INTRODUCTION

Disparities in health exert a tremendous burden on the health care community and on society in general. Although significant improvements in the overall health of the population in the United States have been realized over the last four decades, these improvements have not been realized by all segments of the population. In general, improvements in overall population health can be attributed to the combined effect of improved living standards, advancements in medical care and treatment, improved access to medical care resources, and increased awareness about health risks in the general population. A great deal of research has documented health-related disparities among population subgroups defined by gender, age, race and ethnicity, socioeconomic status, and geographic location. Accordingly, a broad range of initiatives have been undertaken to address population health disparities. In addition, the U.S. Department of Health and Human Services has boldly expanded their Healthy People 2000 objectives from reducing health disparities to the elimination of health disparities by 2010 (USDHHS, 2000).

Health disparities, represented by differences in the incidence, prevalence, and mortality from disease and other adverse conditions may result from variations in social, cultural, behavioral, biologic, genetic, and environmental factors among population subgroups and geographic locations. Due to the considerable health disparities experienced by racial/ethnic minorities, particularly African Americans, there has been a great deal of effort at the national level to address, reduce, and eliminate these disparities. Recent evidence has identified significant geographic disparities within racial/ethnic groups for major causes of death and illness (Pickle, et al 1996; Casper et al. 2000; Halverson et al 2004). Although geographic (place-based) disparities have been widely documented less effort has been devoted to addressing place-based health disparities at the federal level, despite the fact that they occur among racial/ethnic minorities as well.

There is a considerable body of research that suggests adverse socioeconomic status (SES) is a key factor in observed health disparities among population subgroups in the U.S. Poverty, income levels, employment status/unemployment, and educational attainment are used, either individually or combined, as measures of socioeconomic status. Strong associations between SES and health outcome measures have been consistently reported for both individual and area-level data (Anderson et al, 1997;Kahn et al, 1999;Winkleby et al, 1999; Pickering, 1999; Chen et al, 2006). The majority of studies analyze these relationships in aggregate as opposed to identifying differences in relationships across space. However, strong associations between SES and health outcomes do not exist for all diseases and are not uniform for all populations defined by ethnicity, gender, age, and geography. Nor is there a clear understanding of how SES influences the health outcomes of more disadvantaged individuals. A major regional study *An Analysis of Disparities in Health Status and Access to Medical Care in the Appalachian Region* found that there are geographically variable associations between aggregate socioeconomic conditions and observed health outcomes in the Appalachian region.

In order to develop a better understanding of observed health disparities in the Appalachian region, this study examines the relationships between health disparities in premature mortality for major cause of death and measures of socioeconomic condition. Premature deaths, those occurring before the age of 64, are considered preventable. Therefore observed disparities in premature mortality represent a blight on the public health landscape that needs critical attention and intervention.

Unlike individual measures of SES, measures of socioeconomic condition within local areas generally reflect the ability of local social, economic, and political structures to support a high-quality of life. Understanding socioeconomic disparities in health is important to identify underlying factors that give rise to conditions that both impede and facilitate behavioral, medical, and social interventions aimed at improving health at the local level. Although measures of socioeconomic condition tend to be highly correlated with one another, local variations in the interactions of multiple measures may reveal important anomalies that will elucidate the relationship between socioeconomic conditions and health outcomes and identify unique.

In general, the Appalachian region has lagged economically from other parts of the U.S. Relatively high levels of unemployment, low regional incomes, and educational deficits continue to contribute to a lower standard of living than enjoyed in many areas of the U.S. (Isserman, 1997). However, there are significant levels of socioeconomic diversity within Appalachia. For example, metropolitan areas in the region have more diversified economies, higher per capita incomes, and greater access to medical care than non-metropolitan areas (Barnett et al, 1998). Local socioeconomic differences within the Appalachian region may contribute to regional disparities in health outcomes since those areas having diminished access to social, economic, and medical care resources apparently experience more adverse outcomes.

The Appalachian region experiences significant excess premature mortality when compared with the non-Appalachian U.S. Given that adverse socioeconomic conditions are much of what defines Appalachia as a place, it seems likely that these conditions contribute to a generally poor health profile. Poverty, unemployment, low incomes, limited economic diversity, limited transportation and social infrastructure, rural isolation, and rugged and mountainous terrain are 'characteristic' traits for much of the region. There is, however, significant within-region diversity in both socioeconomic conditions and health outcomes (Halverson, et al, 2004). It is not presently clear, how these variations are associated with one another across the region.

This study investigates associations between measures of socioeconomic condition and rates of premature mortality for leading causes of death for counties in the U.S. with a focus on the Appalachian region. The overall goal is to elucidate relationships between observed health outcomes in the region and underlying socioeconomic conditions that may be contributing factors. This study represents a continuation of a recent study conducted for the Appalachian Regional Commission entitled *An Analysis of Disparities in Health Status and Access to Medical Care in the Appalachian Region*. The specific goals of this study were to 1) document place-based disparities in premature

mortality for leading causes of death, and 2) examine associations between rates of mortality and underlying socioeconomic conditions that may contribute to observed disparities.

This report focuses on premature mortality for several reasons. First mortality derived from death certificate data is the most comprehensive and reliable source of public health related information available. Due to the general uniformity in death certificate data, comparable mortality rates can be generated for all U.S. counties. Second, the Appalachian region exhibits considerable excesses in premature mortality when compared with the non-Appalachian U.S. and also contains some of the highest premature mortality rate counties in the U.S. (Halverson et al, 2004).

Given the broad geographic scope of this study, it represents an initial examination of these relationships. More detailed local analyses will need to be performed in order to refine our understanding of relationships between socioeconomic conditions and health outcomes at the local level.

This study represents an ecological analysis in that it deals with population level outcomes and conditions as opposed to individual level. Within local populations the influence of area-level variables such as poverty may have different influences on individuals who reside in those areas and may be dependent on other local variables such as social infrastructure, access to medical care resources, cultural norms, or local patterns of/or individual health-related behavior. Likewise, the failure to identify associations between, for example, area-level poverty and premature mortality does not mean that within those areas, individuals who live in poverty do not experience more adverse health outcomes. However, area-level measures do provide a measure of the local context and are suggestive of local conditions that contribute to population-level health outcomes. Contextual approaches examine the social conditions that affect those who share a particular environment. The social contexts of local areas are largely defined by *socioeconomic* conditions that exist in these areas. Socioeconomic conditions are tied to forms of social organization and the productive process, which are important factors influencing the collective social, physical, and mental well-being of populations. There is an inherent relationship between the quality-of-life of places and economic issues of those places. The nature of local economic industrial structure, together with their relationship to regional and national economies, have helped to define occupational and income opportunities, labor relations, social relationships and community infrastructure, local cultural practices and values, and overall levels of individual and community well-being. Although many risk factors may collectively impact a population, economic conditions provide a fundamental basis for local social context.

Theoretically, individuals of lower economic status are at greater risk of poor health because they suffer from some level of *deprivation* that results in lack of basic needs (food, clothing, housing), access to medical care and resources, access to recreational/physical activities, employment opportunities, etc. Individuals who suffer from various forms of deprivation may also be more likely to adopt higher risk health behaviors (Winkleby et al, 1999).

Collective deprivation measured by aggregate measures among populations may also reflect deleterious conditions in local areas. For example, low incomes are indicative of places with less well-developed economies, and therefore more likely to have fewer resources for public health infrastructure.

Section I.   Data and Methods

Mortality

Death certificate data for the years 1995-2001 were obtained through the National Vital Statistics System maintained by the National Center for Health Statistics. The specific causes of death which were analyzed are all-causes, heart disease, all-site cancers, and cerebrovascular disease (stroke). Heart disease, cancer, and cerebrovascular disease represent the top four causes of death nationally. Death certificate data for 1995-1997 were coded using the International Classification of Disease - 9th revision (ICD-9), and for 1998-2001 ICD-10.   For each decedent, underlying cause of death, age and county of residence at the time of death were abstracted from computerized death certificate files. These death counts were used as the numerators for calculating mortality rates. The study population consisted of all persons ages 35 to 64, who resided in the United States during the period 1995-2001. Deaths which occur in the 35 to 64 age-groups are considered premature and preventable.

Population count data for all counties in the U.S., used as denominators in mortality rate calculations, were obtained from the Bureau of the Census for the years 1995-2001. These age-specific intercensal estimates were calculated by the Bureau of the Census through extrapolation of linear trends in population growth and inter-county migration patterns.

Deaths rates were age standardized using the direct method of adjustment, with the 2000 U.S. population as the standard. County-level, age-adjusted death rates, were generated for all U.S. counties using a spatial "smoother" based on a *distance weighted, spatial moving average*. Spatial smoothing is used in this analysis to reduce the statistical variability of county death rates and to compensate for sparse populations and small numbers of deaths for some population subgroups in certain parts of the country. Spatial smoothing involves calculating spatial moving averages of county rates. Using this method, an age-adjusted rate in a single county represents an average of the mortality experience of that county and all of its neighboring counties. In this analysis the contribution of neighboring counties has been weighted by using the distance from the geographic center of each county to those of neighboring counties. Un-smoothed age-adjusted death rates were also generated and used to establish associations with socioeconomic indicators (see Section III).

Socioeconomic Condition

There are a number of variables that reflect local socioeconomic conditions including income, poverty rates, and unemployment rates. Five indicators of socioeconomic condition were selected for this analysis: Percent Urban Population, Median Family Income, Unemployment Rate, Percent of Persons Living in Poverty, and the Percent of Persons without Health Insurance. The selection of these indicators was guided by literature that has established associations between these variables and health outcomes (Brenner, 1987; Adler et al, 1994; Kahn et al, 1999, Pickle et al, 2002; Avis et al , 2006). Although these variables are often correlated with one another, local variations often occur (Wood, 2004).

These data were either directly extracted or derived from data obtained from the Area Resource File (ARF), 2005 release. The Area Resource File (ARF) was compiled by the National Center for Health Workforce Information & Analysis, Bureau of Health Professions, Health Resources and Services Administration, Department of Health and Human Services. Because relative differences in area-based socioeconomic measures are generally consistent over time, the most recent year was used for each indicator; 2000 for all but Median Family Income which was calculated for 1999.

Indicators of socioeconomic condition serve as markers representing conditions which either mediate or exacerbate social deprivation and thereby influence population health at the local level. However, it is important to recognize that area- level measures of socioeconomic condition mask important individual variations in socioeconomic status for those areas. Unemployment rates, for example, often fluctuate dramatically over time and may be dependent on local industrial infrastructure and volatility. Therefore in this analysis, unemployment represents a static point in time rather that reflect absolute differences over time. Unemployment rates do provide a general indicator of local economic development, in that, local areas with well developed and diversified economies are likely to be more resilient in times of economic distress and therefore tend to maintain lower rates of unemployment.

Unemployment

Rates of unemployment are calculated as the number of people actively seeking work divided by the total number of people in the civilian labor force. High rates of unemployment have been shown to be highly correlated with adverse public health outcomes in general (Brenner, 1987). For individuals, unemployment may result in economic hardships that limit lifestyle choices, options for health insurance, as well as access to medical care resources. When communities suffer persistently high rates of unemployment, social infrastructures that serve these communities may be difficult to establish and those that exist may break down. Unemployment is generally a key indicator of local and regional development. However, the true burden of unemployment is hard to estimate due to the fact that standard unemployment definitions do not include the long-term unemployed, often referred to as discouraged workers, or those individuals who are involuntarily employed part-time, and those not seeking work due to disability.

Income

Family income is the sum of income received by all family members in a household. Median family income indicates that point at which incomes of half of the families are higher and half are lower. Median family income has been used independently as a measure of economic development (Nielsen and Alderson, 1997) and represents a key component of aggregate socioeconomic status. Income is critical for maintaining basic life necessities as well as achieving 'quality of life' expectation and aspirations. In a society where quality of life expectations and aspirations are fueled by an increasingly ubiquitous media (magazines, television, radio), income becomes a primary factor in achieving those expectations.

Income, as an individual measure of SES reflects both the absolute level of financial resources available to each person as well as the relative level of these resources among a population of individuals. Income is critical in order to maintain life's essentials such as food, clothing, housing, medical care, and leisure activities. For local areas, a collection of individuals with high incomes in indicative of areas with relatively stable economies and translates into tax bases that are essential for maintaining public infrastructures such as schools, roads, law enforcement, social services, parks, etc. In addition, viable and stable economies are able to support a number of service and retail related industries (grocery stores, gas stations, department stores, restaurants etc.) which serve to further enhance the public 'quality–of-life'.

There are, however, several problems with using income as a collective measure of either wealth or deprivation. Income levels alone do not capture relative differences in the cost of living. Therefore it is difficult to compare the relative importance of high or low income between, for example, rural Appalachia and New York City. In addition, although median family income is a better measure of the distribution of income in a local population than the mean income, it still does not capture the local distribution of wealth. Local inequalities in income, the degree of separation between the highest and lowest income earning groups in a population, may have 'extraindividual or contextual effects that structure the social environment in ways that affect the health of a population' (Kennedy et al., 1998).

Poverty

Poverty statistics provide a basic indicator of the socioeconomic status of a population within a given area. In general, poverty is one of the most important social determinants of health and well-being. While poverty does not influence all diseases in the same way, strong positive relationships have been consistently shown between poverty and cardiovascular diseases, gastrointestinal disease, chronic respiratory disease, as well as other adverse health outcomes, including accidental and violent deaths, and suicide (Adler and Ostrove, 1999; Hopper and Guttmacher, 1979). Typically, individuals who live in poverty have limited lifestyle choices and may therefore be more susceptible to anxiety and stress associated with economic hardship. High levels of poverty may also

indicate poorly developed economies, with fewer opportunities for gainful employment. Poverty statistics provide a basic indicator of the socioeconomic status of populations within given areas and represents one of the most important social determinants of health and well-being (Adler and Ostrove, 1999; Hopper and Guttmacher, 1979).

Individuals are classified below poverty if their total individual income was less than the poverty threshold specified for the applicable family size, age of householder, and number of related children less than 18 present as defined by the federal government's official poverty definition (DeNavas-Walt et al, 2004). For this analysis the percent of persons living below the poverty threshold is used for the year 2000.

Percent Urban Population

Percent urban population is used in this study to reflect the general distribution of local populations within counties. In places where the percent urban population is low, many people are more likely to be distributed in relatively isolated rural areas and may diminished access to socioeconomic and public health resources than people living in more urban areas.

The Census Bureau defines urbanized areas as those which have a population concentration of at least 50,000 inhabitants, generally consisting of a central city and the surrounding, closely settled, contiguous territory. Included in urban population are persons living in places of 2,500 or more inhabitants outside urbanized areas (ARF, 2005). In places where the percent urban population is low, many people are likely to be distributed in relatively isolated rural areas and are more likely to have reduced access to medical care facilities that are more typically found in urban settings. Transportation infrastructures and access to public transportation are more likely to benefit urban residents than rural counterparts. Urbanized areas are also likely to have more developed social, economic, and political, and medical care infrastructures to serve local populations.

Percent of Persons without Health Insurance

Given the rising costs of medical care in the U.S., health insurance is critical for many individuals and families seeking both primary preventive care and treatment for injury and disease. In areas where large proportions of individuals do not have health insurance, medical care costs may be prohibitive and result in neglect of primary preventive care that reduces the onset and impact of disease.

The 2000 Estimates of Persons with and without Health Insurance; and Estimates of Persons with and without Health Insurance under age 18 data are from the Bureau of Census' *Small Area Health Insurance Estimates (SAIHE)* file.

These measures of socioeconomic condition provide a general context within which individuals live their daily lives and within which institutions and regulatory systems are developed that are related to health care, education, public safety, working conditions, and local and regional economic development. Although region-wide these variables are highly correlated, local differences in the interaction of these variables may help to explain differential rates of premature mortality in areas which are similar on a single (or several) indicators.

Measures of Association

The focus of the majority of research which links socioeconomic status/conditions and health outcomes is in seeking generalized associations. For this reason, associations are usually derived in aggregate. For example, if a socioeconomic information and health status information are available for individuals, these data are 'lumped' together in order to identify the general pattern of association (high poverty = poor health). Less attention is usually given to anomalies in established associations (high poverty = good health), because the majority of observations follow a general pattern. Similarly, studies which examine area-level data often use a similar methodology. For example, when examining county-level indicators of socioeconomic status against health outcomes for those counties, a general approach is to aggregate the data to derive a general association/pattern (high county poverty = high county death rate).

The following analysis relies on a measure of bivariate spatial autocorrelation to assess associations between rates of premature mortality and indicators of socioeconomic conditions. The underlying assumption is that the local patterns of premature mortality will be associated with the underlying socioeconomic conditions of the local area. The specific method uses Local Indicators of Spatial Association (LISA) to derive localized estimates of association (Anselin 1995). LISAs are a measure of spatial autocorrelation. Spatial autocorrelation is traditionally is used to assess the degree to which values of one variable are similar among locations which are close to one another. In this regard LISAs aid in the identification of local clusters. For this analysis a bivariate implementation of the LISA is calculated to assess the degree to which values of one variable (premature mortality) are similar to values of another variable (socioeconomic condition) among locations in close proximity.

LISAs are calculated as a Local Moran Statistic using the following formula:

$$I_{i,(x,z)} = m_{i,x} \quad W_i \, m_{i,z}$$

where: $m_{i,x}$ is the z-score standardized dataset (x) being tested for region i.
           $m_{i,z}$ is the z-score standardized dataset (z) being tested for region i.
and W_i is the spatial weight set defining the neighborhood

The unit of analysis is a neighborhood of counties defined by a ***spatial weights matrix***. For these analyses the spatial weights matrix has been defined by first-order contiguity. In other words, each 'neighborhood' is defined by a county and its' immediate neighbors with which it shares a boundary. First-order contiguity was selected for these analyses in order to minimize the local neighborhood and thereby maximize the local specificity of observed associations. Underlying these analyses is the basic assumption that areas (counties) are more like those in close proximity than those that are further away.

Using bivariate LISAs, associations are examined for overall premature mortality as well as disease-specific premature mortality. The distribution of each variable is transformed to z-scores to calculate the LISAs in order to take advantage of the symmetric properties of the standard- normal distribution. Z-scores are obtained by calculating the number of standard deviations that each value is from the distribution mean.

One of the key features of this method is that critical values of local spatial autocorrelation are identified by plotting the value of the first variable at each location against the weighted average of the comparison value (where the weights are determined by the spatial weights matrix). This plot is referred to as the **Moran Scatterplot** and an example is provided in Figure 1.

Figure 1. Moran Scatterplot of Local Associations between Premature Heart Disease Mortality and Percent of Persons Below Poverty level.
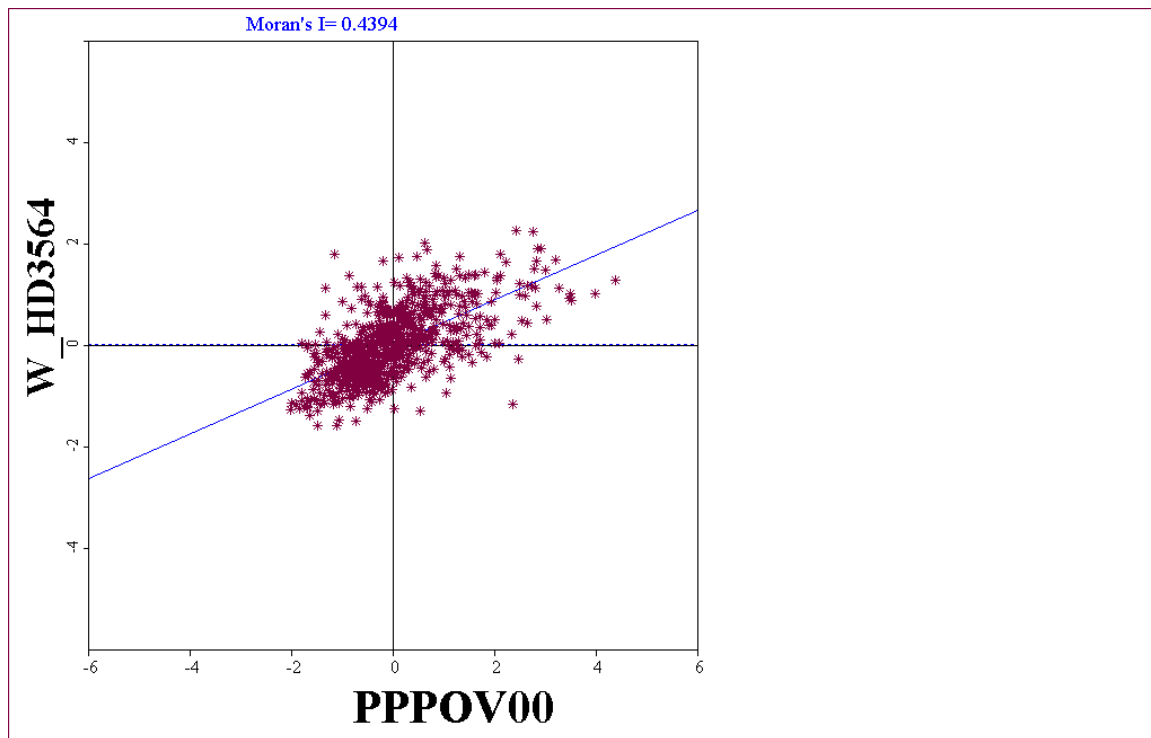
Key associations are identified in each of the four quadrants of the Moran Scatterplot. For example, the upper right quadrant identifies areas where high values on one variable are associated with high values on the other variable (High-High), the lower left quadrant identifies areas where low values on one variable are associated with low values on the other variable (Low-Low).

Due to the dependence on the global mean, separate analyses were conducted for the U.S. and Appalachian region. For the analyses of Appalachia, a region was constructed by including counties whose boundaries were within 50 miles of the A.R.C. designated counties in Appalachia. This was done to ensure an adequate number of neighbors were included in the calculation of LISAs for the region. For this study LISAs were generated for the .05 level of statistical significance**.

The inclusion of counties within 50 miles of the ARC designated boundary, also includes a number counties that contain major U.S. cities in addition to Appalachian cites of Pittsburgh and Birmingham. It is possible that the inclusion accentuates the extremes of the distribution of socioeconomic factors used in this analysis (see Appendix 1).

**Statistical significance of associations is determined by Monte-Carlo randomizations of the Local Moran statistic (I) value. For this analysis 999 permutations are used to establish statistical significance. Only those associations which are significant at the .05 level of significance are used in this analysis.