | | | | | |
| Year | All CVD* | Hypertension | CHD | HF | Stroke | Hyper. as RF** |
| 2010 | 171.7 | 23.6 | 73.2 | 9.7 | 25.6 | 25.4 |
| 2015 | 195.7 | 27.2 | 82.8 | 11.3 | 29.7 | 29.3 |
| 2020 | 220.0 | 31.0 | 92.0 | 13.0 | 34.0 | 33.3 |
| 2025 | 246.1 | 35.1 | 101.5 | 15.1 | 38.9 | 37.8 |
| 2030 | 275.8 | 39.8 | 112.3 | 17.4 | 44.4 | 42.8 |
| % Change | 61 | 69 | 53 | 80 | 73 | 69 |

Source: Heidenreich et al. 2011.
CHD indicates coronary heart disease; HF indicates heart failure.
* This category includes hypertension, CHD, HF, stroke, and cardiac arhythmias, rheumatic heart disease, cardiomyopathy, pulmonary heart disease, and other or ill-defined "heart" diseases. It does not include hypertension as a risk factor.
** Hypertension as Risk Factor: this category includes a portion of the costs of complications associated with hypertension, including CHF, CHD, stroke, and other CVD. The costs of hypertension as a risk factor should not be summed with other CVD conditions to calculate the costs of all CVD.

For all these reasons, CVD have received considerable attention in fields such as medicine, economics and sociology. Researchers agree that socioeconomic variables are important determinants of CVD. In fact, although the percentage of people with related pathologies - e.g., high cholesterol and obesity - is extremely high, significant disparities among certain groups clearly emerge. For instance, vulnerable groups to heart disease and stroke include old African Americans, Hispanic Americans and individuals belonging to low socioeconomic classes (Wing, 1988 and Sundquist et al., 2001). A related issue concerns the existence of a risk differential between rural and urban areas. For example, during 1985-1995, declines in mortality rates for premature coronary heart disease in African Americans and whites were slower in the rural South than their counterparts in other geographic areas (NIH, 2002). For men, the highest heart-disease related deaths occur in the South's most rural counties (Eberhardt et al., 2001). Please note that we use the terms rural vs. urban and metropolitan vs. non-metropolitan interchangeably, throughout the paper.

Figure 1 shows a county-level map of heart mortality in the U.S. over the period 2007-2009. According to this map, the highest mortality rates are observed in the Great Basin, along the border between the Interior Lowlands and the Coastal Plain, and in the Appalachian region. If we look at Figure 2, we can see that these areas are less populated than other areas such as the Coastal Range, the North-East Coastal Plain and Florida. Moreover, densely populated areas are also characterized by higher levels of income per-capita and higher fractions of physicians specialized in cardiovascular diseases. In general, such observed risk factors for cardiovascular diseases are typically more abundant in sparsely populated areas. However, two important questions remain: Is the mortality gap between urban and rural areas explained solely by observed factors? Is urbanization a Pareto improving policy in terms of heart mortality?

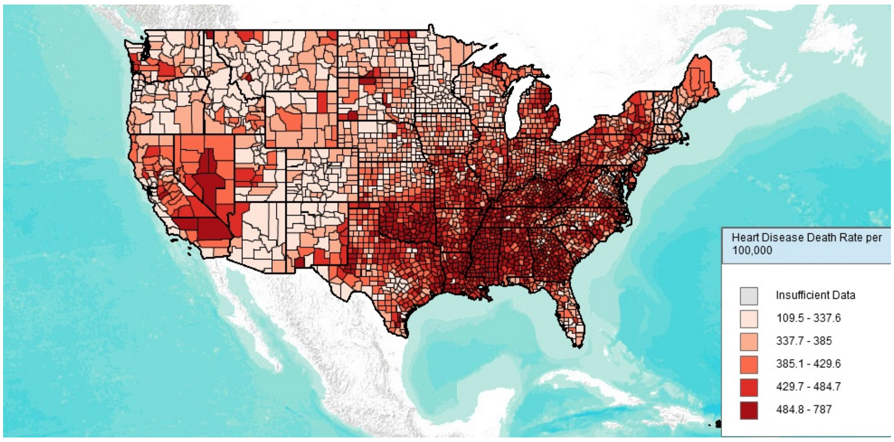


Figure 1: Map of heart disease in the U.S. (2007-2009)

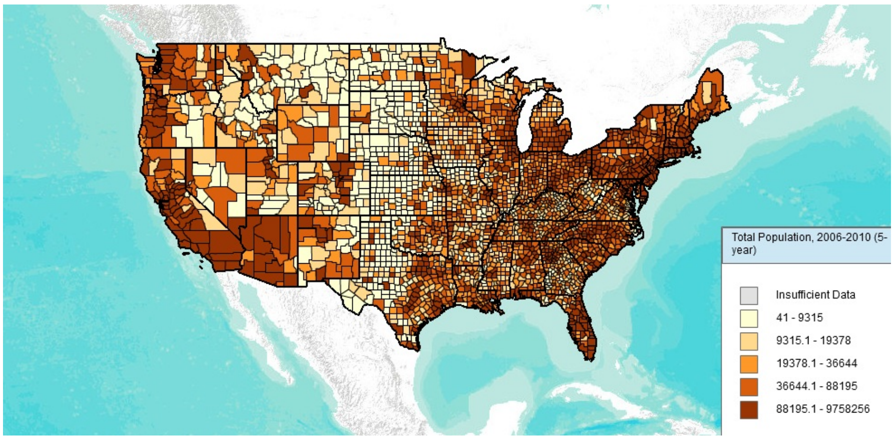


Figure 2: Total population in the U.S. (2006-2010)

Zuniga et al. (2003) argue that capacities of health care providers may heighten the disparity in heart disease and stroke incidence in metropolitan versus nonmetropolitan areas. The authors point out that rural populations have behaviors and attitudes that enhance the risk of coronary heart disease and stroke. According to them, urban lifestyles change individuals' perception of heart disease risk. This means that, ceteris paribus, an individual should face a lower risk in metropolitan areas than in rural ones only because of her perceptions.

This work aims to test these findings. In particular, by using the famous Blinder-Oaxaca decomposition method, we divide the risk gap between metropolitan and nonmetropolitan areas into three components: a component due to different endowments in risk factors between rural and urban areas, a component due to differences in the marginal impact of these risk factors, and an interaction term.[3] The second and third components are associated with the presence of factors which are not explained by the model, but interact with the explanatory variables. Following Zuniga et al. (2003), we mainly refer to these unexplained components as perceptions and attitudes, that is, unobservable lifestyles.[4]

Our study contributes towards the investigation of the cardiovascular mortality differentials between rural and urban areas. Specifically, we compile a dataset incorporating the determinants of CVD along with all types of cardiovascular diseases at a low level of regional disaggregation (county) in the US. After taking care of the common support requirement regarding rural and urban counties, we arrive at several interesting results. First, we show that traditional risk factors are predominant in rural areas, in line with the existing literature. Second, unobserved risk factors increase significantly mortality risk in urban counties, while they lower CVD mortality in rural ones. Third, our findings remain intact when we consider gender-specific mortality rates. In particular, unobserved risk factors fully compensate for the the effects of the risk factors. Fourth, when we disaggregate CVD mortality into different types, the unobserved risk factors account for the rural-urban gap in all cases, with the exception of infarction and stroke. Finally, unobserved determinants of mortality rates are particularly influential in nonmetro areas adjacent to metro ones.

In light of this evidence, the ongoing urbanization observed all over the world is predicted to have contradicting effects on the health status of the population in terms of CVD: on the one hand, socioeconomic factors tend to reduce mortality rates, whereas unobserved predictors are likely to increase them. This dilemma paves the way for policies oriented towards specific socioeconomic risk factors regarding health promotion. Such initiatives are likely to be more effective if implemented with emphasis on diverse community needs. On the basis of the health economics literature, from these policies, we can reasonably expect benefits in terms of productivity and growth. The merit of this paper is to show that unobserved predictors may change the impact of traditional risk factors on CVD mortality. Future research should explore the nature of these unobserved variables.

The rest of the paper is organized as follows. Section 2 describes the original dataset we built. Section 3 presents the decomposition technique we use to identify the role of unobserved predictors. Section 4 provides the results of our analysis and Section 5 concludes.

# 2 Data

To decompose the differences in heart mortality rates between metropolitan and nonmetropolitan areas, we built an original dataset based on several sources for U.S. counties. We apply the Metropolitan Statistical Area (MSA) definition used by the U.S. Office of Management and Budget to characterize a county as urban or rural according to a relatively high/low population density in its core

[3] Decomposition analysis can be used to study group differences in continuous and unbounded outcome variable. For example, O'Donnell et al. (2008) use it to analyze health inequalities by poverty status.

[4] The notion of perception towards CVD risk has been dealt by numerous studies in the medical literature (e.g. Foss et al. 1996; Rimal, 2001; Celentano et al., 2003; van der Weijden, 2007; Honko et al., 2008 among others). Note that our model includes observable lifestyles such as binge drinking, smoking and poor eating habits.

and the economic ties throughout the area. Alternative definitions where counties are classified according to population and adjacency are taken from the U.S. Department of Agriculture (Economic Research Service, ERS). The definitions of all the variables together with a complete list of sources are available in the Appendix.

As dependent variable we use the CVD mortality rate (per 100,000 inhabitants, aged 35 and over) for 2005-2007, due to five heart diseases, namely coronary heart disease, acute myocardial infarction, cardiac arrhythmia, heart failure and hypertension, with the addition of stroke (both ischemic and hemorrhagic). These variables are provided by National Vital Statistics System (2005). Obviously, we cannot ignore the possibility of spatial correlation between CVD due to possible dependence on spatially varying risk factors. Therefore, our dependent variable is spatially smoothed to produce more reliable estimates of mortality in each area.

Table 2 presents the descriptive statistics for our dependent variables. The last column of Table 2 reports the incidence of each disease on the overall mortality rate. As it can be seen, coronary heart disease, heart failure and hypertension are the first three causes of cardiovascular mortality.

Table 2: Dependent variables (descriptive statistics)

| Variable | Obs | Mean | Std. Dev. | Min | Max | % total |
|---|---|---|---|---|---|---|
| Heart disease 2005 | 3015 | 423.5 | 100.0 | 100.3 | 882.1 | 100 |
| Heart disease 2005 (men) | 3137 | 522.1 | 106.2 | 149.7 | 1169.8 | 61.1 |
| Heart disease 2005 (women) | 3134 | 333.1 | 76.1 | 66.5 | 749.9 | 38.9 |
| Coronary 2005 | 3138 | 270.9 | 71.5 | 76.8 | 686.2 | 27.4 |
| Infarction 2005 | 3126 | 114.3 | 55.7 | 11.9 | 462.4 | 11.5 |
| Arrhythmia 2005 | 2825 | 105.4 | 39.5 | 24.5 | 545.5 | 9.6 |
| Heart failure 2005 | 3074 | 212.3 | 57.3 | 69.5 | 745.5 | 21.1 |
| Hypertension 2005 | 3137 | 204.0 | 81.1 | 32.2 | 843.6 | 20.7 |
| Stroke 2005 | 3131 | 96.3 | 19.9 | 35 | 256.8 | 9.7 |
| Heart disease 2007 | 3014 | 399.5 | 98.4 | 138.3 | 901.5 | 100 |
| Heart disease 2007 (men) | 3135 | 494.1 | 105.5 | 105.1 | 1029.7 | 61.4 |
| Heart disease 2007 (women) | 3134 | 310.7 | 72.0 | 115.3 | 645.1 | 38.6 |

Figure 3 shows the distribution of CVD mortality in both MSA and non-MSA counties. From this figure, we can see that some rural counties exhibit CVD mortality rates that are particularly high. These values are perfectly in line with the previous evidence of a high mortality risk in rural areas.

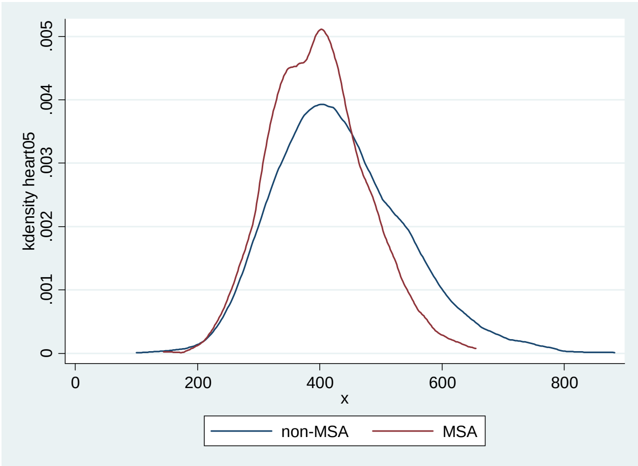


Figure 3: CVD mortality rates (MSA vs non-MSA)

Our analysis is based on 33 independent variables, which are drawn from the literature on cardiovascular diseases. These risk factors can be classified into six broad groups: demographic characteristics, economic variables, human capital factors, medical and behavioral factors, social and environmental controls. Since the reduction of CVD mortality involves lifestyle modifications, drug treatment and effective management of underlying medical conditions, we must consider both prevention and treatment variables. Therefore, we use past variables (mainly from the 1990 U.S. Census) for long-term determinants, recent variables (1990-2000) for changes in the composition of the society, and current ones (2000-2005) for ongoing risk factors.

Given the importance of examining socioeconomic and racial/ethnic disparities jointly, we consider a number of variables related to local demography (Braveman, et. al, 2009). These factors include population size, ethnicity controls (the fractions of African Americans and Asian Americans), people aged over 65 and net migration. All variables in this group are denoted in levels (1990) and changes (1990-2000); we exclude the change in net migration because of collinearity with its stock.

A second group of variables central to our research refers to local economic conditions. Previous articles suggest that not only poor and less educated people are more likely than wealthy and well-educated ones to die from CVD, but that this gap may be widening (Cooper et al., 2000). Therefore, we include the median household income, the percentage of population below the poverty line as well as their changes over 1990-2000. Since the empirical evidence suggests that income inequality and unemployment are associated with CVD mortality (see, among others, Massing et al., 2004 and Henriksson et al., 2003), we also add the unemployment rate (2005) and the Gini coefficient (2000).

As we have already mentioned, education is another important dimension that should be taken into account when we study CVD mortality. Our human capital variables include the fraction of adult population with tertiary and secondary education as well as the change in tertiary and secondary education graduates from 1990 to 2000. These changes allow us to account for the social consequences of human capital variation. Moreover, since the service sector is characterized by high job stress, the percentage of workforce employed in professional occupations and the number of programming engineers (proxied by the number of students enrolled in engineering programs in 1990) are also considered (Smith et al., 1999).

The fraction of obese individuals and diabetic medicare enrollees, together with measures of observable habits such as binge drinking, smoking and fast-food eating, are taken into account. To control for psychological diseases, we include the average number of reported mentally unhealthy days per month. Other medical variables are the number of physicians per capita, the number of primary care providers, the percentage of adults without health insurance, and the number of discharges from hospitals for ambulatory care sensitive conditions. We also incorporate health expenditure in two periods (1992 and 2002) to estimate both prevention and treatment effects on CVD (Govil et al., 2009).

Finally, an array of social variables are included, such as the fraction of married men (1990) and of married women (only for 2000 due to collinearity), together with their changes between 1990 and 2000 (Kiecolt-Glaser and Newton, 2001). The percentage of adults that report the lack of social/emotional support is used to proxy the level of social capital characterizing the area. In addition, climate and environmental factors are captured using three complementary measures: number of days in 2005 that air quality was unhealthy due to ozone, number of days in 2005 that air quality was unhealthy due to fine Particulate Matter (PM), and an index of natural amenities (Peters et al. 2000).[5]

Table 3 reports the summary statistics for all independent variables used in the article. Notice that, measures such as binge drinking, smoking, and the lack of social support reduce the sample size to 1941 counties. These variables are very important to explain CVD mortality, so they cannot be simply dropped. Therefore, our analysis must account for possible sample selection problems.[6]

[5] The natural amenities index is a measure of the physical characteristics of a county. This index reflects environmental qualities most people prefer, and it is based on the following dimensions: warm winter, winter sun, temperate summer, low summer humidity, topographic variation, and water area.

[6] When all independent variables are included, less populated counties seem to be underrepresented.

Table 3: Independent variables (descriptive statistics)

| Variable | Obs | Mean | Std. Dev. | Min | Max |
|---|---|---|---|---|---|
| | | Demography | | | |
| Population | 3142 | 88623.05 | 351389.1 | 1532 | 8863124 |
| Pop. growth | 3142 | 0.096 | 0.133 | -0.551 | 1.068 |
| Net migration 1995 | 3144 | 0.0002 | 0.005 | -0.139 | 0.073 |
| African-Americans | 3142 | 0.086 | 0.143 | 0 | 0.862 |
| Change Afr.-Am. | 3142 | 0.001 | 0.017 | -0.099 | 0.272 |
| Asians | 3139 | 0.007 | 0.025 | 0 | 0.630 |
| Change Asians | 3139 | -0.002 | 0.011 | -0.144 | 0.299 |
| Age 65+ | 3142 | 0.148 | 0.044 | 0.009 | 0.341 |
| Change age 65+ | 3142 | -0.001 | 0.014 | -0.092 | 0.085 |
| | | Economic variables | | | |
| Median income | 3141 | 24.022 | 6.643 | 8.595 | 59.284 |
| Income growth | 3139 | 0.394 | 0.084 | -0.138 | 0.752 |
| Below poverty line | 3142 | 0.167 | 0.079 | 0.022 | 0.631 |
| Change below p.l. | 3139 | -0.063 | 0.231 | -1.113 | 1.134 |
| Gini 2000 | 3143 | 0.434 | 0.038 | 0.314 | 0.605 |
| Unemployment 2005 | 3220 | 5.602 | 2.195 | 1.8 | 20.9 |
| | | Human Capital | | | |
| Tertiary edu. | 3142 | 0.136 | 0.066 | 0.037 | 0.537 |
| Secondary edu. | 3142 | 0.698 | 0.104 | 0.317 | 0.962 |
| Professionals | 3144 | 0.352 | 0.066 | 0.160 | 0.674 |
| Engineers | 3144 | 0.001 | 0.005 | 0 | 0.112 |
| Change tertiary | 3142 | -0.027 | 0.021 | -0.146 | 0.075 |
| Change secondary | 3142 | -0.188 | 0.052 | -0.324 | -0.029 |

Table 3 (cont.): Independent variables (descriptive statistics)

| Variable | Obs | Mean | Std. Dev. | Min | Max |
|---|---|---|---|---|---|
| | | Medical-behavioral factors | | | |
| Diabetics | 3047 | 79.745 | 7.736 | 29.27 | 100 |
| Drinking | 2615 | 13.623 | 5.170 | 0 | 35.3 |
| Smokers | 2458 | 22.413 | 5.931 | 0.54 | 47.62 |
| Obese | 3144 | 25.187 | 3.300 | 12.3 | 37.9 |
| Mental health | 2914 | 3.457 | 1.027 | 0.4 | 8.3 |
| Uninsured | 3144 | 17.999 | 6.089 | 7.1 | 46.8 |
| Physicians | 3142 | 0.002 | 0.003 | 0 | 0.162 |
| Primary care | 3144 | 85.553 | 59.556 | 0 | 814.751 |
| Ambulatory care | 3078 | 90.509 | 36.063 | 24.159 | 318.617 |
| Fast food exp. | 3144 | 469.772 | 65.937 | 0 | 708 |
| Health exp. 1992 | 3141 | 5157.392 | 36106.04 | 0 | 1018152 |
| Health exp. 2002 | 3141 | 11004.16 | 74848.82 | 0 | 1990013 |
| | | Social variables | | | |
| Married men 1990 | 3139 | 0.234 | 0.024 | 0.103 | 0.299 |
| Married women 2000 | 3141 | 0.228 | 0.026 | 0.089 | 0.326 |
| Change mar. men | 3139 | 0.001 | 0.018 | -0.067 | 0.209 |
| Change mar. women | 3139 | -0.003 | 0.010 | -0.067 | 0.141 |
| No social support | 2104 | 18.997 | 5.118 | 5.55 | 50.71 |
| | | Environmental variables | | | |
| Amenities | 3112 | 0.079 | 2.339 | -6.4 | 11.17 |
| Ozone days | 3113 | 2.923 | 7.374 | 0 | 110 |
| PM days | 3113 | 2.846 | 4.356 | 0 | 58 |

# 3 Risk Factors and CVD Mortality: A Decomposition Analysis

This section briefly resumes the decomposition method popularized by Blinder (1973) and Oaxaca (1973). Let $Y$ denote the CVD mortality rate. By considering metropolitan counties ($j = 1$) and nonmetropolitan counties ($j = 0$), we first estimate the following linear model:

$$Y_{i,j} = c_{i,j} + D'_{i,j}d_j + E'_{i,j}e_j + H'_{i,j}h_j + M'_{i,j}m_j + S'_{i,j}s_j + Z'_{i,j}z_j + u_{i,j}, \quad (1)$$

where $i$ is the county index, $c$ is a constant term, $H$ stands for human capital and captures educational dimensions, $D$ is a matrix of demographic characteristics, $E$ includes relevant economic variables, $M$ is a matrix of medical and behavioral factors, $S$ is a matrix of social characteristics, $Z$ is a matrix of environmental factors, and $u_{i,j} \sim N(0, \sigma_j)$ is the usual error term. Lowercase letters denote the corresponding vectors of coefficients, that is, the marginal impacts of the determinants on mortality rates. However, a simple OLS regression does not answer the following question: Is the mortality gap completely explained by differences in the endowment of risk factors?

To address this issue, we use a counterfactual approach that allows us to divide the risk gap between rural and urban areas into two parts: one "explained" by areas' differences in risk factors and an "unexplained" part containing the effects of group differences in unobserved and unobservable predictors. Any decomposition method is based on the "common support" assumption. According to this assumption, the distribution of each covariate must have the same support for both groups. However, by definition, the population size is always higher in metro areas than in nonmetro areas, that is, the common support assumption does not hold for the population level. A direct way to solve this problem, as well as for the sample selection problem mentioned above (see Reimers, 1983), is to subtract the effect of population size from the mortality differential and then decompose the adjusted differential.[7] Therefore, our results concern the decomposition of the mean difference in the mortality risk adjusted by the population effect.

By using a more compact notation for the entire set of covariates ($X_{i,j}$), except for the population size, model (1) can be rewritten as follows:

$$\widetilde{Y}_{i,j} = Y_{i,j} - \beta_{pop} Pop = X'_{i,j}\beta_j + u_{i,j}, \tag{2}$$

where $\beta_j$ contains the slope parameters and the intercept. We can express the mean difference in the adjusted mortality risk as the difference in the linear predictions at the group-specific means of the regressors. From (2) and the assumption that $E(u_{i,0}) = E(u_{i,1}) = 0$, we have:

$$E(\widetilde{Y}_{i,1}) - E(\widetilde{Y}_{i,0}) = E(X_{i,1})'\beta_1 - E(X_{i,0})'\beta_0. \tag{3}$$

Following Winsborough and Dickinson (1971), Jones and Kelley (1984) and Daymont and Andrisani (1984), equation (3) can be modified to identify the contribution of areas' characteristics to the adjusted mortality difference. In particular, it can be rearranged to obtain a "three-fold" decomposition:

$$\begin{aligned} E(\widetilde{Y}_{i,1}) - E(\widetilde{Y}_{i,0}) \quad = \quad & [E(X_{i,1}) - E(X_{i,0})]'\beta_0 + \\ & + E(X_{i,0})'(\beta_1 - \beta_0) + \\ & + [E(X_{i,1}) - E(X_{i,0})]'(\beta_1 - \beta_0) \end{aligned} \tag{4}$$

In equation (4), the first term of the right hand side represents the part of the risk gap that is due to differences in the covariates between rural and urban areas (the "effect of characteristics"). The second term captures the part of the risk gap that is due to differences in the marginal impacts of risk factors (the "effect of coefficients"). The last part is an interaction term accounting for the composite effect of differences in endowments and coefficients, since these differences operate simultaneously. We also distinguish men's mortality rates from women's mortality rates in order to analyze if results are driven by gender composition differences.

Barsky et al. (2002) and Fortin et al. (2011) have recently noted that the Oaxaca–Blinder decomposition represents a consistent estimator of the population average treatment effect on the treated. Moreover, Kline (2011) shows that the classic Blinder-Oaxaca estimator is equivalent to a propensity score

[7] The percentage of MSA counties that lie in the common support is 1.05%, whereas the percentage of non-MSA counties in the common support is 0.42%. This small coverage impeeds the use of the method proposed in Ñopo (2008).

reweighting estimator based on a linear model for the treatment odds, and satisfies a 'double robustness' property. This implies that the effect of coefficients determines a causal relationship between urbanization and mortality risk.

We conclude the analysis with a battery of robustness checks. Some of these checks are devoted to test the validity of our main conclusions, others are also intended to shed some light on a series of complementary questions. First, we drop those variables that reduce the sample size notably. In this way, many non represented counties enter the sample. Note that we are aware of the trade-off between misspecification bias and sample selection bias. Second, we use a different time period for the dependent variable in order to exclude cohort effects or other temporary phenomena affecting the results. Third, we use as dependent variables five different heart diseases, namely coronary heart disease, acute myocardial infarction, cardiac arrhythmia, heart failure and hypertension, with the addition of stroke (both ischemic and hemorrhagic). This allows us to identify if any specific diseases drive our results. Finally, we investigate whether the effects of unobserved risk factors can be attributed to specific areas. To do this, we distinguish metropolitan counties using the population size of their metro area, and nonmetropolitan counties by the degree of urbanization and adjacency to a metro area.

# 4 Results

## 4.1 Least Squares Estimates

We initially establish the correlation between regressors and regressands. Table 4 contains the least squares estimates for metropolitan and nonmetropolitan statistical areas. This table also reports separate regressions for men and women. A positive relationship between population size and heart mortality emerges only for nonmetropolitan counties. This means that agglomeration leads to negative externalities only in nonmetropolitan areas. On the contrary, the correlation between heart mortality and population growth is negative. This effect, can be explained by the fact that, controlling for net migration, the rate of population growth is related to the change in the age distribution of population. As expected, the share of African Americans is associated with higher heart mortality, which is consistent with the existing evidence showing a higher risk for this ethnic group. The share of elderly and its change are negatively correlated with mortality, since old people are more vulnerable to degenerative diseases (Repetto and Comandini, 2000) and less vulnerable to chronic diseases (Manton, 2008).

Concerning the role of economic variables, Table 4 shows that, in metropolitan counties, mortality risk is sensitive to income growth. Those metropolitan counties that experience higher economic growth also exhibit a stronger reduction of heart mortality. Table 4 also reveals the existence of a negative relationship between heart mortality and the fraction of population below the poverty line in nonmetropolitan areas. By using data on individual heart diseases, we can see that economic variables affect specific diseases, and this is why we do not find significant effects when we consider all diseases together. For instance, median income and the fraction of of population below the poverty line are strongly related to stroke and infarction (see Gillum and Mussolino, 2003).[8]

The share of population with tertiary education and its growth have strong negative effects on heart mortality in 2005. This is consistent with the lifestyle people with higher education adopt which is conducive to heart disease prevention compared to people with lower education level (Kilander et al., 2001). In urban areas, the share of men involved in professional activities has a positive impact on CVD mortality. Similarly, areas with higher fractions of programming engineers are associated with higher mortality rates. As argued by Kalimo (1999), cognitive occupations involve many health-promoting features, but the rapid increase in the demand for cognitive and non-cognitive skills as well as the emergence of new professional subcultures emphasizing excessive commitment to work may cause stress and burnout problems. Therefore, after having controlled for health-promoting features, the estimated coefficients reflect this second channel.

Table 4 also suggests that the heart mortality risk increases with the number of smokers, especially in nonmetropolitan areas. Indeed, smoking is a major cause of coronary artery disease. Similarly, obesity is positively correlated to heart diseases. At first glance, a negative correlation between heart mortality and the number of diabetic patients might be surprising. However, this relationship is not so awkward if we consider that chronic conditions (diabetes and heart disease) imply drinking and eating restrictions.

Because of the high level of saturated or trans fats present in much of the fast food diet, the expenditure in fast foods is positively correlated with the heart mortality for individuals living in nonmetropolitan areas. Public health expenditure in 1992 reduces mortality since it can facilitate disease prevention, while expenditure in 2002 boosts mortality, because it is associated with the cure of people with heart diseases.

In nonmetropolitan areas, the number of primary care providers is negatively related with mortality rates. This evidence emphasizes the need of medical infrastructure in rural counties. Vice versa, the number of discharged patients from hospitals for ambulatory care sensitive conditions is positively associated with mortality rates. Finally, mental diseases are associated with heart mortality in rural counties, whereas urban centers suffer from high levels of air pollution. Overall, our variables explain 58-66% of the total variation in heart death rates, which is satisfactory given that we use cross-section data.

[8] The marginal impact of poverty on stroke is $-76.9$ with a probability value equals to 0.004. As expected, we have also found a negative effect of poverty on the infarction rate, although the significance level is at 10%. Detailed results are available from the authors upon request.

Table 4: OLS regression (heart mortality, 2005)

| | Non-MSA | MSA | Non-MSA | MSA | Non-MSA | MSA |
|---|---|---|---|---|---|---|
| | All | All | Men | Men | Women | Women |
| Constant | 425.633*** | 123.254 | 763.200*** | 268.395* | 269.935*** | 20.139 |
| | (110.921) | (124.213) | (119.384) | (140.493) | (75.050) | (109.734) |
| Population | 0.000** | 0 | 0.000*** | 0 | 0.000*** | 0 |
| | (0.000) | (0.000) | (0.000) | (0.000) | (0.000) | (0.000) |
| Pop. growth | -154.143*** | -28.581 | -172.507*** | -84.909** | -98.595*** | -21.571 |
| | (42.633) | (40.743) | (39.461) | (33.269) | (31.707) | (34.644) |
| Net migration | 3,569.50 | 303.864 | 6,567.91 | 469.722 | -3,273.38 | 183.371 |
| | (3,976.743) | (277.953) | (4,286.051) | (329.327) | (2,976.727) | (240.845) |
| African Am. | 109.698*** | 152.839*** | 111.078*** | 178.420*** | 108.021*** | 127.635*** |
| | (32.031) | (31.670) | (33.310) | (35.814) | (23.194) | (28.673) |
| Ch. Afr. Am. | 467.063** | -99.208 | 591.489*** | -73.793 | 668.261*** | -79.852 |
| | (209.359) | (117.932) | (218.962) | (134.493) | (158.509) | (103.728) |
| Asians | -789.398 | 165.36 | -717.875 | 161.488 | -515.324 | 247.002** |
| | (568.619) | (134.157) | (589.809) | (141.178) | (476.675) | (120.828) |
| Ch. Asians | 1,472.186* | 377.01 | 722.118 | 386.376 | 1,197.285* | 543.610** |
| | (876.562) | (257.215) | (857.288) | (280.029) | (662.916) | (233.590) |
| Aged 65+ | -231.255* | -286.520** | -30.599 | -266.513*** | -291.102*** | -323.619*** |
| | (129.889) | (112.057) | (124.378) | (100.495) | (98.180) | (98.456) |
| Ch. Aged 65+ | -890.483*** | -798.516** | -577.441** | -899.000*** | -593.275*** | -676.630** |
| | (289.255) | (344.713) | (256.350) | (285.493) | (221.194) | (297.741) |
| Median income | -1.658 | 0.85 | -2.621* | -0.612 | -1.348 | 1.078 |
| | (1.361) | (0.911) | (1.384) | (0.966) | (0.986) | (0.803) |
| Below p.l. | -198.724** | 136.819 | -251.623** | 113.602 | -45.498 | 155.358 |
| | (100.735) | (115.088) | (110.054) | (127.541) | (70.559) | (99.600) |
| Gini 2000 | 86.054 | 280.366** | -8.966 | 199.023 | 73.513 | 298.125*** |
| | (119.490) | (123.670) | (123.205) | (130.008) | (81.373) | (113.259) |
| Unemp. 2005 | 1.811 | -2.301 | 5.305*** | 1.212 | 0.62 | -2.087 |
| | (1.723) | (2.759) | (1.769) | (3.040) | (1.304) | (2.101) |
| Income growth | 20.267 | -186.051*** | -38.003 | -262.705*** | 19.864 | -170.120*** |
| | (43.406) | (60.493) | (46.162) | (62.906) | (30.359) | (51.456) |
| Ch. Below p.l. | 0.239 | -4.81 | -10.353 | 4.099 | 11.147 | -7.838 |
| | (18.420) | (20.033) | (18.205) | (21.799) | (12.788) | (15.858) |
| Tertiary edu. | -153.818 | -525.907*** | -155.33 | -650.825*** | -21.205 | -399.312*** |
| | (121.720) | (153.613) | (129.146) | (169.912) | (87.925) | (139.450) |
| Ch. Ter. Edu. | -446.035** | -594.743** | -505.577** | -604.423** | -142 | -363.414 |
| | (208.260) | (275.227) | (214.372) | (301.661) | (145.838) | (248.639) |
| Secondary edu. | -29.405 | 112.661 | -201.899** | 133.668 | -82.069 | 174.014 |
| | (98.650) | (128.699) | (99.715) | (139.029) | (69.909) | (106.920) |
| Ch. Sec. Edu. | 141.511 | 365.698 | 66.486 | 504.472** | -91.821 | 318.184 |
| | (154.798) | (236.011) | (150.168) | (239.815) | (109.808) | (195.124) |
| Professionals | -117.251 | 139.127 | -22.928 | 348.169*** | -88.813 | 87.401 |
| | (81.462) | (109.028) | (76.495) | (117.635) | (54.047) | (93.367) |
| Prog. Engineers | 512.051* | 644.756*** | 824.175** | 501.414* | 496.147** | 670.580*** |
| | (282.669) | (247.989) | (328.952) | (285.732) | (197.617) | (224.235) |

Robust standard errors in parentheses. Significance level: * 10%; ** 5%; *** 1%

Table 4 (cont.): OLS regression (heart mortality, 2005)

| | Non-MSA | MSA | Non-MSA | MSA | Non-MSA | MSA |
|---|---|---|---|---|---|---|
| | All | All | Men | Men | Women | Women |
| Drinkers | -0.398 | -0.898 | -0.053 | -0.662 | -0.628* | -1.096** |
| | (0.512) | (0.591) | (0.526) | (0.658) | (0.365) | (0.506) |
| Smokers | 1.250** | 0.736 | 1.335** | 1.404* | 0.780** | 0.057 |
| | (0.526) | (0.674) | (0.555) | (0.720) | (0.389) | (0.604) |
| Obese | 2.097** | 2.049** | 1.492 | 2.749** | 1.928*** | 1.711* |
| | (0.953) | (1.034) | (0.995) | (1.156) | (0.706) | (0.923) |
| Diabets | -0.231 | -1.561*** | -0.322 | -2.225*** | -0.102 | -1.289*** |
| | (0.316) | (0.572) | (0.312) | (0.662) | (0.213) | (0.490) |
| Uninsured | 0.261 | -1.091 | 0.25 | -0.865 | 0.5 | -0.586 |
| | (0.579) | (0.680) | (0.590) | (0.757) | (0.390) | (0.617) |
| Primary care | -0.192*** | 0.001 | -0.123** | 0.141 | -0.142*** | -0.038 |
| | (0.052) | (0.079) | (0.055) | (0.090) | (0.035) | (0.068) |
| Ambulatory | 0.775*** | 1.059*** | 0.600*** | 0.954*** | 0.507*** | 0.924*** |
| | (0.099) | (0.144) | (0.089) | (0.172) | (0.066) | (0.126) |
| Mental | 13.094*** | 4.848 | 14.463*** | 2.512 | 12.311*** | 4.627 |
| | (2.981) | (3.793) | (2.923) | (4.249) | (2.101) | (3.314) |
| Physicians | 5,159.60 | -698.757 | 7,707.506** | -2,774.89 | 2,899.86 | -172.784 |
| | (3,235.880) | (1,727.885) | (3,416.265) | (1,944.402) | (2,299.067) | (1,642.452) |
| Fast food exp. | 0.059* | 0.016 | 0.133*** | 0.044 | 0.067*** | -0.006 |
| | (0.033) | (0.040) | (0.033) | (0.044) | (0.024) | (0.031) |
| Health exp. 92 | -124.406** | -93.63 | -87.773 | -179.463** | -118.300*** | -74.501 |
| | (59.249) | (61.176) | (67.461) | (71.577) | (44.097) | (51.367) |
| Health exp. 02 | 67.668** | 47.014* | 39.237 | 112.803*** | 59.926*** | 19.199 |
| | (26.933) | (26.315) | (31.425) | (29.769) | (18.988) | (21.821) |
| Mar. men 1990 | 128.062 | 287.763 | -492.849 | 144.669 | 166.521 | 70.81 |
| | (447.672) | (410.632) | (466.826) | (434.807) | (314.499) | (343.008) |
| Mar. women 2000 | -43.843 | 364.443 | 256.071 | 614.432 | 395.917 | 606.744 |
| | (468.835) | (508.229) | (484.956) | (480.138) | (319.943) | (427.757) |
| Ch. Mar. Men | -106.917 | 406.903 | -164.26 | 529.870* | 21.195 | 223.118 |
| | (190.540) | (258.529) | (193.461) | (307.546) | (132.642) | (230.767) |
| Ch. Mar. Wom. | 375.538 | -673.274 | -33.567 | -585.989 | 117.027 | -595.718 |
| | (616.420) | (580.442) | (619.878) | (567.934) | (418.577) | (499.991) |
| No social sup. | 0.305 | 0.871 | -0.301 | 0.557 | -0.576 | 0.87 |
| | (0.555) | (0.684) | (0.561) | (0.765) | (0.394) | (0.598) |
| Amenities | 0.219 | 0.075 | -1.69 | -0.211 | -0.097 | -0.171 |
| | (1.362) | (1.290) | (1.482) | (1.427) | (0.976) | (1.084) |
| Ozone days | -0.02 | 0.401** | -0.835 | 0.437** | 0.469 | 0.389*** |
| | (0.513) | (0.163) | (0.600) | (0.178) | (0.439) | (0.149) |
| PM days | -0.056 | 0.398 | 1.024 | 0.48 | -0.156 | 0.505* |
| | (0.693) | (0.316) | (0.728) | (0.329) | (0.576) | (0.275) |
| Latitude | -2.340*** | 1.487* | -2.633*** | 1.852* | -2.731*** | 1.056 |
| | (0.757) | (0.871) | (0.773) | (0.983) | (0.545) | (0.717) |
| Longitude | -0.128 | 0.042 | 0.278 | 0.257 | 0.178 | 0.146 |
| | (0.258) | (0.312) | (0.277) | (0.322) | (0.189) | (0.264) |
| Observations | 1187 | 748 | 1209 | 748 | 1209 | 748 |
| R-squared | 0.61 | 0.63 | 0.64 | 0.65 | 0.66 | 0.58 |

Robust standard errors in parentheses. Significance level: * 10%; ** 5%; *** 1%

## 4.2 Decomposition Analysis: Main Results

In this section, we present our decomposition analysis. Table 5 provides the Blinder-Oaxaca decomposition described by equation (4). Estimates and standard errors are reported in columns 1 and 2, respectively. The predicted mortality risk is lower in metropolitan counties than in nonmetropolitan ones. The unadjusted mortality differential is about 28 deaths per 100,000 inhabitants, whereas the adjusted differential is 19.7 deaths per 100,000 inhabitants. In line with the existing literature, the endowment of risk factors is lower in urban centers than in rural counties. This effect leads to a mortality differential of 40.2 deaths per 100,000 inhabitants. On the contrary, unobserved risk factors generate a risk differential favoring rural areas. In particular, a mortality differential of -36.3 comes from the existence of significant differences in the impact of observed risk factors. The main variables responsible for the sign of the unexplained component are: population, median income, change in secondary education, the number of primary care providers, and air pollution due to PM.[9] The remaining interaction effect is not statistically significant. However, given the magnitude of the adjusted differential, we know that the combination of differences in characteristics and coefficients partially compensates the effect of coefficients. The interaction effect could be attributed to the first two components by using a two-fold decomposition. In this case, we would obtain an explained effect of 30.7 deaths per 100,000 inhabitants and an unexplained effect of -11 deaths per 100,000 inhabitants (both effects are statistically significant).[10]

Columns 3-6 propose the same decomposition of columns 1 and 2 for men and women, separately. Given the lack of information on gender-specific risk factors, we implicitly assume that men and women share the same risk factors. By looking at the R-squared reported at the end of Table 4, we can say that this assumption seems to be rather sensible, especially for rural counties. Table 5 shows that, conditional on our specification, results are qualitatively the same for both men and women. Notice that, for women, the adjusted differential is not significant. This means that, for women, the unexplained component almost neutralizes the relative abundance of risk factors in rural areas. On the contrary, for men, the unexplained component is significantly lower than the explained one. Therefore, we can conclude that risk factors in rural areas are particularly dangerous for men.

[9] This result comes from a detailed decomposition of the adjusted mortality differential. These estimates and the corresponding inference are very sensitive to measurement errors and scale effects, therefore, we avoid any possible speculation based on these results. However, the detailed decomposition for column 1 of Table 4 can be found in Appendix B.

[10] We have also clustered the errors at the State level to control for specific regulations or medical plans affecting the distribution of mortality. However, results do not change.

Table 5: Decomposition (Heart mortality, 2005-2007)

| | Total | | Men | | Women | |
|---|---|---|---|---|---|---|
| | Predicted | S.E. | Predicted | S.E. | Predicted | S.E. |
| Differential | | | | | | |
| Non MSA | 420.761*** | 2.999 | 522.123*** | 3.226 | 330.764*** | 2.317 |
| MSA | 392.759*** | 2.774 | 493.497*** | 3.230 | 317.650*** | 2.292 |
| Difference | 28.002*** | 4.085 | 28.626*** | 4.565 | 13.114*** | 3.259 |
| Adjusted | 19.665*** | 5.795 | 19.337*** | 6.257 | 3.252 | 4.467 |
| Decomposition | | | | | | |
| Endowments | 40.215*** | 5.898 | 52.259*** | 6.760 | 26.713*** | 5.096 |
| Coefficients | -36.339*** | 12.859 | -34.240** | 13.588 | -28.411*** | 9.414 |
| Interaction | 15.788 | 11.307 | 1.317 | 12.235 | 4.950 | 8.631 |
| Obs. | 1935 | | 1957 | | 1957 | |
| Significance level: * 10%; ** 5%; *** 1%. | | | | | | |

## 4.3 Robustness Checks and Additional Results

This section provides the results of a variety of robustness checks. We first drop those explanatory variables that considerably reduce the sample size, namely binge drinking, smoking, and the lack of social support. By looking at Table 6, we can see that a larger sample leads to a larger difference between the explained and the unexplained components. Although the price of reducing a sample selection bias is higher misspecification bias, we can say that, at least qualitatively, the signs of the estimated components and their standard errors confirm our main results. Moreover, the effect of coefficients seems to be extremely stable. On the one hand, this means that the impact of binge drinking, smoking, and the lack of social support on heart mortality does not change significantly between rural and urban counties. On the other hand, by removing these variables, we include in the sample high mortality (rural) counties.

Table 6: Decomposition (Heart mortality, 2005-2007)

| | Total | | Men | | Women | |
|---|---|---|---|---|---|---|
| | Predicted | S.E. | Predicted | S.E. | Predicted | S.E. |
| Differential | | | | | | |
| Non MSA | 433.513*** | 2.413 | 534.854*** | 2.531 | 339.752*** | 1.829 |
| MSA | 398.159*** | 2.704 | 499.097*** | 3.075 | 320.965*** | 2.158 |
| Difference | 35.355*** | 3.624 | 35.758*** | 3.983 | 18.787*** | 2.828 |
| Adjusted | 29.833*** | 4.894 | 28.328*** | 5.152 | 11.263*** | 3.714 |
| Decomposition | | | | | | |
| Endowments | 54.042*** | 5.699 | 65.925*** | 6.360 | 35.159*** | 4.751 |
| Coefficients | -31.906*** | 10.715 | -33.917*** | 10.757 | -26.582*** | 7.656 |
| Interaction | 7.697 | 9.927 | -3.680 | 10.340 | 2.686 | 7.441 |
| Obs. | 2772 | | 2809 | | 2809 | |
| Significance level: * 10%; ** 5%; *** 1%. | | | | | | |

The next check aims to exclude the hypothesis that our results depend on the time period we have selected. In fact, cross-section analyses may be affected by temporary phenomena such as cohort effects. Since most of the covariates originate from decennial census data, we can only change the dependent variable. Table 7 reports our decomposition findings when we use the average heart mortality rate as dependent variable over the period 2007-2009. Again, rural counties benefit from a significant unexplained component, while urban areas exhibit lower endowments of risk factors.

Table 7: Decomposition (Heart mortality, 2007-2009)

| | Total | | Men | | Women | |
|---|---|---|---|---|---|---|
| | Predicted | S.E. | Predicted | S.E. | Predicted | S.E. |
| Differential | | | | | | |
| Non MSA | 397.483*** | 2.897 | 496.065*** | 3.154 | 308.156*** | 2.173 |
| MSA | 365.299*** | 2.681 | 460.581*** | 3.137 | 292.589*** | 2.181 |
| Difference | 32.185*** | 3.948 | 35.484*** | 4.449 | 15.567*** | 3.079 |
| Adjusted | 26.337*** | 5.708 | 27.747*** | 6.155 | 7.005 | 4.293 |
| Decomposition | | | | | | |
| Endowments | 46.130*** | 5.694 | 57.747*** | 6.582 | 31.786*** | 4.854 |
| Coefficients | -27.239*** | 12.894 | -29.230*** | 13.333 | -24.066*** | 9.244 |
| Interaction | 7.445 | 11.215 | -0.770 | 11.928 | -0.716 | 8.395 |
| Obs. | 1938 | | 1956 | | 1957 | |
| Significance level: * 10%; ** 5%; *** 1%. | | | | | | |

In order to investigate whether different types of heart disease behave differently, Table 8 reports the results for six different types of CVD, namely, coronary heart disease, acute myocardial infarction, cardiac arrhythmia, heart failure, hypertension, and stroke. This table confirms the validity of our main findings. With the exception of infarction and stroke, the effect of coefficients is negative and significant for all cardiovascular diseases. Unobserved predictors are extremely important for hypertension and coronary heart disease. Infarction and stroke are the only heart diseases in which unobserved predictors do not affect the mortality gap. In particular, for myocardial infarction and stroke, the impact of economic variables such as the share of individuals below the poverty line and the fraction of professionals on the mortality differential do not significantly change between rural and urban counties.[11]

[11] As before, these findigs originate from detailed decomposition. Results are available from the authors upon request.

Table 8: Decomposition for different heart diseases (2005-2007)

| Types | Coronary | | Infarction | | Arrhythmia | |
|---|---|---|---|---|---|---|
| | Predicted | S.E. | Predicted | S.E. | Predicted | S.E. |
| Differential | | | | | | |
| Non MSA | 271.295*** | 2.121 | 114.295*** | 1.618 | 106.516*** | 1.350 |
| MSA | 256.229*** | 2.287 | 89.603*** | 1.380 | 105.714*** | 1.328 |
| Difference | 15.066*** | 3.119 | 24.693*** | 2.127 | 0.802 | 1.894 |
| Adjusted | 9.755** | 4.700 | 26.105*** | 3.548 | -8.783*** | 3.396 |
| Decomposition | | | | | | |
| Endowments | 37.081*** | 5.345 | 29.296*** | 3.384 | 4.211 | 3.319 |
| Coefficients | -23.856** | 10.842 | 0.304 | 9.075 | -19.677** | 8.546 |
| Interaction | -3.471 | 9.809 | -3.496 | 7.832 | 6.682 | 7.410 |
| Obs. | 1957 | | 1957 | | 1938 | |
| | | | | | | |
| Types | Failure | | Hypertension | | Stroke | |
| | Predicted | S.E. | Predicted | S.E. | Predicted | S.E. |
| Differential | | | | | | |
| Non MSA | 217.069*** | 1.809 | 203.194*** | 2.676 | 96.122*** | 0.597 |
| MSA | 188.335*** | 1.653 | 205.661*** | 2.525 | 90.801*** | 0.638 |
| Difference | 28.735*** | 2.451 | -2.467 | 3.680 | 5.321*** | 0.873 |
| Adjusted | 18.782*** | 3.883 | -16.028*** | 5.901 | 3.758*** | 1.339 |
| Decomposition | | | | | | |
| Endowments | 23.447*** | 3.750 | 12.405** | 6.084 | 6.407*** | 1.395 |
| Coefficients | -20.079** | 9.461 | -26.856* | 15.688 | -3.565 | 3.262 |
| Interaction | 15.414* | 8.236 | -1.578 | 14.129 | 0.916 | 2.897 |
| Obs. | 1930 | | 1957 | | 1956 | |
| Significance: * 10%; ** 5%; *** 1%. | | | | | | |

Jones and Goza (2008) underline the need for separate CVD analyses for rural, suburban and urban residents. Therefore, by using the Rural-Urban Continuum Code (2000) provided by the Economic Research Service of the U.S. Department of Agriculture, we divide rural areas into two groups: rural areas with at least 20,000 inhabitants (big non-MSA), and rural areas with less than 20,000 inhabitants (small non-MSA). We also distinguish rural areas that are adjacent to metropolitan counties and rural areas that do not border with metropolitan areas. Columns 1 and 2 of Table 9 show that the mortality risk is higher in small rural areas than in urban centers, despite the compensating effect of unobserved risk factors. In contrast, columns 3 and 4 show that the adjusted mortality differential between MSA and big non-MSA is not significant. This means that after having excluded the positive effect of population on mortality in big rural areas, these counties are equally dangerous with urban ones. Given this evidence, we can say that observed risk factors are more abundant in big rural areas than urban centers, but this disadvantage is largely compensated by the effect of unobserved risk factors.

Table 9: Small and Big non-MSA (2005-2007)

| Ref. group | Small non-MSA | | Big non-MSA | |
|---|---|---|---|---|
| | Predicted | S.E. | Predicted | S.E. |
| Differential | | | | |
| Ref. group | 422.393*** | 3.420 | 411.355*** | 4.934 |
| MSA | 391.326*** | 2.846 | 392.197*** | 2.774 |
| Difference | 31.067*** | 4.449 | 19.157*** | 5.661 |
| Adjusted | 22.288*** | 6.368 | 5.861 | 10.499 |
| Decomposition | | | | |
| Endowments | 45.764*** | 6.798 | 25.085*** | 5.317 |
| Coefficients | -44.178*** | 16.564 | -34.872** | 17.345 |
| Interaction | 20.701 | 14.808 | 15.648 | 12.147 |
| Obs. | 1662 | | 1134 | |
| Significance level: * 10%; ** 5%; *** 1%. | | | | |

Finally, in Table 10, we test whether unobserved risk factors can be attributed to proximity. For instance, living in a small village close to an urban area may be less stressful than living in crowded metropolitan centers. Columns 1 and 2 of Table 10 show that nonmetro areas adjacent to metro ones enjoy beneficial effects from unobserved risk factors. In contrast, columns 3 and 4 show that the effect of unobserved risk factors on heart mortality is smaller and less significant in nonadjacent rural areas. Moreover, this effect disappears if we perform a two-fold decomposition. An interesting topic for future research could be the differences between living in metropolitan counties vs neighboring nonmetropolitan ones.

Table 10: Adjacent and non-Adjacent non-MSA (2005-2007)

| Ref. group | Adj. non-MSA | | Non-adj. non-MSA | |
|---|---|---|---|---|
| | Predicted | S.E. | Predicted | S.E. |
| Differential | | | | |
| Ref. group | 424.570*** | 3.723 | 413.214*** | 4.264 |
| MSA | 392.664*** | 2.776 | 390.824*** | 2.844 |
| Difference | 31.906*** | 4.644 | 22.390*** | 5.126 |
| Adjusted | 21.146*** | 7.186 | 17.002** | 7.300 |
| Decomposition | | | | |
| Endowments | 40.373*** | 5.576 | 38.690*** | 7.184 |
| Coefficients | -41.200*** | 15.255 | -34.656* | 18.696 |
| Interaction | 21.973* | 12.346 | 12.967 | 16.216 |
| Obs. | 1447 | | 1349 | |
| Significance level: * 10%; ** 5%; *** 1%. | | | | |

# 5 Conclusion

By using U.S. county-level data and the famous Blinder-Oaxaca decomposition, this paper investigates the role of observed and unobserved risk factors in explaining rural-urban differences in cardiovascular mortality. After having adjusted the mortality gap for differences in population levels between rural and urban counties, we have divided the estimated difference in mortality into three components: a component due to differences in the endowments of risk factors, a component due to discrepancies in the marginal impacts of these factors, and an interaction term. We arrived at several interesting results. First, observed risk factors are relatively more abundant in rural areas, in line with the recent medical literature. With some differences, this result holds when we replace overall mortality with male and female mortality rates. Second, rural areas are characterized by a favorable impact regarding unobserved risk factors. Third, when we distinguish between different types of CVD mortality, the unobserved risk factors account for the rural-urban gap in most instances, with the exception of infarction and stroke. Fourth, after having controlled for population effects, densely populated rural counties are not more dangerous than urban ones. Finally, unobserved risk factors are less present in nonmetro areas neighboring with metro ones.

This analysis has some limitations. For instance, by assuming the invariance of coefficients for the construction of the counterfactual component, decomposition methods inherently follow a partial equilibrium approach. That is, we cannot reject the hypothesis that a change in the observed risk factors will influence the unexplained component too. Moreover, the Blinder-Oaxaca decomposition only investigates differences in the mean of an outcome variable.

Despite these limitations, the merit of this paper is to show that unobserved predictors may change the impact of traditional risk factors on CVD mortality. Future research should explore the nature of these unobserved variables.

## References


[1] Barsky, R., Bound, J., Charles, K. K. and Lupton, J. P. (2002). 'Accounting for the blackwhite wealth gap: a nonparametric approach', Journal of the American Statistical Association, Vol. 97, pp. 663-673.

[2] Blinder A. 1973. Wage discrimination: Reduced form and structural variables. Journal of Human Resources 8(4): 436-55.

[3] Braveman P, Cubbin C, Egerter S, Williams DR, Pamuk E. 2010. Socioeconomic Disparities in Health in the United States: What the Patterns Tell Us. American Journal of Public Health 100(S1): S186-S196.

[4] Celentano A., Panico S., Palmieri V., Guillaro B., Brancati C., Di Palma EspositoN., Arezzi E., Setola C., Piccinocchi G., Canfora G., Rubba P., Di Minno G., Mancini M. 2003. Citizens and family doctors facing awareness and management of traditional cardiovascular risk factors: Results from the global cardiovascular risk reduction project. Nutrition, Metabolism & Cardiovascular Diseases, 13, 211-217.

[5] Centers for Disease Control and Prevention (CDC). Preventing heart disease and stroke. Addressing the nation's leading killers, 2002.

[6] Cooper R., Cutler J., Desvigne-Nickens P., Fortmann S.P, Friedman L., Havlik R., Hogelin G., Marler J., McGovern P., Morosco G., Mosca L. Pearson T., Stamler J., Stryer D., Thom T. 2000. Trends and Disparities in Coronary Heart Disease, Stroke, and Other Cardiovascular Diseases in the United States. Circulation, 102, 3137-3147.

[7] Daymont, T. N., and P. J. Andrisani. 1984. Job P, College Major, and the Gender Gap in Earnings. The Journal of Human Resources 19: 408–428.

[8] Downes T, Greenstein S. 2007. Understanding Why Universal Service Obligations May Be Unnecessary: The Private Development of Local Internet Access Markets. Journal of Urban Economics 62(1): 2–26.

[9] Fortin, N., Lemieux, T. and Firpo, S. (2011). 'Decomposition methods in economics', in Ashenfelter O. and Card D. (eds), Handbook of Labor Economics, Vol. 4A, Elsevier, San Diego–Amsterdam, pp. 1–102.

[10] Foss F.A., Dickinson E., Hills M., Thomson A., Wilson V., Ebrahim S. 1996. Missed opportunities for the prevention of cardiovascular disease among British hypertensives in primary care, British Journal of General Practice, 46, 571-575.

[11] Gillum RF, Mussolino ME. "Education, poverty, and stroke incidence in whites and blacks: the NHANES I Epidemiologic Follow-up Study." J Clin Epidemiol. 2003 Feb;56(2):188-95.

[12] Govil SR, Weidner G, Merritt-Worden T, Ornish, D. 2009. Socioeconomic Status and Improvements in Lifestyle, Coronary Risk Factors, and Quality of Life: The Multisite Cardiac Lifestyle Intervention Program. American Journal of Public Health 99(7): 1263-1270.

[13] Heidenreich PA, Trogdon JG, Khavjou OA, et al. Forecasting the future of cardiovascular disease in the United States: a policy statement from the American Heart Association. Circulation. 2011;123:933-44. Epub 2011 Jan 24.

[14] Henriksson KM, Lindblad U, Agren B, Nilsson-Ehle P, R\aa stam L., 2003. Associations between unemployment and cardiovascular risk factors varies with the unemployment rate: the Cardiovascular Risk Factor Study in Southern Sweden (CRISS). Scand J Public Health, 31(4), 305-11.

[15] Homko C.J., Santamore W.P., Zamora L., Shirk G., Gaughan J., Cross R., Kashem A. , Petersen S., Bove A.A. 2008. Cardiovascular disease knowledge and risk perception among underserved individuals at increased risk of cardiovascular disease. Journal of Cardiovascular Nursing, 23(4), 332-337.

[16] Jones, A.., and Goza F. 2008. Rural, suburban and urban differences in the self-diagnosis of coronary heart disease in the United States. J.biosoc.Sci, 40: 895–909.

[17] Jones, F. L., and J. Kelley. 1984. Decomposing Differences Between Groups. A Cautionary Note on Measuring Discrimination. Sociological Methods and Research 12: 323–343.

[18] Kalimo, R. 1999. Knowledge jobs–how to manage without burnout?. Scandinavian Journal of Work, Environment & Health, 25(6): 605-609.

[19] Kiecolt-Glaser JK, Newton TL. 2001. Marriage and health: His and hers. Psychological Bulletin 127(4): 472-503.

[20] Kilander L, Berglund L, Boberg M, Vessby B, Lithell H. 2001. Education, Lifestyle Factors, and Mortality from Cardiovascular Disease and Cancer. A 25-year follow-up of Swedish 50-year-old men. International Journal of Epidemiology 30: 1119-1126.

[21] Kline, P. (2011). 'Oaxaca-Blinder as a reweighting estimator', American Economic Review, Vol. 101, pp. 532–537.

[22] Levine R, Renelt D. 1992. A sensitivity analysis of cross-country growth regressions. American Economic Review 82: 942–963.

[23] Lorentzen P, McMillan J, Wacziarg R. 2008. Death and development. Journal of Economic Growth 13: 81–124.

[24] Manton KG. 2008. Recent Declines in Chronic Disability in the Elderly U.S. Population: Risk Factors and Future Dynamics. Annual Review of Public Health 29: 91-113.

[25] Massing MW, Rosamond WD, Wing SB, Suchindran CM, Kaplan BH, Tyroler HA., 2004. Income, income inequality, and cardiovascular disease mortality: relations among county populations of the United States, 1985 to 1994. South Med J., 97(5), 475-84.

[26] National Institutes of Health and National Heart, Lung, and Blood Institute. Morbidity and Mortality: 2002 Chart Book on Cardiovascular, Lung, and Blood Diseases. U.S. Department of Health and Human Services, Public Health Service, National Institutes of Health, May 2002.

[27] Ñopo H. 2008. Matching as a tool to decompose wage gaps. Review of Economics and Statistics, 90, 290-299.

[28] Oaxaca R. 1973. Male-Female Wage Differentials in Urban Labor Markets. International Economic Review 14(3): 693-709.

[29] O'Donnell O, van Doorslaer E, Wagstaff A, Lindelow M. 2008. Analyzing Health Equity Using Household Survey Data: A Guide to Techniques and Their Implementation. Washington, DC: TheWorld Bank.

[30] Peters A, Liu E, Verrier R, Richard et al. 2000. Air Pollution and Incidence of Cardiac Arrhythmia. Epidemiology 11(1): 11-17.

[31] Reimers, C. W. 1983. Labor Market Discrimination Against Hispanic and Black Men. The Review of Economics and Statistics 65: 570–579.

[32] Repetto L, Comandini D. 2000. Cancer in the Elderly: Assessing Patients for Fitness. Critical Reviews in Oncology/Hematology 35: 155–160.

[33] Rimal, R.N. 2001. Perceived Risk and Self-Efficacy as Motivators: Understanding Individuals' Long-Term Use of Health Information, Journal of Communication, December, 633-654.

[34] Roger VL, Go AS, Lloyd-Jones DM, et al. 2012. Heart disease and stroke statistics—2012 update: a report from the American Heart Association. Circulation. 125(1):e2–220.

[35] Sala-i-Martin X, Doppelhofer G, Miller RI. 2004. Determinants of long-term growth: a Bayesian averaging of classical estimates (BACE) approach. American Economic Review 94(4): 813–835.

[36] Smith MJ, Conway FT, Karsh B-T. Occupational stress in human computer interaction. Indust Health 1999; 37: 157-173.

[37] Suhrcke M, Rocco L. 2008. Microeconomic consequences of non-communicable diseases at the individual or household level. World Health Organization, mimeograph.

[38] Suhrcke M, Urban D. 2010. Are cardiovascular diseases bad for economic growth? Health Economics 19: 1478–1496

[39] Sundquist J, Winkleby MA, Pudaric S. 2001. Cardiovascular disease risk factors among older black, Mexican-American, and white women and men: An analysis of NHANES III, 1988-1994. Third National Health and Nutrition Examination Survey. Journal of the American Geriatrics Society 49:109-116.

[40] Swift R. 2011. The relationship between health and GDP in OECD countries in the very long run. Health Economics 20: 306–322.

[41] United States, Bureau of Labor Statistics. ftp://ftp.bls.gov/pub/special.requests/cew/ (accessed June 6, 2012)

[42] United States, Census Bureau. 2011. 1990 Decenial Census. http://www.census.gov/main/www/ cen1990.html (accessed March 25, 2012).

[43] United States, Census Bureau. 2011. 2000 Decenial Census. http://www.census.gov/main/www/ cen2000.html (accessed March 30, 2012).

[44] United States, Department of Agriculture, Economic Research Service (ERS). http://www.ers.usda.gov (accessed March 30, 2012).

[45] United States, Department of Health and Human Services. 2000. Healthy People 2010. 2nd ed. With Understanding and Improving Health and Objectives for Improving Health. 2 vols. Washington, DC: U.S. Government Printing Office.

[46] United States, Office of Management and Budget 2008. Update of Statistical Area Definitions and Guidance on their Uses: 1–2.

[47] University of Wisconsin, Population Health Institute. http://uwphi.pophealth.wisc.edu (accessed April 2, 2012)

[48] Van der Weijden T., van Steenkiste B., Stoffers H.E.J.H., Timmermans D.R.M., Grol R. 2007. Primary prevention of cardiovascular diseases in general practice: mismatch between cardiovascular risk and patients' risk perceptions. Medical Decision Making 27, 754–761.

[49] Zuniga M, Anderson D, Alexander K. 2003. Heart Disease and Stroke in Rural America. Rural Healthy People 2010: A companion document to Healthy People 2010. Volume 1. College Station, TX: The Texas A&M University System Health Science Center, School of Rural Public Health, Southwest Rural Health Research Center.

[50] Wing S. 1988. Social inequalities in the decline of coronary mortality. American Journal of Public Health 78(11):1415-1416.

[51] Winsborough, H. H., and P. Dickinson. 1971. Components of Negro-White Income Differences. In Proceedings of the Social Statistics Section, 6–8. Washington, DC: American Statistical Association.

# Appendix A. Variable Descriptions

Table A.1: Description and Sources of Variables

| Variable | Definition | Source |
|---|---|---|
| MSA/Non-MSA County | Using the Metropolitan Statistical Area (MSA) definition, a county is defined as urban according to a relatively high population density in its core and close economic ties throughout the area. | U.S. Office of Management and Budget (2008) |
| Mortality rates of Cardiovascular Diseases (CVDs) | Mortality rates per 100,000 inhabitants due to CVDs, (International Classification of Diseases, 10th Rev.), namely coronary heart disease, acute myocardial infarction, cardiac arrhythmia, heart failure, hypertension and stroke (both ischemic and hemorrhagic) (2005-2007; 2007-2009) | National Vital Statistics System, U.S. Centers for Disease Control and Prevention |
| Population | Total population as of Decennial Census (1990 and 2000) | US Census |
| African American | Fraction of African American (1990 and 2000) | US Census |
| Asian American | Fraction of Asian American (1990 and 2000) | US Census |
| Population over age 65 | Fraction of county population over 65 as of Decennial Census (1990 and 2000) | US Census |
| Net migration | Net migration to county | US Census |
| Median income | Median county household income as of Decennial Census (1990 and 2000) | US Census |
| Below poverty line | Fraction of population below poverty line (1990 and 2000) | US Census |
| Gini 2000 | Gini coefficient of household income inequality (2000) | University of Wisconsin Population Health Institute |
| Unemployment 2005 | Annual average unemployment rates in 2005. | Bureau of Labor Statistics, U.S. Department of Labor |
| Tertiary education | % population tertiary graduates (1990 and 2000) | US Census |
| Secondary education | % population high school graduates (1990 and 2000) | US Census |
| Programming Engineers | Per capita number of students enrolled in engineering programs at local universities (1990) | Downes and Greenstein (2007) |
| Physicians | Per capita number of physicians (2005) | American Medical Association |

Table A.2: Description and Sources of Variables

| Variable | Definition | Source |
|---|---|---|
| Smoking | Percent of adults that reported currently smoking (2002-2008) | University of Wisconsin Population Health Institute |
| Drinking | Percent of adults that report binge drinking (2002) | University of Wisconsin Population Health Institute |
| Diabets | Percent of Diabetic Medicare enrollees receiving HbA1c test (2003-2006) | University of Wisconsin Population Health Institute |
| Obesity | Percent of adults that report BMI >= 30 (2004) | U.S. Department of Health and Human Services |
| Uninsured | Percent of adults 18-64 without insurance (2005) | U.S. Department of Health and Human Services |
| Primary care | Number of primary care providers (PCP) in patient care (2006) | University of Wisconsin Population Health Institute |
| Ambulatory | Discharges for ambulatory care sensitive conditions/Medicare Enrollees (2005-2006) | University of Wisconsin Population Health Institute |
| Mental | Average number of reported mentally unhealthy days per month | University of Wisconsin Population Health Institute |
| Expenditure in fast food | Per capita expenditure in fast food (2002) | Economic Research Service, U.S. Department of Agriculture |
| Expenditure in health | Per capita expenditure in health (1992 and 2002) | U.S. Census of Governments |
| Married men | The percentage of married men (1990 and 2000) | US Census |
| Married women | The percentage of married women (1990 and 2000) | US Census |
| No social support | Percent of adults that report not getting social/emotional support | University of Wisconsin Population Health Institute |
| Amenities | The natural amenities scale is a measure of the physical characteristics of a county area that enhance the location as a place to live. The scale is constructed by combining six measures of climate, topography, and water area that reflect environmental qualities most people prefer. These measures are warm winter, winter sun, temperate summer, low summer humidity, topographic variation, and water area. | Economic Research Service, U.S. Department of Agriculture |
| Ozone days | Number of days in 2005 that air quality was unhealthy due to ozone | University of Wisconsin Population Health Institute |
| PM days | Number of days in 2005 that air quality was unhealthy due to fine particulate matter | University of Wisconsin Population Health Institute |

# Appendix B. Detailed Decomposition for the Coefficients Effect

Table B1: Detailed decomposition for coeff. effect (Table 5)

| | All | Men | Women |
|---|---|---|---|
| Constant | 302.379* | 494.805*** | 249.795* |
| Population | -19.263*** | -13.439* | -11.817** |
| Pop. Growth | 0.723 | 1.350 | -0.765 |
| Net migration | -4.036 | -6.300 | -1.835 |
| African Am. | 2.689** | 3.159*** | 3.552*** |
| Ch. Afr. Am. | -14.237* | -13.113 | -11.367* |
| Asians | -5.895 | -1.807 | -3.518 |
| Ch. Asians | 6.668 | 28.463 | 3.923 |
| Aged 65+ | -0.169 | 0.591 | 0.153 |
| Ch. Aged 65+ | -76.570 | -61.334 | -74.072** |
| Median income | -40.007** | -43.546** | -23.948* |
| Below p.l. | -82.828 | -88.659 | -95.744* |
| Gini 2000 | 20.412* | 20.320 | 13.440 |
| Unemp. 2005 | 74.729*** | 81.388*** | 68.813*** |
| Income growth | 0.343 | -0.982 | 1.290 |
| Ch. Below p.l. | 70.868* | 94.372** | 72.014** |
| Tertiary edu. | -5.638 | -3.748 | -8.395 |
| Ch. Ter. Edu. | -108.016 | -255.139* | -194.705** |
| Secondary edu. | 49.655 | 97.009 | 90.811* |
| Ch. Sec. Edu. | -102.531** | -148.41*** | -70.472* |
| Professionals | -0.229 | 0.556 | -0.300 |
| Prog. Engineers | 7.535 | 9.165 | 7.050 |
| Drinkers | 11.033 | -1.475 | 15.498 |
| Smokers | 1.169 | -30.168 | 5.220 |
| Obese | 108.016** | 154.601*** | 96.397** |
| Diabets | 21.413 | 17.661 | 17.212 |
| Uninsured | -21.653** | -29.552*** | -11.689 |
| Primary care | -21.003* | -26.26574* | -30.923*** |
| Ambulatory | 27.954* | 40.514** | 26.046** |
| Mental | 15.419 | 27.589** | 8.087 |
| Physicians | 20.053 | 41.714* | 34.212** |
| Fast food exp. | -1.387 | 4.132 | -1.974 |
| Health exp. 92 | 1.739 | -6.193* | 3.428 |
| Health exp. 02 | -36.138 | -144.260 | 21.658 |
| Mar. men 1990 | -90.385 | -79.333 | -46.672 |
| Mar. women 2000 | 0.916 | 1.237* | 0.360 |
| Ch. Mar. Men | -2.811 | -1.480 | -1.910 |
| Ch. Mar. Wom. | -10.809 | -16.393 | -27.625** |
| No social sup. | 0.068 | -0.703 | 0.035 |
| Amenities | -3.590 | -10.851* | 0.677 |
| Ozone days | -2.377 | 2.842 | -3.458 |
| PM days | -145.632*** | -170.670*** | -144.078*** |
| Latitude | 15.083 | -1.888 | -2.817 |
| Longitude | -36.339*** | -34.24** | -28.411*** |

Significance level: * 10%; ** 5%; *** 1%.