Heart Mortality and Urbanization: The Role of Unobserved Predictors.

Thomas Bassetti* Nikos Benos† Stelios Karagiannis‡

January 23, 2013

Abstract

This article explores rural - urban differences in cardiovascular mortality. Using U.S. county-level data and a well-established decomposition technique, we study how observed and unobserved risk factors affect the mortality rate in metropolitan and nonmetropolitan areas. Results indicate that observed and unobserved variables have opposite effects on the risk gap. In particular, traditional risk factors are less abundant in metropolitan counties, in line with the existing literature. However, unobserved factors are less dangerous in sparsely populated rural areas. Our results continue to hold when we distinguish between different types of heart diseases and examine men and women separately. These findings have important consequences in terms of urbanization and migration policies.

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 60 million of individuals in the United States are afflicted with some form of cardiovascular disease (CVD). Specifically, heart disease and stroke are respectively the first and third leading causes of death in the United States. All together, these diseases contribute to 40% of deaths in the U.S.[1] Every year about 935,000 Americans have a heart attack. Of these, 610,000 are a first heart attack. 325,000 happen in people who have already had a heart attack (Roger et al., 2012). From an economic point of view, coronary heart disease alone costs the United States $108.9 billion each year (Heidenreich et al., 2011). This sum includes the cost of health care services, medications, and lost productivity. In this respect, 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 the labor productivity and economic growth. By looking at high income countries, Suhrcke and Urban (2010) find that CVD mortality rates play a negative role on subsequent five-year growth rates.[2]

For all these reasons, CVDs have received tremendous attention in fields such as medicine, economics and sociology. Researchers agree that socio-economic variables are important determinants of CVDs. In fact, although the percentage of people with related pathologies - e.g., high cholesterol, hypertension and obesity - is extremely high, important disparities among certain groups clearly emerge. For instance, in 1995 the death rate of African-American males from cardiovascular diseases was 42 percent higher than white males. Analogously, the rate for African-American females was 65 percent higher than white females. Other vulnerable groups to heart disease and stroke are: older African Americans, Hispanic Americans, and individuals belonging to low socio-economic classes (Wing, 1988 and Sundquist et al., 2001).

Another important issue concerns the existence of a risk differential between rural and urban areas. In fact, from 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).

[1] Source: Centers for Disease Control and Prevention (CDC, 2002).

[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).

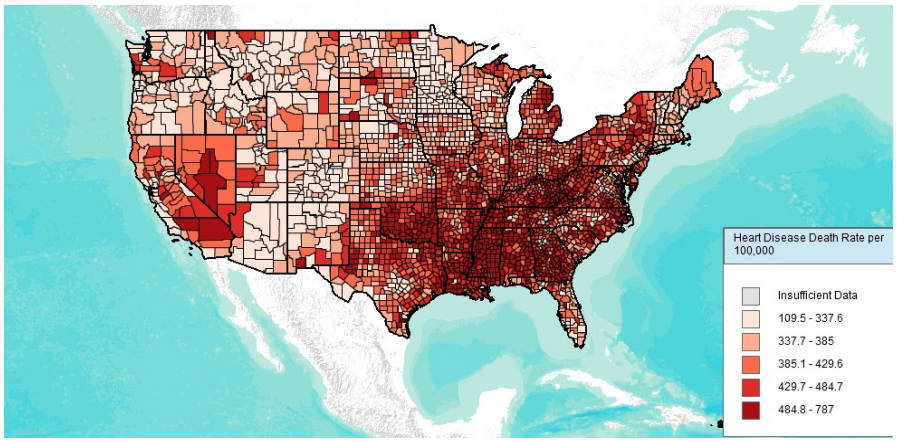

Figure 1: Map of heart disease in the U.S. (2007-2009)

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 Costal Range, the North-East of 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 disease. In other words, some modifiable risk factors for cardiovascular disease are usually more abundant in sparsely populated areas. However, two important questions remain: Is there any role for factors such as emotional well-being in explaining mortality rates? Is urbanization a Pareto improving policy in terms of CVD mortality?

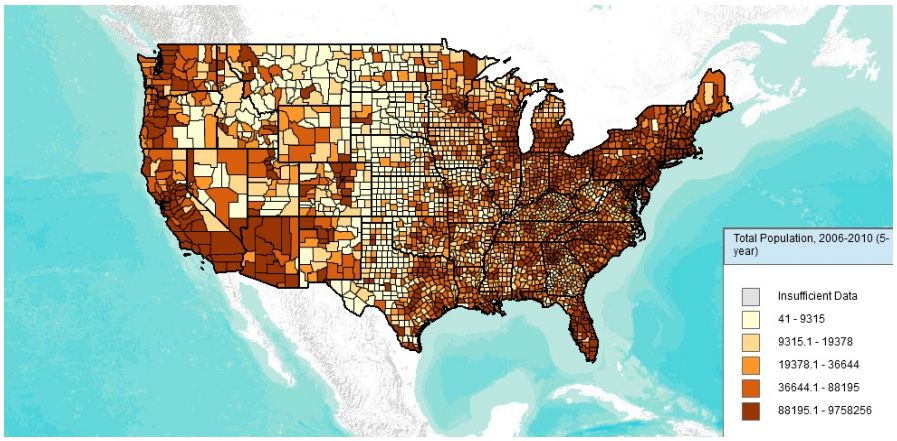

Figure 2: Total population in the U.S. (2006-2010)

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 that are not explained by the model, but that interact with the explanatory variables. Following Zuniga et al. (2003), we should refer to these unexplained components as perceptions and attitudes, that is, unobservable lifestyles.[4] After having correct for the existence of sample selection effects in rural areas, we arrive at several interesting results. First, in line with the existing literature, we show that traditional risk factors are more predominant in rural areas. Second, unobserved risk factors notably increase the mortality risk in populated rural counties, while they lower CVD mortality in less populated rural counties. Third, our results hold for almost all heart diseases. Finally, the main conclusions of this paper do not change when we consider gender-specific mortality rates.

Our study contributes towards the investigation of the cardiovascular mortality differentials between rural and urban areas in several important ways. First, to the best of our knowledge, this is the first work which uses decomposition techniques to explore the distinct roles of observed and unobserved factors in the explanation of these mortality differentials. Second, we compile a new dataset in order to examine most determinants of CVDs reported in the literature along with all types of cardiovascular diseases. We do so, using information at a low level of regional disaggregation (county data). In this way, we are able to shed light into the role of unobserved predictors regarding the rural-urban health disparities.

This study has clear policy implications. In fact, since sparsely populated rural areas are safer than densely populated ones, to reduce the CVD mortality gap between rural and urban areas as well as the overall CVD mortality rate, a government should facilitate rural migrations from more populated areas to less populated ones. Alternatively, policy-makers could facilitate the transformation of these areas into urban areas. On the basis of the health economics literature, from this policy, we can reasonably expect an increase in productivity and then

[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] Our model includes observable lifestyles such as binge drinking, smoking or poor eating habits.

a reduction of inequality.

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 about 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 and the economic ties throughout the area. Alternative definitions of areas 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 variables, we use the log of CVD mortality rates (per 100,000 inhabitants, aged 35 or over) due to 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). These variables are provided by National Vital Statistics System (2005) and they are logged in order to make their distribution comparable between rural and urban counties. Obviously, we cannot ignore the possibility of spatial correlation between CVDs in nearby areas due to possible dependence on spatially varying risk factors. Therefore, our dependent variable is spatially smoothed to produce better (more stable, less noisy) estimates of the risk in each area.

Table 1 summarizes the descriptive statistics for our dependent variables. These values are perfectly in line with the previous evidence of a higher mortality risk in rural areas. The last column of Table 1 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 death for cardiovascular diseases.

Table 1: Dependent variables (descriptive statistics)

Variable	Obs	Mean	Std. Dev.	Min	Max	% total
Heart disease 2005	3014	6.021	0.238	4.608	6.782	100
Heart disease 2005 (men)	3132	6.237	0.205	5.009	7.065	51.91
Heart disease 2005 (women)	3130	5.783	0.228	4.197	6.620	48.09
Coronary 2005	3132	5.568	0.263	4.341	6.531	27.42
Infarction 2005	3120	4.635	0.453	2.477	6.136	11.52
Arrhythmia 2005	2823	4.599	0.339	3.199	6.302	9.62
Heart failure 2005	3070	5.324	0.258	4.241	6.614	21.06
Hypertension 2005	3132	5.247	0.379	3.472	6.738	20.65
Stroke 2005	3126	4.548	0.200	3.555	5.548	9.73
Heart disease 2007	3013	5.960	0.245	4.929	6.804	100
Heart disease 2007 (men)	3130	6.180	0.215	4.655	6.937	51.97
Heart disease 2007 (women)	3129	5.713	0.228	4.748	6.469	48.03

Figure 3 shows the distribution of CVD mortality in both MSA and non-MSA counties. From this figure, we can see how some rural counties exhibit CVD mortality rates that are particularly high. We want to investigate whether these counties belong to a specific sub-group of nonmetropolitan areas.

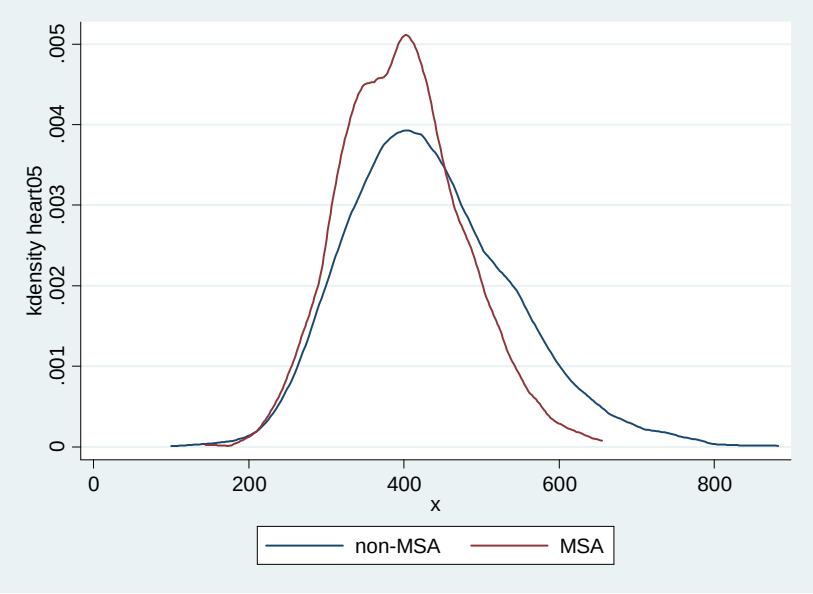

Figure 3: CVD mortality rates (MSA vs non-MSA)

The independent variables are drawn from a well established literature on cardiovascular diseases. These variables can be classified in six broad groups: demographic characteristics, economic variables, human capital factors, behavioral factors, social and environmental controls. Since reduction of CVD mortality involves lifestyle modifications, drug treatment and effective management of

any overt underlying medical condition, we must consider both prevention and treatment variables. Therefore, we use past variables (mainly from 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 or 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 fraction of African American and the fraction of Asian American), people aged more than 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 not only that poor and less educated people are more likely than wealthy and well-educated people to die from CVD, but that this gap may be widening. Therefore, we include the median household income, the percentage of population below poverty line as well as their changes over 1990-2000. We incorporate health expenditure in two periods (1992 and 2002) to estimate both prevention and treatment effects on CDVs (Govil et al., 2009).

As we have already mentioned, education is another important dimension to take 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 from 1990 to 2000. These changes allow us to account for the social consequences of human capital variation. Moreover, since the services sector is a high-risk sector for 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) 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. 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.

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 not getting 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]

[5] The natural amenities index is a measure of the physical characteristics of a county. This

Tables 2a and 2b 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 reduced the sample size to 1941 counties. These variables are so important to explain CVD mortality that they cannot be simply dropped. Therefore, our analysis must take into account the sample selection bias due to the fact that less populated counties are underrepresented.

Table 2a: Independent variables (descriptive statistics)

Variable	Obs	Mean	Std. Dev.	Min	Max
		Demography			
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
Net migration 1995	3144	0.0002	0.005	-0.139	0.073
Population	3142	10.147	1.374	5.869	15.997
Change pop.	3142	0.096	0.133	-0.551	1.068
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	3142	0.024	0.007	0.009	0.059
Change med. Inc.	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
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
		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

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.

Table 2b: 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
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
Social variables					
Married men 1990	3139	0.234	0.024	0.103	0.299
Married men 2000	3141	0.235	0.023	0.100	0.488
Married women 1990	3139	0.231	0.025	0.103	0.298
Married women 2000	3141	0.228	0.026	0.089	0.326
Change men	3139	0.001	0.018	-0.067	0.209
Change 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 well known Blinder–Oaxaca decomposition technique. Let Y denote the logged 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 a county index, c is a constant term, H stays 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. Small letters denote the relative vectors of coefficients, that is, the marginal impacts of each determinant on mortality rates. However, a simple OLS regression does not answer to the following question: Is the risk gap completely explained by differences in the endowment of risk factors? To address this issue, we employ the methodology popularized by Blinder (1973) and Oaxaca (1973). That is, we use a counterfactual approach to divide the risk gap between rural and urban areas into a part

that is "explained" by areas differences in risk factors and an "unexplained" part containing the effects of group differences in unobserved and unobservable predictors.

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

$$Y_{i,j} = X'_{i,j}\beta_j + u_{i,j}, \tag{2}$$

where β_j contains the slope parameters and the intercept. We can express the mean difference in the mortality risk as the difference in the linear prediction 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:

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

Given the sample selection problem mentioned in Section 2, we also use a correction method based on the procedure by Heckman (1976, 1979).[6] The appendix discusses how to handle the selectivity bias correction within the decomposition of the raw mortality gap. 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 differences to the overall mortality difference. In particular, it can be rearranged to obtain a "three-fold" decomposition:

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

In equation (4), the first addend 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"). Finally, the last addend is an interaction term accounting for the composite effect of differences in endowments and coefficients, since these differences operate simultaneously. We also separate men's mortality rates from women's mortality rates in order to analyze if previous 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 (PATT). Moreover, Kline (2011) shows that the classic Blinder-Oaxaca estimator is equivalent to a propensity score 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.

[6] See among others Reimers (1983) and Oglobin (1999).

We conclude the analysis with a battery of robustness checks. Some of these checks are only 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, underrepresented counties can enter the sample. Obviously, we are aware of the existing 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 others temporary phenomena affecting the analysis. Third, we distinguish metropolitan counties by the population size of their metro area, and nonmetropolitan counties by degree of urbanization and adjacency to a metro area or areas. We aim to test whether unobserved lifestyles in nonmetropolitan counties adjacent to metropolitan ones are similar to urban ones. Finally, we use as dependent variable 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 the specific diseases driving our results.

4 Results

4.1 Least Squares Estimates

We initially establish the correlation between regressors and regressand. Tables 3a and 3b contain the least squares estimates for metropolitan and nonmetropolitan statistical areas. These tables also report separate regressions for men and women.[7]

Table 3a shows that 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. However, in metropolitan areas, population affects heart mortality through net migration. On the contrary, the correlation between heart mortality and population growth (new births) is negative. As expected, the share of African Americans is associated with a 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 so to chronic diseases (Manton, 2008).

Concerning the role of economic variables, Table 3a shows that economic development is negatively associated with heart mortality only in nonmetropolitan areas. Vice versa, in metropolitan counties, the mortality risk is sensible to the change in the income level. This evidence is not surprising once we consider that the definition of MSA is based on the income level. However, even if MSA counties are rather homogeneous in terms of income levels, those counties that experience a higher economic growth will also exhibit a stronger reduction of

[7] In the appendix, we also report the OLS estimates after having corrected for the sample correction. Additional tables are avalible upon request from the corresponding author.

heart mortality. Table 3a also reveals the existence of a negative relationship between heart mortality and the fraction of population below the poverty in nonmetropolitan areas. By using data on single heart diseases, it is possible to see that this negative relationship is exclusively due to stroke and infarction (see Gillum and Mussolino, 2003).[8] Because of the high level of saturated or trans fats found in much of the fast food items, the expenditure in fast foods is positively correlated with the heart mortality for men 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.

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 line with that, the relationship between the change in secondary education and CVD mortality is positive and significant in metropolitan counties.

In urban areas, the share of professionals has a positive impact on CVD mortality in MSA areas, and this is true for both men and women (Table 3b). Similarly, areas with higher fractions of programming engineers seem to be 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 3b also suggests that the heart mortality risk increases with the number of (men) smokers. Indeed, smoking is a major cause of coronary artery disease. Similarly, obesity is positively correlated to heart diseases. Although it may seem surprising at first view, a negative correlation between heart mortality and alcohol consumption is not so strange if we consider that chronic conditions (e.g., diabetes and heart disease) imply drinking and eating restrictions. The same reasoning holds for the relationship between the number of diabetic patients and heart mortality.

The number of primary care providers is negatively correlated with mortality rates, whereas the opposite is true for the number of discharged patients from hospitals for ambulatory care sensitive conditions. Finally, in line with the existing literature, we found that metropolitan centers suffer from the lack of social capital and the high levels of air pollution. Overall, our variables explain 59-67% of the total variation in heart death rates, which is satisfactory given that we use cross-section data.

[8] According to our sample, the marginal impact of poverty on stroke is -0.665 with a probability value equals to 0.007. As expected, we have also found a negative effect of poverty on the infarction rate, although the significance level is at 10%.

Table 3a: OLS regression (heart mortality, 2005)

	Non-MSA All	MSA All	Non-MSA Men	MSA Men	Non-MSA Women	MSA Women
Constant	5.884***	5.732***	6.231***	6.120***	5.381***	5.357***
	(0.300)	(0.312)	(0.275)	(0.277)	(0.263)	(0.333)
Population	0.046***	0.004	0.046***	-0.001	0.048***	0.003
	(0.011)	(0.008)	(0.010)	(0.007)	(0.010)	(0.009)
Ch. Pop.	-0.416***	-0.139	-0.354***	-0.204***	-0.348***	-0.157
	(0.102)	(0.101)	(0.076)	(0.070)	(0.088)	(0.103)
Net migration	10.310	1.162***	9.663	0.971***	-10.951	1.257***
	(10.241)	(0.334)	(8.880)	(0.314)	(9.046)	(0.352)
African Americans	0.194**	0.336***	0.204***	0.316***	0.257***	0.351***
	(0.078)	(0.074)	(0.066)	(0.067)	(0.072)	(0.082)
Ch. Afr. Am.	0.800*	-0.384	0.963**	-0.232	1.762***	-0.402
	(0.444)	(0.262)	(0.395)	(0.243)	(0.418)	(0.286)
Asians	-2.124	0.468	-1.369	0.343	-2.297	0.806**
	(1.564)	(0.321)	(1.299)	(0.283)	(1.667)	(0.355)
Ch. Asians	3.775	1.536**	1.172	1.060*	3.687	2.391***
	(2.418)	(0.701)	(1.903)	(0.624)	(2.470)	(0.771)
Aged 65+	-0.448	-0.782***	0.044	-0.563***	-0.827***	-0.984***
	(0.306)	(0.281)	(0.260)	(0.205)	(0.278)	(0.304)
Ch. Aged 65+	-2.260***	-2.525***	-1.230**	-2.000***	-1.886***	-2.531***
	(0.660)	(0.871)	(0.504)	(0.607)	(0.606)	(0.926)
Median income	-6.133*	0.401	-7.362**	-2.189	-5.742*	1.291
	(3.440)	(2.148)	(3.176)	(1.886)	(3.273)	(2.279)
Below p.l.	-0.454**	0.344	-0.364	0.152	-0.096	0.560*
	(0.225)	(0.271)	(0.224)	(0.238)	(0.204)	(0.292)
Ch. Income	-0.036	-0.603***	-0.167*	-0.593***	-0.038	-0.669***
	(0.105)	(0.155)	(0.096)	(0.126)	(0.097)	(0.159)
Ch. Below p.l.	0.000	-0.007	-0.015	0.005	0.043	-0.008
	(0.046)	(0.052)	(0.037)	(0.045)	(0.041)	(0.050)
Fast food exp.	0.000	0.000	0.000***	0.000	0.000	0.000
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Health exp. '92	-0.279**	-0.263*	-0.185	-0.359**	-0.332**	-0.282*
	(0.139)	(0.158)	(0.129)	(0.147)	(0.136)	(0.166)
Health exp. '02	0.166***	0.149**	0.098*	0.242***	0.183***	0.102
	(0.061)	(0.067)	(0.058)	(0.060)	(0.057)	(0.068)
Tertiary edu.	-0.872**	-1.440***	-0.794**	-1.537***	-0.413	-1.224***
	(0.362)	(0.389)	(0.317)	(0.343)	(0.377)	(0.433)
Ch. Ter. Edu.	-1.759***	-1.546**	-1.570***	-1.431**	-0.797	-1.077
	(0.559)	(0.725)	(0.504)	(0.637)	(0.519)	(0.810)
Secondary edu.	0.096	0.436	-0.167	0.294	-0.121	0.653*
	(0.235)	(0.331)	(0.199)	(0.270)	(0.209)	(0.338)
Ch. Sec. Edu.	0.666*	1.378**	0.414	1.209**	-0.061	1.423**
	(0.377)	(0.625)	(0.304)	(0.486)	(0.324)	(0.629)
Professionals	-0.233	0.627**	-0.088	0.917***	-0.197	0.545*
	(0.206)	(0.278)	(0.165)	(0.238)	(0.180)	(0.295)

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

Table 3b: OLS regression (heart mortality, 2005)

	Non-MSA All	MSA All	Non-MSA Men	MSA Men	Non-MSA Women	MSA Women
Prog. Engineers	1.617**	1.824***	1.853***	1.013	2.194***	2.387***
	(0.729)	(0.665)	(0.633)	(0.617)	(0.691)	(0.745)
Smokers	0.005***	0.003	0.003***	0.003**	0.004***	0.001
	(0.001)	(0.002)	(0.001)	(0.001)	(0.001)	(0.002)
Obese	0.007***	0.006**	0.004**	0.006**	0.008***	0.006**
	(0.002)	(0.003)	(0.002)	(0.002)	(0.002)	(0.003)
Drinkers	-0.003*	-0.003**	-0.001	-0.002	-0.004***	-0.005***
	(0.001)	(0.002)	(0.001)	(0.001)	(0.001)	(0.002)
Diabets	-0.001	-0.005***	-0.001	-0.005***	0.000	-0.005***
	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)
Uninsured	0.002	-0.002	0.002	-0.001	0.003**	-0.001
	(0.001)	(0.002)	(0.001)	(0.001)	(0.001)	(0.002)
Physicians	12.567	-0.309	13.030*	-3.283	11.119	0.695
	(8.763)	(4.905)	(6.819)	(4.106)	(7.810)	(6.065)
Primary care	-0.000***	0.000	-0.000*	0.000	-0.001***	0.000
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Ambulatory	0.002***	0.002***	0.001***	0.002***	0.001***	0.003***
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Mar. men 1990	0.594	0.742	-0.666	0.291	0.745	0.311
	(1.092)	(1.018)	(0.933)	(0.888)	(0.956)	(1.022)
Mar. women 2000	-0.482	0.753	0.532	0.866	1.108	1.741
	(1.142)	(1.280)	(0.965)	(1.008)	(0.951)	(1.324)
Ch. Mar. Men	-0.182	1.341**	-0.214	1.147*	0.265	1.146*
	(0.440)	(0.651)	(0.356)	(0.613)	(0.383)	(0.689)
Ch. Mar. Wom.	1.467	-1.659	-0.004	-0.998	0.627	-1.891
	(1.466)	(1.471)	(1.205)	(1.158)	(1.270)	(1.564)
No social sup.	0.001	0.003**	0.000	0.002	-0.001	0.004**
	(0.001)	(0.002)	(0.001)	(0.001)	(0.001)	(0.002)
Amenities	0.004	0.000	-0.001	-0.001	0.003	0.000
	(0.003)	(0.003)	(0.003)	(0.003)	(0.003)	(0.003)
Ozone days	0.001	0.001***	0.000	0.001***	0.003**	0.002***
	(0.001)	(0.000)	(0.001)	(0.000)	(0.001)	(0.000)
PM days	-0.001	0.001	0.001	0.001	-0.002	0.001
	(0.002)	(0.001)	(0.001)	(0.001)	(0.002)	(0.001)
Latitude	-0.006***	0.002	-0.003	0.003	-0.008***	0.002
	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)
Longitude	0.000	0.000	0.001*	0.000	0.001	0.000
	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)
Observations	1187	754	1209	754	1209	754
R-squared	0.60	0.64	0.63	0.67	0.65	0.59

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

4.2 Decomposition Analysis: Main Results

In this section, we present our decomposition analysis. Table 4 provides the decomposition described by equation (4). In particular, columns 1 and 2 report the coefficients and the standard errors for a simple decomposition, without any adjustment for the sample selection bias. On the contrary, columns 3 and 4 present the Blinder-Oaxaca decomposition corrected by the selection bias. In line with the existing literature, the predicted mortality risk is lower in metropolitan counties than in nonmetropolitan ones. Moreover, if we look at columns 1 and 2, we would conclude that the entire difference is explained by the endowment effect: metropolitan areas are safer than nonmetropolitan ones. This result would confirm previous studies based on simple regression models. However, columns 3 and 4 of Table 4 provide additional information. In particular, if we correct for the underrepresentation of small rural counties, the Blinder-Oaxaca decomposition reveals that unobserved predictors play an important role in explaining the risk gap. While differences in the endowment of risk factors are still in favor of metropolitan areas, the unexplained component is, now, in favor of nonmetropolitan areas. This effect is mitigated by the interaction term, which implies that the combination of differences in characteristics and coefficients partially compensates the effect of coefficients. We can conclude that highly populated nonmetropolitan areas are responsible for the lack of evidence in favor of a negative effect of coefficients. The sample selection bias hides the fact that risk factors are less dangerous in rural areas than in urban ones, despite the relative abundance of risk factors in nonmetropolitan counties.[9]

Table 4: Decomposition (Heart mortality, 2005-2007)

	Predicted	S.E.	Predicted	S.E.
Differential				
Non MSA	6.012***	0.007	6.012***	0.007
MSA	5.955***	0.007	5.955***	0.007
Difference	0.057***	0.010	0.057***	0.010
Adjusted	NO	NO	-0.356***	0.133
Decomposition				
Endowments	0.099***	0.015	0.107***	0.016
Coefficients	-0.034[a]	0.023	-0.531***	0.143
Interaction	-0.008	0.026	0.068**	0.028
Significance: [a]13.2%; * 10%; ** 5%; *** 1%. 1941 Obs.				

Table 5 proposes the same decomposition of Table 4 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 3b, we can say that this assumption seems to be rather sensible, especially for rural counties. Table 5 shows that,

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

conditional on our specification, results are qualitatively the same for both men and women. Moreover, now, in the unadjusted decomposition, the interaction term is negative and significant. That is, some unexplained effects emerge even when we do not correct for the selection bias.

Table 5: Decomposition for men and women (Heart mortality, 2005-2007)

	Men				Women			
	Predicted	S.E.	Predicted	S.E.	Predicted	S.E.	Predicted	S.E.
Differential								
Non MSA	6.235***	0.006	6.235***	0.006	5.773***	0.007	5.773***	0.007
MSA	6.186***	0.007	6.186***	0.007	5.742***	0.007	5.742***	0.007
Difference	0.049***	0.009	0.049***	0.009	0.031***	0.010	0.031***	0.010
Adjusted	NO	NO	-0.419***	0.112	NO	NO	-0.420***	0.129
Decomposition								
Endowments	0.105***	0.014	0.104***	0.014	0.088***	0.016	0.093***	0.017
Coefficients	-0.012	0.019	-0.565***	0.120	-0.006	0.021	-0.547***	0.137
Interaction	-0.045**	0.022	0.042*	0.024	-0.051**	0.025	0.034	0.027
Significance: * 10%; ** 5%; *** 1%. 1963 Obs.								

4.3 Robustness Checks and Additional Results

This section provides the results of a variety of additional checks. We first drop those explanatory variables that considerably reduce the sample size. Table 6 shows a consistent reduction in the adjustment term, which is not significant anymore. At the same time, the effect of coefficients in the unadjusted specification becomes negative and significant, with a convergence between this effect and the effect of coefficients obtained after the adjustment. Although we are introducing a misspecification bias, the sign and the standard errors support our main results. The most prudent interpretation suggests that the effect of unobserved risk factors is negative and at least half of the effect of characteristics.

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

	Predicted	S.E.	Predicted	S.E.
Differential				
Non MSA	6.045***	0.005	6.045***	0.005
MSA	5.968***	0.007	5.968***	0.007
Difference	0.078***	0.009	0.078***	0.009
Adjusted	NO	NO	-0.059	0.101
Decomposition				
Endowments	0.130***	0.014823	0.156***	0.015
Coefficients	-0.055***	0.019	-0.248**	0.108
Interaction	0.001	0.022	0.033	0.024
Significance: * 10%; ** 5%; *** 1%. 2952 Obs.				

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 are decennial census data, we can only change the dependent variable. Table 7 reports our decomposition outcomes when we use as dependent variable the average heart mortality rate over the period 2007-2009. Table 7 confirms the fact that the unadjusted gap is in favor of metropolitan areas. Again, columns 1 and 2 of Table 7 exclude the existence of an unexplained effect. This effect becomes significant after having controlled for sample selection effects.[10]

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

	Predicted	S.E.	Predicted	S.E.
Differential				
Non MSA	5.954***	0.007	5.954***	0.007
MSA	5.881***	0.007	5.881***	0.007
Difference	0.073***	0.010	0.073***	0.010
Adjusted	NO	NO	-0.259*	0.139
Decomposition				
Endowments	0.122***	0.016	0.122***	0.016
Coefficients	-0.030	0.024	-0.422***	0.149
Interaction	-0.019	0.026	0.041	0.029
Significance: * 10%; ** 5%; *** 1%. 1944 Obs.				

In order to identify which areas are responsible for the higher mortality risk observed in rural counties, we use an extension of the Blinder-Oaxaca decomposition approach that allows us to consider multi-class groups.[11] By using the Rural-Urban Continuum Code (2000) provided by the Economic Research Service of the U.S. Department of Agriculture, we find that the most risky areas are those classified as nonmetropolitan areas with a population of at least 20,000 inhabitants and adjacent to metropolitan areas. This also explains why a sample in which small rural areas are underrepresented produces very biased results. According to Table 8, the endowments of observed risk factors in our reference group are more dangerous than those observed in other areas. The only exception is for nonmetropolitan areas with a population of at least 20,000 inhabitants and not adjacent to metropolitan areas. In terms of observable risk factors, these areas are less dangerous than nonmetropolitan areas with a population of at least 20,000 inhabitants but adjacent to metropolitan areas.[12] Second, the effect of coefficient is positive for relatively big areas and negative for relatively small areas. This means that, if we take the coefficients of metropolitan areas and we apply them to the reference group, we would obtain an increase in the mortality rates. The opposite is true for rural areas that

[10] For the sake of space and to avoid repetition, we do not report the separate results for men and women when we take the heart mortality over 2007-2009.

[11] We use this technique only as an additional test, because so far we are not able to construct the confidence intervals.

[12] Notice that, now, we do not proceed with any adjustment for the population size.

are relatively small. In other words, for populated rural areas, unobserved risk factors represent a serious problem.

Table 8: Decomposition for rural-urban continuum code 2000 (reference group = 4)

Comp. group	1	2	3	5	6	7	8	9
Ref. group.	6.102	6.102	6.102	6.102	6.102	6.102	6.102	6.102
Comp. group	5.939	5.965	6.051	6.033	6.042	5.985	5.98	6.058
Difference	0.163	0.137	0.051	0.069	0.059	0.117	0.122	0.044
Endowments	0.122	0.092	0.253	-0.026	0.024	0.172	0.049	0.046
Coefficients	0.68	0.231	0.021	0.152	-0.527	-0.221	-0.309	-0.715
Interaction	-0.64	-0.186	-0.224	-0.057	0.563	0.166	0.381	0.713

Rural-urban continuum codes: 1 = metro area with 1 million population or more; 2 = metro area of 250,000 to 1 million population; 3 = metro area of fewer than 250,000 population; 4 = nonmetro area with population of 20,000 or more, adjacent to a metro area; 5 = nonmetro area with population of 20,000 or more, not adjacent to a metro area; 6 = nonmetro area with population of 2,500-19,999, adjacent to a metro area; 7 = nonmetro area with population of 2,500-19,999, not adjacent to a metro area; 8 = nonmetro completely rural or less than 2,500 population, adjacent to metro area; 9 = nonmetro completely rural or less than 2,500 population, not adjacent to metro area.

Given this evidence, in Table 9 we reclassify U.S. counties including non-metropolitan areas with a population of at least 20,000 inhabitants in the urban group. With respect to Table 4, we can see how differences in endowments remain unchanged, the adjusted difference is smaller and differences in coefficients are still significant.

Table 9: Decomposition with new classification

	Predicted	S.E.	Predicted	S.E.
Differential				
Non MSA	6.006***	0.010	6.006***	0.010
MSA	5.980***	0.006	5.980***	0.006
Difference	0.026**	0.011	0.026**	0.011
Adjusted	NO	NO	-0.261*	0.151
Decomposition				
Endowments	0.049***	0.012	0.079***	0.013
Coefficients	-0.035	0.031	-0.398**	0.164
Interaction	0.012	0.031	0.058*	0.033
Significance: * 10%; ** 5%; *** 1%. 1941 Obs.				

Table 10 combines the test of Table 9 with the data used in Table 6. This check allows us to see if with the new areas division and a As in Table 6, the selection bias correction is not needed anymore. Table 10 shows that, now, the mortality gap is smaller and the effect of coefficient almost compensates the effect of characteristics. This means that, with the new classification of counties, CVD mortality tends to be the same in rural and urban areas, since urban areas are less abundant of observed risk factors but more abundant of

unobserved ones. In this respect, we confirm the conclusion of Jones and Goza (2008), when they underline the need of separate analyses for rural, suburban and urban residents.

Table 10: Decomposition with new class. and more obs.

	Predicted	S.E.	Predicted	S.E.
Differential				
Non MSA	6.041***	0.007	6.041***	0.007
MSA	6.008***	0.005	6.008***	0.005
Difference	0.033***	0.009	0.033***	0.009
Adjusted	NO	NO	-0.057	0.104
Decomposition				
Endowments	0.051***	0.010	0.079***	0.010
Coefficients	-0.042**	0.021	-0.179	0.113
Interaction	0.025	0.021	0.043*	0.023
Significance: * 10%; ** 5%; *** 1%. 2952 Obs.				

In order to investigate whether different types of heart disease behave in a different manner, in Table 11, we report the results for six different types of CVDs, 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, the effect of coefficients is negative and significant for all cardiovascular diseases. Unobserved predictors are extremely important especially for hypertension, heart failure and arrhythmia. On the contrary, infarction is the only heart disease in which unobserved predictors do not affect the mortality gap. In other words, unobserved rural lifestyles are useful to prevent all heart diseases, except infarction.

Table 11: Decomposition for different heart diseases (2005-2007)

Types	Coronary		Infarction		Arrhythmia	
	Predicted	S.E.	Predicted	S.E.	Predicted	S.E.
Differential						
Non MSA	5.567***	0.008	4.637***	0.013	4.601***	0.011
MSA	5.517***	0.009	4.422***	0.014	4.607***	0.012
Difference	0.050***	0.012	0.215***	0.019	-0.006	0.016
Adjusted	-0.388**	0.176	-0.132	0.308	-0.574*	0.298
Decomposition						
Endowments	0.153***	0.021	0.256***	0.035	0.066**	0.031
Coefficients	-0.565***	0.187	-0.422	0.330	-0.650**	0.320
Interaction	0.025	0.036	0.034	0.065	0.011	0.061
Types	Failure		Hypertension		Stroke	
	Predicted	S.E.	Predicted	S.E.	Predicted	S.E.
Differential						
Non MSA	5.345***	0.008	5.229***	0.012	4.545***	0.006
MSA	5.209***	0.009	5.272***	0.012	4.491***	0.007
Difference	0.135***	0.012	-0.043**	0.017	0.054***	0.009
Adjusted	-0.563***	0.176	-0.907***	0.277	-0.364***	0.135
Decomposition						
Endowments	0.120***	0.020	0.086***	0.031	0.064***	0.016
Coefficients	-0.767***	0.189	-1.003***	0.299	-0.466***	0.144
Interaction	0.083**	0.036	0.010	0.065	0.038	0.028
Significance: * 10%; ** 5%; *** 1%. 1963 Obs.						

5 Conclusion

By using U.S. county-level data and the famous Blinder-Oaxaca decomposition, this paper investigated the role of observed and unobserved risk factors in explaining rural and urban differences in cardiovascular mortality. We have divided the estimated difference between mortality rates in rural and in urban areas into three components: a component due to differences in the endowments of forty-one traditional risk factors, a component due to differences in the marginal impacts of these factors, and an interaction term. We arrived at several interesting results. First, in line with the recent medical literature, observed risk factors are relatively more abundant in rural areas. This result still holds when we replace the dependent variables with male and female mortality rates. Hypertension is the only CVD for which the mortality rate is lower in rural areas than in urban areas. Second, if we adjust the data to account for sample selection problems, rural areas present favorable level of unobserved risk factors. In particular, nonmetropolitan areas with more than 20,000 inhabitants are responsible for the higher risk attributed by previous studies to rural areas. Densely populated nonmetro areas have both large endowments of observed and unobserved risk factors, that is, they are too big to capture the advantages of

small rural areas in terms of unobserved predictors and too small to take advantage of urban amenities. In these areas, a policy intervention on modifiable characteristics is particularly opportune. Two opposite policies are possible: transforming these areas into urban centers or reducing their population size.

This analysis has some important 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 and other 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] Centers for Disease Control and Prevention (CDC). Preventing heart disease and stroke. Addressing the nation's leading killers, 2002.

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

[6] 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.

[7] 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.

[8] 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.

[9] 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.

[10] Heckman, J. J. 1976. The Common Structure of Statistical Models of Truncation, Sample Selection and Limited Dependent Variables and a Simple Estimator for Such Models. Annals of Econometrics and Social Measurement 5: 475–492.

[11] Heckman, J.J. 1979. Sample Selection Bias as a Specification Error. Econometrica 47: 153–161.

[12] 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.

[13] 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.

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

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

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

[17] 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.

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

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

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

[21] 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.

[22] 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.

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

[24] 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.

[25] Oglobin, C. G. 1999. The Gender Earnings Differential in the Russian Transition Economy, Industrial and Labor Relations Review, Vol. 52 (4): 602-627.

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

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

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

[29] 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.

[30] 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.

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

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

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

[34] 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.

[35] Swift R. 2011. The Relationship Between Health and GDP in Oecd Countries in the Very Long Run. Health Economics 20: 306–322.

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

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

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

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

[40] 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.

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

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

[43] 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.

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

[45] 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 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
Expenditure in fast food	Per capita expenditure in fast food (2002)	Economic Research Service, U.S. Department of Agriculture
Expenditure in restaurant	Per capita expenditure in restaurant (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
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
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. Sample selection correction

This appendix describes the sample selection correction we use. If our observations are randomly sampled, the correction term is zero. Our correction is based on the Heckman two-step regression technique. Therefore, once we perform a Blinder-Oaxaca decomposition, the explained and unexplained components are not affected by the selection bias.

It is essential to adjust for the lack of balance in counties' characteristics when making group comparisons.

Take equation (2),

$$Y_{i,j} = X'_{i,j}\beta_j + u_{i,j},$$

and assume the following selection mechanism:

$$\pi^*_{i,j} = \gamma p_{i,j} + \varepsilon_{i,j},$$

$$\pi_{i,j} = \begin{cases} 1 \text{ if } \pi^*_{i,j} \geq 0 \\ 0 \text{ otherwise} \end{cases}.$$

We also assume that $\varepsilon_{i,j}, u_{i,j} \sim$Bivariate normal $[0; 0; 1; \sigma_{u_j}; \rho_j]$. If $E(u_{i,j}|\varepsilon_{i,j}) = \beta_\lambda \varepsilon_{i,j} \neq 0$, our sample is subject to a selection bias. Here, β_λ measures the covariance between the two error terms: $u_{i,j}$ and $\varepsilon_{i,j}$. Therefore, we can write the following equation:

$$E(Y_{i,j}|p_{i,j}, \varepsilon_{i,j}) = X'_{i,j}\beta_j + \beta_\lambda u_{i,j}.$$

Since $\varepsilon_{i,j}$ is not observable, we must estimate it. First, we construct the following function:

$$\begin{aligned} h(p_{i,j}, \pi_{i,j}) &= E(\varepsilon_{i,j}|p_{i,j}, \pi_{i,j} = 1) \\ &= E(\varepsilon_{i,j}|\varepsilon_{i,j} \geq -\gamma p_{i,j}). \end{aligned}$$

Now, we can define the general inverse Mills ratio:

$$\lambda(\gamma p_{i,j}) = E(\varepsilon_{i,j}|\varepsilon_{i,j} \geq -\gamma p_{i,j}) = \frac{\phi(\gamma p_{i,j})}{\Phi(\gamma p_{i,j})}.$$

Then, we have

$$E(Y_{i,j}|p_{i,j}, \varepsilon_{i,j}) = X'_{i,j}\beta_j + \beta_\lambda \lambda(\gamma p_{i,j}).$$

Following the mainstream, we correct the Blinder-Oaxaca decomposition as follows:

$$\begin{aligned} E(Y_{i,1}) - E(Y_{i,0}) - (\lambda_1\theta_1 - \lambda_0\theta_0) &= [E(X_{i,1}) - E(X_{i,0})]'\beta_0 + \\ &+ E(X_{i,0})'(\beta_1 - \beta_0) + \\ &+ [E(X_{i,1}) - E(X_{i,0})]'(\beta_1 - \beta_0), \end{aligned}$$

where θ_j is an estimate of $\rho_j \sigma_{u_j}$. This correction affects both the explained and the unexplained component. On the contrary, with other correction approaches, the relative magnitude of the biased and unbiased gap may not relate directly to the effect of coefficients. In general, it is not possible to point out any correction approach as the right one, since the appropriate procedure depends on the specific empirical problem.

Appendix C. OLS after Bias Correction

Table C.1: OLS with bias correction (heart mortality 2005)

	non-MSA	MSA
Constant	5.800***	5.536***
Population	0.046***	0.004
Ch. Pop.	-0.411***	-0.165**
Net migration	8.365	0.986*
African Americans	0.208***	0.350***
Ch. Afr. Am.	0.836**	-0.397
Asians	-2.169*	0.495
Ch. Asians	3.747*	1.469*
Aged 65+	-0.446*	-0.769***
Ch. Aged 65+	-2.272***	-2.650***
Median income	-6.444*	0.338
Below p.l.	-0.468**	0.378
Ch. Income	-0.044	-0.579***
Ch. Below p.l.	-0.002	0.003
Physicians	12.801	-0.348
Fast food exp.	0.000	0.000
Health exp. '92	-0.280*	-0.258
Health exp. '02	0.160**	0.139*
Tertiary edu.	-0.900***	-1.565***
Ch. Ter. Edu.	-1.797***	-1.706**
Secondary edu.	0.121	0.492
Ch. Sec. Edu.	0.686*	1.385**
Professionals	-0.230	0.696***
Prog. Engineers	1.600*	1.801*
Smokers	0.004***	0.002
Obese	0.007***	0.006**
Drinkers	-0.002*	-0.003**
Diabets	0.000	-0.004***
Uninsured	0.002	-0.001
Primary care	0.000***	0.000
Ambulatory	0.002***	0.002***
Mar. men 1990	0.522	0.614
Mar. women 2000	-0.427	1.000
Ch. Mar. Men	-0.177	1.316
Ch. Mar. Wom.	1.394	-1.826
No social sup.	0.001	0.004**
Amenities	0.004	0.001
Ozone days	0.000	0.001**
PM days	-0.001	0.001
Latitude	-0.005**	0.003
Longitude	0.001	0.001
Observation	1941	1941

* significant at 10%; ** significant at 5%; *** significant at 1%