How Much Do We Know about Rural-Urban Health Disparities: Lessons from Four Major Diseases in Virginia

Di Zeng[†*], Wen You[†], Bradford Mills[†], Jeffrey Alwang[†], Michael Royster[‡], Kenneth Studer[‡], Rexford Anson-Dwamena[‡]

[†] Department of Agricultural and Applied Economics, Virginia Tech
[‡] Office of Minority Health and Health Equity, Virginia Department of Health

Selected Paper prepared for presentation at the Agricultural & Applied Economics Association's 2012 AAEA Annual Meeting, Seattle, Washington, August 12-14, 2012

[*] Corresponding author: di83@vt.edu

Abstract: Health disparities are abundantly recorded in literature, but is much less understood within a rural-urban context. In this paper, four major diseases in Virginia are studied: cancer, stroke, cardiovascular disease (heart disease) and chronic obstructive pulmonary disease (COPD). Separate count data regressions are estimated at regional level to provide a primary understanding of those factors. A simultaneous equations model with rural-urban specification are then estimated via seemingly unrelated regression (SUR) techniques to take account of possible causalities among these diseases as well as error correlations, which is followed by Blinder-Oaxaca decomposition of the disparity proportions explained by observed characteristics and unobserved mechanisms. The results suggest that regional-level factors are significantly correlated with health disparities between rural and urban areas. The unknown mechanisms behind these linkages are different between rural and urban areas, and explain an even larger proportion of these disparities.

Key Words: Disparities, Rural-Urban, Count Data, SUR, Virginia

1. Introduction

It has long been witnessed that different demographic and socioeconomic groups may differ in their health status, yet this issue has received increasing attention only recently from researchers, policy makers and the general public. For example, in Virginia in 2006, the infant mortality rate of African Americans (13.8 per thousand live births) was more than twices that of Whites (5.5 per thousand live births), and more than three times that of Hispanics (4.1 per thousand live births) and Asian and Pacific Islanders (4.2 per thousand live births) (Virginia Health Information database, 2006). Such disparities in aggregate health outcomes are not local phenomena. For example, it has been shown using transnational data that adults aged between 25 and 50 with a college degree will on average live 5 years longer than those with less than a high school education (Robert Wood Johnson Foundation, 2008).

A health disparity population, as defined by the National Institutes of Health (NIH), is a population where there is a significant disparity (difference) in the overall rate of disease incidence, prevalence, morbidity, mortality, or survival rates in the population as compared to the health status of the general (or a reference) population (NIH, 2000). Many factors may be correlated with health disparities and the mechanisms are complicated. In literature, key factors have been identified as the primary pathways that affect individual health outcomes. These factors include socioeconomic status (Williams et al., 1995; Lantz et al., 1998; Schulz et al., 2000), lack of health insurance (Monheit et al., 2000; Baker et al., 2001), adverse health behaviors (Lantz et al., 2001) and environmental risks (Currie et al., 2011).

Socioeconomic factors have been shown to be associated with health outcome disparities. For example, the Eight Americas Study investigated the differences in health outcomes for eight distinct groups of the U.S. Population classified jointly by race and income. The study shows that the life expectancy gap between male African Americans living in high-risk urban environments (who have the shortest life expectancy) and Asian females (with the longest life expectancy) can be as large as 21 years (Murray et al., 2005).

Many studies have focused on different componnents of socioeconomic factors. For example, it is shown that income is closely related to health outcomes and related disparities. According to the literature, poverty is consistently linked with health disparities such as increased mortality risk (Lochner et al., 2001), lower self-rated health (Kennedy et al., 1998; Blakely et al., 2000; Subramanian and Kawachi, 2003a; Subramanian and Kawachi, 2003b), higher prevalence of depressive symptoms (Kahn et al., 2000), more adverse health-related behaviors (Diez-Roux et al., 2000), and worse infant outcomes (Olson et al., 2010).

Another socioeconomic factor associated with health outcome disparities is education, which is closely related to economic status. Recent literature includes the Robert Wood Johnson Foundation (2008) study mentioned above, and Lleras-Muney (2005), which investigated the relationship between education and adult mortality in the United States. The negative relationship between education and mortality revealed in these two studies is consistent with several earlier investigations, including

Kitagawa and Hauser (1973), Christenson and Johnson (1995), Elo and Preston (1996) and Rogers, Hummer and Nam (2000).

Besides socioeconomic factors, much has also been done investigating other factors that affect health outcomes. One prominent correlate of health disparities is race. Multiple investigations have documented a consistent gap in all measures of health outcome, particularly between African Americans and White Americans (Hahn et al., 1995; Singh et al., 1996; Wong et al., 2002; Smedley et al., 2003; Burchard et al., 2003). Health behaviors are also recorded as determinants of health outcomes. Such behavioral indicators include health-related expenditure (Crémieux et al., 1999; Bokhari et al., 2007), smoking and drinking (Fertig, 2010; Chatterji and Markowitz, 2001; Gavaler et al., 2004).

Although expanding literature on the above aspects that generate health disparities is observed, the relationship between place of residence and health outcomes has been much less recorded among health economic studies. Specifically, little work has been done investigating economic factors that are correlated to rural-urban health disparities. Investigations have been performed on the correlation between limited access to health care services in rural areas and health disparities (Office of Rural Health, American Psychological Association, 1995; Fortney and Warren, 2000; Bull et al., 2001) as well as related policies (Jensen and Royeen, 2002; Strasser, 2003; Nelson and Gingerich, 2010), while little is known about predictors of rural-urban health disparities beyond health care access; if they differ between rural and urban areas and to what extent they explain such disparities.

This study aims to bridge this gap by providing an empirical examination of the predictors of rural-urban health disparities in Virginia using multiple datasets. Specifically, we would like to know: 1) if rural areas are worse off in terms of health outcomes because of socioeconomic disadvantages; 2) if rural and urban areas share the same predictors of health outcomes and how they differ if not; and 3) to what extent can these predictors explain possible health disparities. Also, we are interested in knowing possible policy implications related to the answers of the above. In this study, we focus on four major diseases: cancer, stroke, cardiovascular disease (heart disease hereafter) and chronic obstructive pulmonary disease (COPD hereafter). Regional-level aggregate counts of these risks are used in count data regressions. A simultaneous equations model that specifies four diseases within a binary rural-urban context are then estimated via seemingly unrelated regression (SUR) techniques to take account of possible causalities among these diseases as well as error correlations. Blinder-Oaxaca decomposition (Blinder, 1973; Oaxaca, 1973) is used to examine what proportions of the disparities can be explained by observed and unobserved characteristics, respectively.

2. Modeling Framework

Ideally, patient-level data can be matched with regional-level factors and hierarchical modeling can be employed in this case. Unfortunately, socioeconomic data is not available at the patient level given the specific type of our data, nor does our dataset provide us reasonable control groups that may support such hierarchical analysis since we only have patients in our dataset. Thus, aggregation has to be done and count data model is employed based on patient counts at regional level (ZIP-code level in our case).

In the first stage, four separate count data regression models are estimated. The regional-level counts (*Y*) is regressed on a set of factors, including demographic characteristics (*D*), income and income inequality (*I*), behavioral factors (*B*), education (*E*) and environmental risks (*R*), i.e.

$$Y = f(D, I, B, E, R)$$

D includes population under investigation, the proportion of African Americans and average household size, whom are consistently shown to have inferior health outcomes in literature; and regional-level obesity rate. *I* includes average household income and Gini coefficient measure of income inequality. *B* includes average household expenditure on health insurance, alcohol and tobacco products. *E* is measured by the percentage of college graduates among total population. Finally, given a lack of environmental quality statistics, *R* is measured by a binary indicator that shows if there is a Superfund site within that area. These regressions are supposed to provide a basic understanding of the possible correlation between these factors and regional-level health outcomes, and act as the basic specification of further analysis.

Simple specification like the above may not be able to detect and deal with possible error correlations, which is very likely to be in effect in our case given complicated mechanisms among diseases. Medical theories and empirical studies suggest certain causal relationships among these diseases that need to be considered. Specifically, it is suggested that stroke and COPD may increase patient's probability of developing cancer, and stroke occurs more easily among cancer patients (Grisold et al., 2009; Kornum et al., 2012). It is also found that the risks of stroke and heart disease are increased after COPD exacerbation (Donaldson et al., 2010). Besides, cancer may increase the risk of heart disease (Keating et al., 2006), and heart disease may further increase the risk of stroke (Broderick et al., 1992).

In the second stage, we build a simultaneous equations model based on the notions above. Specifically, we have the following system:

$$
\begin{aligned}
&Y_{CANCER} = f(Y_{STROKE}, Y_{COPD}, D, I, B, E, R)\\
&Y_{STROKE} = f(Y_{CANCER}, Y_{HD}, Y_{COPD}, D, I, B, E, R)\\
&Y_{HD} = f(Y_{CANCER}, Y_{COPD}, D, I, B, E, R)\\
&Y_{COPD} = f(D, I, B, E, R)
\end{aligned}
$$

Since we are interested in the mechanisms that explain rural-urban health outcomes, we have two equations for each disease type for rural and urban subpopulation, respectively. Thus, there are 8 equations which are estimated simultaneously via seemingly unrelated regression (SUR).

Blinder-Oaxaca decomposition is used to investigate the proportions of such disparities due to observed characteristics and unobserved mechanisms, respectively. For each disease-specific rural-urban model pairs in the SUR estimation, by assuming rural areas have inferior health outcomes compared with urban areas, a nonlinear Blinder-Oaxaca decomposition that applies for nonlinear models can be mathematically represented as (Bauer and Sinning, 2008):

$$\Delta = [E_{\beta U}(Y^U \mid X^U) - E_{\beta U}(Y^R \mid X^R)] + [E_{\beta U}(Y^R \mid X^R) - E_{\beta R}(Y^R \mid X^R)]$$

in which Δ is the difference in health outcomes between rural (R) and urban (U) areas. The counterfactual introduced is $E_{\beta U}(Y^R \mid X^R)$, which approximates the imaginary health outcome given still the rural covariates but evaluated using the coefficients of urban estimates The first term, $E_{\beta U}(Y^U \mid X^U) - E_{\beta U}(Y^R \mid X^R)$, then explains the health outcome gap due to observed characteristics (differences in values of covariates), and the second term, $E_{\beta U}(Y^R \mid X^R) - E_{\beta R}(Y^R \mid X^R)$, explains the health outcome gap due to unobserved mechanisms (differences in coefficients).

3. Data Description

The dependent variables in these regression models are regional-level counts of patients of each disease, which comes from the inpatient hospital discharge billing data (2006-2008) from the Virginia Health Information Database with special permission. From this dataset, we get patient-level disease information (in terms of ICD-9 Codes) and residential ZIP codes. Patients of these four diseases are identified using ICD-9 Codes. Since ZIP codes are a fairly small and the only geographical information, we aggregate the patients to this level for analysis.

Multiple sources provide data for the possible factors that are correlated with health outcomes. Most ZIP-code level predictors come from 2007 Demographic Estimates and Projections from Geolytics, Inc., a commercial data provider which projected and estimated the dataset to ZIP-code level based on US Census 2000. This dataset includes population statistics, average household size, income and education and behavioral characteristics such as average alcohol expenditure, smoking expenditure and health insurance expenditure at household level. To avoid severe collinearity, we calculated the share of college graduates among population above 25 years old as an indicator of education level. Also, we estimated the ZIP-code level Gini Coefficients based on the income cohorts as an indicator of income inequality.

Due to the lack of suitable regional-level data on environmental quality, the existence of a superfund site is used. The data comes from National Priority List from Environmental Protection Agency, which gives ZIP-Code level location information for each of the 31 currently active superfund sites across Virginia. For rural-urban specification, we adopt the classification systems proposed by Isserman (2005)[1]. It is applied at ZIP-code level, where both rural and mixed rural areas are considered as rural, and similarly, both urban and mixed urban areas are considered as urban. Finally, ZIP-code level obesity rates are projected from county-level obesity estimates from Center of Disease Control and Prevention. The projection is done in a GIS environment based on the non-nesting area relationship between ZIP-code area and counties. GIS files for both ZIP-code and county areas come from Topologically Integrated Geographic Encoding and Referencing (TIGER), US Census Bureau.

---

[1] The Isserman system is a four-pronged geographic classification:

i. Rural – population density of less than 500 per square mile and 90% of the population is in a rural area or the county has no urban area with population of 10,000 or more;
ii. Urban – population density of at least 500 per square mile, 90% of the population lives in urban areas, and the population in the urbanized areas is at least 50,000;
iii. Mixed Rural – meets neither the rural or urban definition and has a population density of less that 320 per square mile;
iv. Mixed Urban – meets neither the rural or urban definition and has a population density of at least 320 per square mile.

In this analysis, only patients aged between 35 and 64 are included. A primary look at the datasets suggests limited observations younger than 35, and we have too many above 64, the latter of which can fairly be considered as normal compared with disease occurrences among patients age 35-64 who are further from life expectation. Also, we take 3-year average (2006-2008) counts (rounded to nearest integers) for patients in an effort to make the observations smoother. For multiple hospital visits of the same patient for treatment of the same disease, we include only the last visit.

Table 1 shows state-level patient counts (rounded to the nearest integer). Cancer and heart disease occur much more than stroke and COPD, and cancer has the lowest mean patient age. Women are more likely to have cancer while men are more likely to suffer from heart disease. Compared with population share of African Americans in the 35-64 age cohort (18.83%, in 2007), the patient proportion of African Americans of the first three diseases are much higher (28.54%, 33.76%, 30.13%, respectively) and is slightly lower (18.53%) than its population share.

[Table 1 here]

Table 2 presents descriptive ZIP-Code level average incidence proportions of each disease. Under Isserman's classification, we have 616 rural areas and 190 urban areas after data merging and cleaning. Table 2 also compares the mean incidence proportions across rural and urban areas. Inferior health outcomes among rural areas are consistently witnessed.

[Table 2 here]

Descriptive statistics of all the possible factors that may affect aggregate health outcomes are given in Table 3. All factors are of significant different value across rural and urban ZIP-code level averages. Rural areas have lower income and expenditure in every means, lower proportion of African Americans, smaller household size, higher obesity rate, lower education level and higher income inequality. Most factors are observed to be worse than those of urban counterparts except for adverse health behavior indicators. Rural areas tend to have fewer patient counts, which, however, appear to be a larger proportion of population between 35 and 64, as seen in Table 2.

[Table 3 here]

4. Empirical Results and Interpretation

To get a primary understanding of the correlation between disease counts and predictors, we run four separate regression models for each disease in the first stage. Descriptive statistics suggest all disease counts exhibit over-dispersion since patient counts for the few largest ZIP-code areas extend well beyond 100. Vuong tests (Vuong, 1989) show that negative binomial model describes our data better than both hurdle negative binomial model and zero-inflated negative binomial model for all diseases, though most competing pairs appear to have close BIC and AIC numerically. This makes sense since zero is never the most common number of patient counts, nor can it be theoretically generated by any different mechanism. Thus, we only report the estimation results of four separate negative binomial regressions in Table 4.

[Table 4 here]

As discussed before, we have 806 ZIP-code areas in total, with rural and urban ones pooled together. As we see from the Likelihood Ratio Chi-Square tests, all the models are significantly explained jointly by these factors. Also, the dispersion parameters are significant across these models, confirming our preference of over-dispersion settings.

Among the demographic factors, total population has a positive and significant correlation with patient counts in all the models. As expected, all diseases tend to occur at a significantly higher rate where the proportion of African Americans is higher. This confirms the findings of many previous studies that African Americans have worse health outcomes. Average household size is negatively correlated with patient counts, which is also significant, indicating that areas with more larger households are less likely to see those disease incidences. As expected, obesity rate is positively correlated with patient counts, which is significant in all the models.

Income is negatively correlated with patient counts, which, however, is very small in magnitude and only marginally significant for cancer and COPD. Compared with income impacts, inequalities in income distribution play a more significant role. This is reflected by the significant coefficients of Gini coefficient across models. Areas with higher income inequality tend to observe more disease incidences, with population and other factors controlled for.

Intuitively, we expect behavioral factors such as smoking and alcohol expenditures to have positive signs and health insurance expenditure to have a negative sign. The results are consistent to our expectation. While alcohol is a significant factor that is correlated with more disease incidences, smoking does not show any significance across models and health insurance is only significant for COPD.

The population share of college graduates has negative correlations with patient counts, which are highly significant in all models. This suggests that areas with better educated population tend to witness fewer disease incidences when controlling for population and other factors. Finally, although the existence of superfund site within a ZIP-code area is positively correlated with patient counts, it is not significant in any case. This may be because we do not have a better environmental hazard indicator other than only 31 superfund sites among 806 ZIP-code areas.

These estimates offer us a basic understanding of the correlations between possible factors and disease incidences. We see that areas with higher proportion of African Americans, smaller households, higher obesity rates, higher income inequalities, more adverse health behaviors and less educated population tend to observe more disease incidences. Also, though not as significant, more disease incidences are likely to occur in poorer areas. However, these estimates do not take account for possible interactions among these diseases, and cannot provide any insights for understanding rural-urban disparities, i.e. the possibly different mechanisms behind differences in health outcomes, which are of our interest. Specifically, we would like to see how these factors are differently correlated with health outcomes (patient counts) in a rural-urban context with possible interactions among diseases properly considered. For this reason, we further implement simultaneous equations strategy and estimate the model described in Section 2. The results are shown in Table 5.

[Table 5 here]

From the last row in Table 5, it is seen that for the rural-urban paired regression equations for each disease, the chi-square test statistics are all significant at 1% level, suggesting that the null hypothesis that the same model applies to both rural and urban areas is rejected for all four diseases. As expected, the dispersion parameter is significant across all models. Also, the parameter estimates appear to be different from the separate estimation results in Table 4, as discussed in detail below.

In the SUR model, all the proposed possible correlations among disease incidences are positive and most estimates are highly significant. This suggests the existence of certain causal relationships among the occurrences of these disease at the regional-level, which is consistent with previous findings in medical and public health literature at individual level(Grisold et al., 2009; Kornum et al., 2012; Donaldson et al., 2010; Keating et al., 2006; Broderick et al., 1992). Although these specifications can only capture a portion of such complicated causal relationships among disease incidences, they provide useful insights in directing to the understanding of possible mechanisms behind regional-level health outcomes that are highly correlated. Most of these coefficients further appear to be different between paired rural-urban models, suggesting impacts of different magnitude across rural and urban areas. For example, the impact of stroke incidence on cancer incidence in rural areas is 1.08 times that of urban areas. After controlling for incidences of correlated diseases, population is also a significant factor, while health insurance and superfund existence are neither significant nor different between paired rural-urban models.

The proportion of African Americans are positively correlated with patient counts except for COPD patients, significant in the cases of stroke and heart disease but not cancer. The significantly negative impact on COPD outcomes also makes sense as we observed a slightly lower patient share of COPD compared with population share of African Americans. This impact is further tested to be statistically different between rural and urban areas. Household size, in general, only has significantly negative correlations with health outcomes in urban areas, suggesting that larger urban households tend to have lower probabilities of disease incidences. After controlling for more possible factors, obesity rates appear not as important, with only positive correlations with urban cancer incidences and rural COPD incidences, only the latter of which is significantly from that with urban COPD incidences.

One dramatic pattern is observed for the correlations between average household income and patient counts. They are significantly negative only for rural areas, though small in magnitude, suggesting fewer disease incidences in richer rural areas, compared with urban areas where no such relationships are found. Unlike income levels, income distribution measured by Gini coefficient is indifferently significant across most models. This provides support to the notion that economic inequality causes health risks anywhere throughout the state.

Adverse health behaviors, as measured in average household alcohol and tobacco products expenditures, exhibit different patterns. Alcohol consumption is positively correlated with all disease incidences in rural areas, while only moderate significance is found for urban stroke. This suggests that alcohol consumption is much more severe an issue in rural areas, and may point directly to possible policy designs for public health managers. Tobacco products consumption shows no significance in most cases, while is positively correlated with rural COPD outcomes. This again may provide some useful insights in public health policy making.

The negative correlations between college graduate proportions and health outcomes are confirmed with significance. Most appear to differ between rural and urban contexts, suggesting a larger potential impact of education improvement on better health outcomes. For example, a 10% increase in college graduate proportion among people aged 25 or above is correlated with a decrease of 1.5 stroke incidences in a typical urban ZIP-code area, while this correlation may be amplified to 3.7 stroke incidences in a rural ZIP-code area.

These unknown mechanisms that determine health disparities are confirmed different between rural and urban areas. To further understanding the relative contributions of differences in such unobserved mechanisms and differences in observed characteristics, we apply a Blinder-Oaxaca decomposition as discussed in Section 2. Table 6 presents the results for each disease.

[Table 6 here]

In general, differences in observed characteristics can only explain a small portion of rural-urban health disparities; it is the unknown mechanisms that play a bigger role. One extreme is from decomposing the COPD estimates, where almost all the disparities come from unknown mechanisms. Thse results further confirms the importance in exploring the mechanisms that generate rural-urban health disparities before policy making.

5. Concluding Remarks

Rural-urban health disparities exist widely and is confirmed by our dataset. These disparities are believed to bear some causal relationship in which they are predicted by a set of demographic, socioeconomic factors and health behaviors. The analysis of this paper consists of separate individual regression and SUR model estimation of patient counts of four major diseases on possible factors that may have impact on these health outcomes. Our analysis shows that demographic, socioeconomic factors as well as health behaviors can all affect regional-level health outcomes. The most general findings include positive correlations among all kinds of disease incidences, negative correlations between income, education and disease incidences, and positive correlations between population share of African Americans, obesity and alcohol consumption and disease incidences.

The analysis above suggests the existence of different mechanisms that determine rural and urban health outcomes, respectively. Further, such unknown mechanisms explain a even larger portion of rural-urban health disparities as seen through the Blinder-Oaxaca decomposition procedure. This suggests that the different mechanisms that generate heterogeneous health outcomes between rural and urban areas are of great importance both in understanding rural-urban health disparities and designing relevant policies. In fact, some of the found correlations may directly point to relative policies. For example, our analysis shows that rural areas with lower income, larger income inequality, heavier alcohol consumption and fewer college students may benefit the most if relative public health policies are in order. These factors are identified as the most important socioeconomic aspects, the improvement of which may yield significant changes in rural-urban health disparities, and should receive enough attention from health policy makers.

References:

Baker, DW, JJ. Sudano, JM. Albert, EA. Borawski, A. Dor (2001), Lack of health insurance and decline in overall health in late middle age, The New England Journal of Medicine, 345, pp. 1106–1112

Bauer, T, Sinning, M (2008), An extension of the Blinder-Oaxaca decomposition to nonlinear models, AStA Advances in Statistical Analysis 92(2), pp. 197-206

Blakely et al. (2000), What is the lag time between income inequality and health status?, Journal of Epidemiology and Community Health, 54, pp. 318–319

Blinder, AS (1973), Wage discrimination: reduced form and structural estimates. Journal of Human Resources 8, pp. 436–455

Bokhari, FAS, Y. Gai and P. Gottret (2007), Government health expenditures and health outcomes, Health Economics, 16(3), pp. 257-273

Braithwaite, RS, DO. Meltzer, JT. King, D. Leslie, MS. Roberts (2008), What does the value of modern medicine say about the $50,000 per quality-adjusted life-year decision rule? Medical Care, 46, pp. 349–356

Broderick, JP, et al. (1992), Relationship of cardiac disease to stroke occurrence, recurrence, and mortality, Stroke, 23(9), pp.1250-1256

Bull, CN, JA. Krout, E. Rathbone-McCuan, MJ. Shreffler (2001), Access and issues of equity in remote/rural areas, The Journal of Rural Health, 17, pp. 356–359

Burchard, EG, et al. (2003), The importance of race and ethnic background in biomedical research and clinical practice, New England Journal of Medicine, 348, pp. 1170–1175

Chatterji, P and S. Markowitz (2001), The impact of maternal alcohol and illicit drug use on children's behavior problems: evidence from the children of the national longitudinal survey of youth, Journal of Health Economics, 20(5), pp. 703-731

Christensen, BA. and NE. Johnson (1995), Educational inequality in adult mortality: an assessment with death certificate data from Michigan, Demography, 32 (2), pp. 215-29

Crémieux, PY, P. Ouellette and C Pilon (1999), Health care spending as determinants Health Economics, 8(7), pp. 627-639

Currie, J, M. Greenstone and E. Moretti (2011), Superfund cleanups and infant health, American Economic Review, 101(3), pp. 435-41

Cutler, DM, E. Richardson, TE. Keeler, and D. Staiger (1997), Measuring the health of the U.S. population, Brookings Papers on Economic Activity, Microeconomics, 1997, pp. 217-82

Diez-Roux et al. (2000), A multilevel analysis of income inequality and cardiovascular disease risk factors, Social Science and Medicine, 50, pp. 673–687

Donaldson, GC, et al. (2009), Increased Risk of Myocardial Infarction and Stroke Following Exacerbation of COPD, Chest, 137(5), pp. 1091-1097

Dow, W, RF. Schoeni (2008), Economic value of improving the health of disadvantaged Americans, Technical Report for Overcoming Obstacles to Health: Report from the Robert Wood Johnson Foundation to the Commission to Build a Healthier America, 2008 Report

Dunlop, S, PC. Coyte, W. McIsaac (2000), Socio-economic status and the utilization of physicians' services: results from the Canadian National Population Health Survey, Social Science & Medicine, 51, pp. 123-33

Elo, IT, Preston, SH (1996), Educational differentials in mortality: United States, 1979–85, Social Science and Medicine, 42 (1), pp. 47-57

Fertig, AR (2010), Selection and the effect of prenatal smoking, Health Economics, 19(2), pp. 209-226.

Fortney, J, K. Rost., J. Warren (2000), Comparing alternative methods of measuring geographic access to health services, Heath Services & Outcome Research methodology 1(2), pp. 173–184

Gannon, B, B. Nolan (2003), Disability and labour market participation, HRB Working Paper, June 2003

Gavaler, JS, et al. (2004), Directions for unraveling the issue of alcohol and health disparities: findings from the Postmenopausal Health Disparities Study, Alcohol, 32(1), pp. 69-75

Grisold, W, S. Oberndorfer, W. Struhal (2009), Stroke and cancer: a review, Acta Neurologica Scandinavica, 119(1), pp. 1-16

Hahn, RA, S. Eberhardt (1995), Life expectancy in four U.S. racial/ethnic populations: 1990, Epidemiology, 6, pp. 350–355

Isserman, A (2005), In the national interest: defining rural correctly for research and policy, International Regional Science Review 28 (4), pp. 465-499

Jensen GM., and CB. Royeen (2002), Improved rural access to care: dimensions of best practice, Journal of Interprofessional Care, 16, pp. 117–128

Kahn et al. (2000), State income inequality, household income, and maternal mental and physical health: cross sectional national survey, British Medical Journal, 321, pp. 1311–1315

Keating, NL, AJ. O'Malley and MR. Smith (2006), Diabetes and Cardiovascular Disease During Androgen Deprivation Therapy for Prostate Cancer, Journal of Clinical Oncology, 24(27), pp. 4448-4456

Kennedy et al. (1998), Income distribution, socioeconomic status and self-rated health:

a US multilevel analysis, British Medical Journal, 317, pp. 917–921

Kitagawa, EM, PM. Hauser (1973), Differential mortality in the United States: a study in socioeconomic epidemiology, Harvard University Press, Cambridge, MA

Kornum, JB, et al. (2012), Chronic obstructive pulmonary disease and cancer risk: A Danish nationwide cohort study, Respiratory Medicine, 106(6), pp.845-852

Lantz, PM et al. (1998), Socioeconomic factors, health behaviors, and mortality: results from a nationally representative prospective study of US adults, Journal of the American Mathematical Association, 279, pp. 1703–1708

Lantz, PM et al. (2001), Socioeconomic disparities in health change in a longitudinal study of US adults: The role of health-risk behaviors, Social Science & Medicine, 53, pp. 29–40

LaVeist, TA, DJ. Gaskin, P. Richard (2009), The economic burden of health inequalities in the United States, The Joint Center for Political and Economic Studies, 2009 Report

Lleras-Muney, A (2005), The relationship Between Education and Adult Mortality in the US, Review of Economic Studies, 72(1), pp189-221

Lochner et al. (2001), State level income inequality and individual mortality risk: a prospective multilevel study, American Journal of Public Health, 91, pp. 385–391

Mackenbach, JP, WJ. Meerding, AE. Kunst (2007), Economic implications of social-economic inequalities in health in the European Union, European Commission, 2007 Report

Meerding et al. (2005), Health problems lead to considerable productivity loss at work among workers with high physical load jobs, Journal of Clinical Epidemiology, 58, pp. 517-23

Monheit, AC, J.P. Vistnes (2000), Race/ethnicity and health insurance status: 1987 and 1996, Medical Care Research & Review, 57 (Suppl. 1), pp. 11–35

Murray, CJL, S. Kulkarni, M. Ezzati (2005), Eight Americas: new perspectives on U.S. health disparities, American Journal of Preventive Medicine, 29 (Suppl. 1), pp. 4-10

Nelson JA, BS. Gingerich (2010), Rural health: access to care and services, home health care management & practice, 22(5), pp. 339–343

Oaxaca, R (1973), Male-female wage differentials in urban labor markets, International Economic Review 14, pp. 693-709

Office of Minority Health & Public Health Policy (2008), Unequal Health across the Commonwealth: A Snapshot, Virginia Health Equity Report 2008, Virginia Department of Health

Office of Rural Health, American Psychological Association (1995), Caring for the rural community: an interdisciplinary curriculum, American Psychological Association, Washington, DC

Olson ME. et al. (2010), Impact of income and income inequality on infant health outcomes in the United States, Pediatrics, 126 (6), pp. 1165-1173

Robert Wood Johnson Foundation (2008), Overcoming obstacles to health: report from the Robert Wood Johnson Foundation to the Commission to build a healthier America, Princeton, RWJ Foundation, NJ

Rogers, RG., RA. Hummer, and CB. Nam (2000), Living and Dying in the USA, Academic Press

Schulz, A et al. (2000), Social inequalities, stressors and self reported health status among African American and white women in the Detroit metropolitan area, Social Science & Medicine, 51, pp. 1639–1653

Singh, GK, SM. Yu (1996), Trends and differentials in adolescent and young adult mortality in the United States, 1950 through 1993, American Journal of Public Health, 86, pp. 560–564

Smedley, BD, AY. Stith, AR. Nelson (2003), Unequal treatment: confronting racial and ethnic disparities in health care, National Academy Press, Washington DC

Strasser R. (2003). Rural health around the world: challenges and solutions, Family Practice, 20, pp. 457–463

Subramanian et al. (2003a), Income inequality as a public health concern: where do we stand? Commentary on Mellor, J., Milyo, J. “Is exposure to income inequality a public health concern?”, Health Services Research, 38 (1), pp. 153–167

Subramanian et al. (2003b), Income inequality and health: multilevel analysis of Chilean communities, Journal of Epidemiology and Community Health,57 (11), pp. 844–848

Sudano, JJ, DW. Baker (2006), Explaining US racial/ethnic disparities in health declines and mortality in late middle age: The roles of socioeconomic status, health behaviors, and health insurance, Social Science & Medicine, 62 (4), pp. 909-922

Vuong, QH (1989), Likelihood ratio tests for model selection and non-nested hypotheses, Econometrica, 57: pp. 307–333

Williams, DR, C. Collins (1995), US socioeconomic and racial differences in health: Patterns and explanations, Annual Review of Sociology, 21, pp. 349–386

Wong, MD et al. (2002), Contribution of major diseases to disparities in mortality, New England Journal of Medicine, 347, pp. 1585–1592

Table 1 Statewide Patient Counts of Four Diseases Among Population Aged 35-64

	Annual Average	Mean Age	Female	African Americans	Female African Americans
Cancer	19,097	51.59	12,477	5,450	3,893
Stroke	6,954	54.18	3,338	2,348	1,235
Heart Disease	24,357	53.75	9,499	7,338	3,442
COPD	4,343	55.45	2,402	805	444

Source: author's calculation.

Table 2 Rural-Urban Disparities in ZIP-code Level Incidence Proportions of Four Diseases

	Pooled Incidence Rate	Rural ZIP-code Area (n=616)	Urban ZIP-code Area (n=190)	Difference
Cancer	6.78‰	7.07‰	5.87‰	1.20‰***
Stroke	2.83‰	3.07‰	2.12‰	.95‰***
Heart Disease	9.20‰	9.96‰	6.81‰	3.15‰***
COPD	2.67‰	3.23‰	1.10‰	2.23‰***

Note: *** indicates the differences are significant at 1% level in a *t*-test.
Source: author's calculation.

Table 3 Descriptive Statistics of Possible Factors at ZIP-code Level Mean

	Pooled (n=806)	Rural (n=616)	Urban (n=190)	Differences
Cancer patient counts	22.24 (30.42)	11.02 (16.23)	58.63 (36.63)	-47.61*** (.000)
Stroke patient counts	8.04 (11.49)	4.17 (6.05)	20.59 (15.37)	-16.42*** (.000)
Heart disease patient counts	27.20 (37.16)	15.05 (21.73)	66.58 (48.01)	-51.54*** (.000)
COPD patient counts	5.05 (7.66)	3.80 (6.35)	9.08 (9.85)	-5.28*** (.000)
Population aged 35-64	3905.16 (5271.36)	1817.41 (2529.42)	10673.86 (6104.95)	-8856.45***
Proportion of African Americans aged 35-64 (%)	17.204 (.173)	15.80 (16.01)	21.75 (20.43)	-5.95***
Average household size	2.500 (.235)	2.487 (.177)	2.543 (.361)	-0.056***
Obesity rate (%)	27.58 (2.651)	28.09 (2.197)	25.94 (3.27)	2.15***
Average household income	53917.13 (21367.62)	49284.58 (15293.69)	68936.35 (29780.25)	-19651.77***
Gini coefficient	.395 (.046)	.401 (.039)	.376 (.059)	.025***
Alcohol expenditure	453.67 (116.48)	426.66 (90.72)	541.26 (144.64)	-114.60***
Tobacco products expenditure	276.90 (14.61)	274.21 (12.69)	285.65 (16.86)	-11.44***
Health insurance expediture	1461.30 (154.73)	1429.37 (128.23)	1564.83 (185.41)	-135.46***
Proportion of college graduates (%)	20.50 (15.05)	15.58 (9.01)	36.46 (19.12)	-20.89***
Existence of Superfund site	.030 (.170)	.024 (.154)	.047 (.213)	-.023

Standard errors are reported in parentheses. *** indicates 1% significance level.

Table 4 Separate Estimation of Negative Binomial Regressions

	Cancer	Stroke	Heart Disease	COPD
Pop	.0002***	.002***	.002***	.0002***
	(.000)	(.000)	(.000)	(.000)
African%	.008***	.012***	.00613***	-.014***
	(..002)	(..002)	(.002)	(.002)
HH Size	-.423**	-.699***	-.409**	-.643**
	(.183)	(.189)	(.184)	(.252)
Obesity%	.031**	.032**	.038**	.045**
	(.015)	(.016)	(.015)	(.020)
HH Income	-6.94e-06*	-4.22e-06	-4.50e-06	-1.51e-05*
	(.000)	(.000)	(.000)	(.000)
Gini	2.863**	2.435**	3.193***	4.189***
	(1.173)	(1.204)	(1.175)	(1.603)
Alcohol	.005**	.005**	.006**	.009***
	(.002)	(.002)	(.002)	(.003)
Tobacco	.022	.017	.025	.042
	(.023)	(.024)	(.023)	(.029)
Insurance	-.002	-.003	-.004	-.010**
	(.004)	(.004)	(.004)	(.005)
College%	-.011***	-.019***	-.018***	-.029***
	(.004)	(.005)	(.005)	(.006)
Superfund	.089	.108	.075	-.012
	(.159)	(.161)	(.163)	(.210)
Cons	--3.838	-1.385	-1.975	-.378
	(2.599)	(2.655)	(2.646)	(3.198)
Log α	-.714***	-.892***	-.650***	-.441***
	(.063)	(.081)	(.058)	(.080)
No. of Obs	806	806	806	806
Log Likelihood	-2794.246	-2069.154	-3005.994	-1857.445
LR χ^2	935.56***	836.34***	867.21***	558.56***
	(.000)	(.000)	(.000)	(.000)
Pseudo R^2	.143	.168	.126	.131

Standard errors are reported in parentheses. *, **, *** indicate 10%, 5% and 1% significance level, respectively.

Table 5 Estimation Results of Seemingly Unrelated Regression (SUR)

	Cancer		Stroke		Heart Disease		COPD	
	Rural (n=616)	Urban (n=190)	Rural (n=616)	Urban (n=190)	Rural (n=616)	Urban (n=190)	Rural (n=616)	Urban (n=190)
Cancer			.0114	.036***	.036***	.012***		
			(.009)†	(.009)†	(.008)†††	(.002)†††		
Stroke	.088***	.027***						
	(.015)†††	(.004)†††						
Heart			.022***	.005***				
Disease			(.006)†††	(.002)†††				
COPD	.024***	.0002	.003***	.0002	.052***	.011**		
	(.008)†††	(.004)†††	(.008)	(.005)	(.008)†††	(.005)†††		
Pop	.000***	.000***	.00002	.000***	.000	.000***	.000***	.000***
	(.000)†††	(.000)†††	(.000)†††	(.000)†††	(.000)†††	(.000)†††	(.000)†††	(.000)†††
African	.001	-.002	.006**	.003	.005**	.004	-.021***	-.006*
%	(.002)	(.002)	(.003)	(.002)	(.002)	(.003)	(.003)†††	(.003)†††
HH Size	.026	-.265	-.145	-.462***	.224	-.379**	.266	-.813***
	(.229)	(.196)	(.233)	(.138)	(.203)††	(.194)††	(.399)††	(.217)††
Obesity	.010	.036***	.021	.003	.011	.008	.088***	-.011
%	(.016)	(.012)	(.018)	(.010)	(.016)	(.012)	(.026)†††	(.020)†††
Income	-.000***	-.000	-.000*	.000	-.000**	.000	-.000***	-.000
	(.000)†††	(.000)†††	(.000)†††	(.000)†††	(.000)†††	(.000)†††	(.000)†††	(.000)†††
Gini	3.569***	2.436	2.774**	2.265**	3.113***	3.783**	4.023**	6.045***
	(1.298)	(1.549)	(1.378)	(1.149)	(1.178)	(1.582)	(1.959)	(1.194)
Alcohol	.004***	.001	.006***	.002**	.003*	-.002	.008**	-.002
	(.001)†	(.003)†	(.002)†	(.001)†	(.002)	(.005)	(.003)††	(.004)††
Tobacco	.022	-.023	.017	-.055	.010*	-.053	.023***	-.052
	(.033)†††	(.073)†††	(.020)	(.048)	(.006)	(.073)	(.007)††	(.038)††
Insuranc	.001	.004	-.002	.008	-.001	.008	-.008	.009
e	(.003)	(.010)	(.003)	(.007)	(.003)	(.010)	(.005)	(.010)
College	.003	-.011*	-.002**	-.019***	-.010***	-.019***	-2023**	-.049***
%	(.006)	(.006)	(.006)††	(.005)††	(.005)†††	(.006)†††	(.009)††	(.007)††
Super-	-.091	-.108	-.340	.024	-.271	.044	-.389	-.082
fund	(.264)	(.112)	(.352)	(.098)	(.213)	(.137)	(.320)	(.130)
Cons	-9.431**	2.092	-4.612**	7.246	-4.130*	6.385	-2.416	4.273
	(2.487)	(8.108)	(2.317)† †	(5.217)† †	(2.326)	(8.186)	(3.967)	(8.050)
Log α	-1.16***	-2.34***	-1.38***	-3.40***	-1.13***	-2.39***	-.44***	-1.75***
	(.108)	(.237)	(.139)	(.294)	(.095)	(.271)	(.106)	(.202)
χ^2 Test	263.563***		251.190***		261.180***		207.589***	
	(.000)		(.000)		(.000)		(.000)	

Standard errors are reported in parentheses. *, **, *** indicate 10%, 5%, 1% significance level of coefficients, respectively. †, ††, ††† indicate 10%, 5%, 1%

significance level of pairwise t-tests of coefficient equality. Population and income impacts are very small and are rounded to the 3rd decimal place with signs.

Table 6 Estimation Results of Seemingly Unrelated Regression (SUR)

	Explained	Unexplained
Cancer	34.7%	65.3%
Stroke	33.8%	67.2%
Heart Disease	45.1%	54.9%
COPD	1.2%	98.8%