Objective To account for variations in death rates in population subgroups of the USA.

Methods Factors associated with age-adjusted death rates in 366 metropolitan and non-metropolitan areas of the United States were examined for 1990–92. The rates ranged from 690 to 1108 per 100,000 population (mean = 885 ± 78 per 100,000).

Findings Least squares regression analysis explained 71% of this variance. Factors with the strongest independent positive association were ethnicity (African-American), less than a high school education, high Medicare expenditures, and location in western or southern regions. Factors with the strongest independent negative associations were employment in agriculture and forestry, ethnicity (Hispanic) and per capita income.

Conclusion Additional research at the individual level is needed to determine if these associations are causal, since some of the factors with the strongest associations, such as education, have long latency periods.

Keywords Mortality rate; Socioeconomic factors; Ethnic groups; Geography; Health services accessibility; Regression analysis; Analysis of variance; United States (source: MeSH, NLM).

Methods

The study population consisted of 320 primary metropolitan statistical areas (MSA) and 46 areas that included non-metropolitan counties within state boundaries (four states had no non-metropolitan counties). Data were aggregated from the county level to the MSA level, using the definition of MSA boundaries that were in effect in 1996. All remaining non-metropolitan counties were pooled into one non-metropolitan area for each state. These geographical units are referred to as “metropolitan statistical areas or non-metropolitan balance of states” (MSA/NBS), and were chosen as units of analysis because of their intermediate size between states and counties.

The independent contribution of demographical and socioeconomic factors, as well as medical care access, to variation in age-adjusted death rates across the 366 MSA/NBS was estimated using linear regression analysis of data from the Bureau of Health Professions Area Resource File (10). The Area Resource File is composed mostly of summed county totals and population-weighted averages for adjacent counties within each of the multi-county areas. The dependent variable is the annual number of deaths per 100,000 population averaged over 1990, 1991 and 1992, and age-standardized by applying local age-specific death rates to the 1990 US population age distribution.

We chose to analyse differences in death rates using variables known to exhibit strong geographical gradients, including racial or ethnic identity (11), socioeconomic status (12), rural/urban differences (13, 14) and medical services (8). Specific independent variables were: census region; gender; racial/ethnic composition (such as Black and Hispanic, which...
are not mutually exclusive categories); socioeconomic composition (e.g. the proportion of adults aged 25 years or older who completed high school; per capita income; percentage unemployed; and the Gini-coefficient of inequality in household incomes in 1990 (15); urban/rural composition (metropolitan/non-metropolitan dummy variable and percentage of labour force in agriculture, forestry, and fisheries); and medical service measures (physician-to-population ratio, Medicare payments per person over 65 years of age, and the number of hospital inpatient days per person for short-term general medical and surgical procedures). The dependent variable was age-standardized, but not the independent variables, which may produce biased regression coefficient estimates (16). An alternative method suggested by Rosenbaum and Rubin for controlling for age structure and reducing bias led to similar results and conclusions. Nevertheless, this reinforces the imprecise nature of estimates produced from aggregate data.

To reduce the effects of multi-colinearity among several regressors, we constructed a unidimensional scale from these highly correlated variables using principal components factor analysis (17). The unidimensional scale of socioeconomic disadvantage included percentage income, the Gini coefficient of household income, percentage unemployed, and percentage with fewer than 12 years of schooling (Chronbach’s alpha factor = 0.76). Increasing values of the factor are associated with increasing levels of economic disadvantage. All 366 cases were used in the analysis, since outliers showed little influence, with one exception: in the equation that included physicians per capita, two cases showed extremely high levels of physicians and extremely low levels of mortality: the Iowa City, Iowa MSA and the Rochester, Minnesota MSA, both of which are small cities with large hospitals and specialty clinics. With these outliers excluded, there was no association between physician supply and mortality (see footnote to Table 2). This was the only major change in coefficients due to the influence of extreme outliers.

After estimating the regression equations, we used the final model to generate hypothetical predicted values of mortality, assuming associated characteristics were held equal to the national average at all places. Mortality values for each MSA/NBS were calculated as the sum of the regression residual, plus the predicted value of the regression equation with one or more regressors of interest fixed at the national average; remaining regressors could take the observed values at each place. Using the ArcView GIS system, these predicted values were then used to produce maps of hypothetical death rates with associated variables equalized across all areas.

### Results

Table 1 shows the summary statistics for the variables in the analysis. The age-adjusted death rate had a mean value of 885 per 100,000 population, with a standard deviation of 78 and a range of 690–1108. The results of the linear regression with all 14 independent variables ranked by the size of their positive and negative effects are shown in Table 2. The unstandardized coefficients indicate the average difference in age-adjusted death rates between MSA/NBS that differ by one unit of the independent variable. For more accurate comparison we also show standardized coefficients which indicate the average number of standard deviations difference in age-adjusted death rates between MSA/NBS that differ by one standard deviation of the independent variable. Typically, standardized coefficients less than 0.10 (in absolute value) are “trivial” or have non-significant t-statistics.

From the standardized coefficients, we see that the variables with the largest relative association with mortality are African-American ethnicity (0.44) and adults with less than high school education (0.29). From the unstandardized coefficients, a difference between geographical areas of one percentage point in Black population is associated with an average increase of 3.32 deaths per 100,000 population; and for adults with less than high school education, a difference of one

<table>
<thead>
<tr>
<th>Variable</th>
<th>N</th>
<th>Mean</th>
<th>SD</th>
<th>Minimum</th>
<th>Maximum</th>
</tr>
</thead>
<tbody>
<tr>
<td>Age-adjusted death rate (per 100 000 population)</td>
<td>366</td>
<td>885.19</td>
<td>77.76</td>
<td>689.67</td>
<td>1108.31</td>
</tr>
<tr>
<td>Female (%)</td>
<td>366</td>
<td>50.26</td>
<td>1.09</td>
<td>39.97</td>
<td>52.49</td>
</tr>
<tr>
<td>West</td>
<td>366</td>
<td>0.20</td>
<td>0.40</td>
<td>0.00</td>
<td>1.00</td>
</tr>
<tr>
<td>Midwest</td>
<td>366</td>
<td>0.24</td>
<td>0.43</td>
<td>0.00</td>
<td>1.00</td>
</tr>
<tr>
<td>South</td>
<td>366</td>
<td>0.40</td>
<td>0.49</td>
<td>0.00</td>
<td>1.00</td>
</tr>
<tr>
<td>Income per capita</td>
<td>366</td>
<td>17.09</td>
<td>3.27</td>
<td>8.90</td>
<td>32.34</td>
</tr>
<tr>
<td>Gini coefficient</td>
<td>366</td>
<td>0.41</td>
<td>0.02</td>
<td>0.36</td>
<td>0.48</td>
</tr>
<tr>
<td>Unemployment (%)</td>
<td>366</td>
<td>5.71</td>
<td>1.99</td>
<td>1.65</td>
<td>19.33</td>
</tr>
<tr>
<td>Less than a high school education (%)</td>
<td>366</td>
<td>24.32</td>
<td>7.08</td>
<td>8.69</td>
<td>53.39</td>
</tr>
<tr>
<td>African-American (%)</td>
<td>366</td>
<td>9.80</td>
<td>10.24</td>
<td>0.08</td>
<td>45.76</td>
</tr>
<tr>
<td>Hispanic (%)</td>
<td>366</td>
<td>6.94</td>
<td>12.44</td>
<td>0.32</td>
<td>93.90</td>
</tr>
<tr>
<td>Agriculture, forestry or fishing worker (%)</td>
<td>366</td>
<td>2.95</td>
<td>2.58</td>
<td>0.32</td>
<td>16.26</td>
</tr>
<tr>
<td>Non-metropolitan resident</td>
<td>366</td>
<td>0.13</td>
<td>0.33</td>
<td>0.00</td>
<td>1.00</td>
</tr>
<tr>
<td>Doctors per 100 000 population</td>
<td>366</td>
<td>179.51</td>
<td>111.82</td>
<td>61.01</td>
<td>1523.01</td>
</tr>
<tr>
<td>Doctors per 100 000 population (minus outliers)</td>
<td>364</td>
<td>173.46</td>
<td>74.37</td>
<td>61.01</td>
<td>670.48</td>
</tr>
<tr>
<td>Hospital days per capita</td>
<td>366</td>
<td>1.03</td>
<td>0.45</td>
<td>0.27</td>
<td>3.90</td>
</tr>
<tr>
<td>Medicare payments per person over 65 yrs old (US$ 1000s)</td>
<td>366</td>
<td>3.06</td>
<td>0.51</td>
<td>1.77</td>
<td>4.81</td>
</tr>
<tr>
<td>Socioeconomic disadvantage factor</td>
<td>366</td>
<td>0.00</td>
<td>1.52</td>
<td>-3.17</td>
<td>8.14</td>
</tr>
</tbody>
</table>

* Number of observations.

b SD = Standard deviation.
percentage point is associated with an average increase of 3.17 deaths per 100,000 population, all else being equal.

Other factors that have independent positive associations (in order of standardized magnitude) are: Medicare payments per capita; location in the west and south regions (compared to the northeast region); hospital days per 1000 population; and non-metropolitan areas. Those with negative associations are: working in agriculture, forestry, and fishing; Hispanic ethnicity; per capita income; the Gini coefficient; and physicians per 100,000 population. On average, MSA/NBS in the south have 75.6 more age-adjusted deaths per 100,000 population than those in the northeast region. This model explains 31% of the variance.

Excluding these outliers, the association fell from a statistically significant –0.16 to a non-significant –0.03 (fewer deaths for each standard deviation difference in physician supply). This was the only major change in coefficients due to the outliers.

In Model 4, MSA/NBS with higher-than-average socioeconomic disadvantage scores also have higher-than-average mortalities, with an average difference of 18.79 deaths per 100,000 population for every standard deviation difference in the socioeconomic disadvantage factor. This model accounted for 60% of the variance.

In Model 5, the effect of the supply of physicians was examined. This component has a small negative effect on mortality, with 0.06 fewer deaths for each physician added per 100,000 population. The variance explained only increases by 1% to 0.61. However, when two extreme outliers were excluded, this effect changed to a non-significant positive 0.02.

Finally, in Model 6 independent differences in death rates between metropolitan areas (MSA) and non-metropolitan areas (NBS) were examined. On average, there were 8.83 fewer deaths per 100,000 population for non-metropolitan areas, but the effect is small relative to the standard error. The variance explained does not increase from Model 5 to Model 6.

Two maps were produced to illustrate these effects more clearly (Fig. 1 and Fig. 2). Fig. 1 shows the baseline distribution of age- and gender-adjusted death rates, with generally higher rates in the south and lower rates in the Midwest. Nevada has an unusually high mortality rate, as does West Virginia, even given the generally high mortality characteristic of the South. Texas, Oklahoma and Missouri have average death rates, in contrast to the deep south-east, and Florida has lower-than-average death rates. Fig. 2 shows the impact of assigning average levels of race/ethnicity and socioeconomic disadvantage, in addition to age and gender, for all areas. The most obvious effect is the improvement in the south-east, with the expected death rate in Tennessee, Kentucky and West Virginia.
falling to the average range, and that in the deeper south-east even falling below average.

**Discussion**

This analysis highlights the large, non-random geographical variation in age-adjusted death rates in the United States. Much attention devoted to the quality of the health care system necessarily focuses on process measures (18, 19), but large differences in such a fundamental outcome as mortality should stimulate a re-examination of measurement priorities to identify the causes of variation. Although the differences reported here focus only on the mortality component, if a health-related quality of life measure were added to mortality, the variation would undoubtedly be greater (20, 21). For example, the five-year difference in male life expectancy across British social classes increased to a difference of nine quality-adjusted life years when the EuroQuol measure was combined with the life year component (22). It is therefore essential that data collection and analysis of health-related quality of life make information on variation available.

---

**Table 3. Linear regression of age-adjusted death rates on selected independent variables**

<table>
<thead>
<tr>
<th>Variable</th>
<th>Model 1</th>
<th>Model 2</th>
<th>Model 3</th>
<th>Model 4</th>
<th>Model 5</th>
<th>Model 6</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Coeff</td>
<td>SE</td>
<td>Coeff</td>
<td>SE</td>
<td>Coeff</td>
<td>SE</td>
</tr>
<tr>
<td>Constant</td>
<td>348.35</td>
<td>171.01</td>
<td>632.69</td>
<td>145.86</td>
<td>628.90</td>
<td>144.70</td>
</tr>
<tr>
<td>Female (%)</td>
<td>10.13</td>
<td>3.39</td>
<td>4.00</td>
<td>2.90</td>
<td>4.15</td>
<td>2.87</td>
</tr>
<tr>
<td>South</td>
<td>75.60</td>
<td>171.01</td>
<td>29.16</td>
<td>9.52</td>
<td>26.64</td>
<td>9.16</td>
</tr>
<tr>
<td>African-American (%)</td>
<td></td>
<td></td>
<td>4.39</td>
<td>0.36</td>
<td>4.21</td>
<td>0.36</td>
</tr>
<tr>
<td>Hispanic (%)</td>
<td></td>
<td></td>
<td>-0.66</td>
<td>0.25</td>
<td>-1.59</td>
<td>0.25</td>
</tr>
<tr>
<td>SDFc</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>18.79</td>
<td>2.14</td>
</tr>
<tr>
<td>Doctors per 100 000</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Non-metropolitan resident</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>R-squared</td>
<td>0.31</td>
<td>0.51</td>
<td>0.52</td>
<td>0.60</td>
<td>0.61</td>
<td>0.61</td>
</tr>
</tbody>
</table>

a Coeff = linear regression coefficient.
b SE = standard error.
c SDF = socioeconomic disadvantage factor.
The utility and limitations of ecological analyses are described elsewhere (23–25). The present analysis cannot approximate relationships at the individual level; instead, relationships between aggregate characteristics are described. Consequently, the correlations described in this paper should not be used to predict effects of interventions at the individual level, without confirmation by individual-level data; and cross-sectional correlations should not be interpreted as causal relationships. The relevance of some ecological or community-level variables, such as income inequality, should not be ignored however, since these variables only (or partly) have a causal effect at an ecological level (26, 27). Indeed, it has recently been suggested that geographical small area analysis may hold the greatest promise for studying the variation in social group health differences (28).

Another limitation of this cross-sectional study is that it does not reflect the different latent periods that are responsible for some of the associations (29, 30); it is almost certain that the association between education and mortality has a very long latent period, and therefore data from earlier periods might display different (and perhaps more valid) relationships. In addition, using age-adjusted death rates masks the mortality at different stages of the lifespan, and can be dominated by mortality at higher ages. Subsequent studies should explore these associations at different life stages (31, 32). It is also possible to think of geographical populations as static, even though the demography is the result of births, deaths, and in and out migration; and analyses to control for the impact of migration on mortality are needed. A recent analysis (33) showed that male migration accounted for nearly all of the differences in death rates at the British district level. Finally, our decision to group together all non-metropolitan counties in a state could mask variation across these counties; future analysis might use some other geographical unit of analysis such as hospital service areas (34).

These issues notwithstanding, our analysis reveals associations that account for 71% of the variance in death rates. While this is considerable, almost a third of the variance is unexplained. A fuller model (e.g. 5) would include variables from the genetic, physical environment and individual behaviour domains. We are currently attempting to organize data, such as air pollution levels, rates of crime, obesity and smoking, for further analysis. In this study, the strongest association of age-adjusted death rate is with the percentage of the population that is African-American. While this is not a new finding (35–37), the strength of this association is independent of income, education, physician supply and region, and deserves attention. Additional variables, such as smoking rates, may explain this association, although they may be closely correlated with other factors in this analysis. The chronic stress induced by racial tension (38, 39) might also be more important than commonly acknowledged. Also, measures that are subject to respondents’ willingness to self-identify with broad social categories, such as African-American and Hispanic (40), may mask significant intraracial differences. However, given the nature of racial mixing in the United States and the renegotiation of ethnic identities, even these categories may lose meaning in the future (11).

The effects of education level and other socioeconomic variables are consistent with previous results on the socioeconomic gradient in mortality (41–43), in contrast to the negative association for the Gini coefficient of income inequality (26, 44). We also expected to find rural/urban differences, but the favourable association of the percentage of labour force working in traditional rural occupations (agriculture, forestry, and fishing), and the opposite independent
Causal mechanism of education on health outcomes

The causal mechanism of education on health outcomes is expected to vary among different contexts. While related social factors underlying education may be. Certainly, socioeconomic factors will have causal implications at the individual level, and thus have policy implications as well. Race per se is not subject to policy manipulation, although related social factors underlying it may be. Certainly, education has health and other social benefits and is one variable that is the subject of local and national policy. While the causal mechanism of education on health outcomes requires more research, the length of its latent period calls for policy attention if improvements in health outcomes are to be accomplished as quickly as possible. The lack of significance of some medical care variables does not mean they are not important in improving health outcomes: they have been important in achieving current health outcomes and will continue to be important. But more detailed analysis is needed to determine which aspects of medical care will contribute most to health outcomes in the coming decades.

Independent of all other regressors in our models, the south and west regions tend to have higher mortality, and the Midwest lower mortality, compared with the north-east region. However, these differences are sensitive to model specification and with successive variable additions, the regions are affected in different ways. The high mortality in the south, so readily apparent in a univariate map (Fig. 1), almost completely disappears when race and socioeconomic differences are controlled, while the lower mortality in the Midwest improves even more. This “Midwest advantage” requires further examination to determine what factors, perhaps genetic or even more. This “Midwest advantage” requires further examination to determine what factors, perhaps genetic or behavioural, are responsible.

While ecological and cross-sectional analysis cannot identify causal relationships, it is likely that race and socioeconomic factors will have causal implications at the individual level, and thus have policy implications as well. Race per se is not subject to policy manipulation, although related social factors underlying it may be. Certainly, education has health and other social benefits and is one variable that is the subject of local and national policy. While the causal mechanism of education on health outcomes requires more research, the length of its latent period calls for policy attention if improvements in health outcomes are to be accomplished as quickly as possible. The lack of significance of some medical care variables does not mean they are not important in improving health outcomes: they have been important in achieving current health outcomes and will continue to be important. But more detailed analysis is needed to determine which aspects of medical care will contribute most to health outcomes in the coming decades.

Conclusion

The possibility of a comprehensive health production function, the “fantasy equation” (48), that relates multiple determinants of health to health outcomes, has been discussed. Some of the methodological difficulties that compromise this approach have already been mentioned, but this approach is more difficult when the factors of production are both public and private, and collaboration across sectors is difficult or impossible (49, 50). The amount of effort that has gone into understanding the health production function is low, given its importance for public and private policy and expenditure (57). Recent global and national efforts to increase support for a broader understanding are welcome, but more will be needed if the goal of optimizing and making more equitable the distribution of length and quality of life to all persons is to be realized in the new millennium.

Acknowledgements

We wish to acknowledge partial support from the Milbank Memorial Fund, the National Institute of Mental Health (NIMH Trainee Grant Program, Grant Number T32-MH18029) and the National Institute of Aging. We also express our appreciation to Judy Knutson for clerical assistance.

Conflicts of interest: none declared.

Résumé

Variation des taux de mortalité dans des sous-populations des Etats-Unis d’Amérique

Objectif Tenir compte des variations des taux de mortalité dans des sous-groupes de population des Etats-Unis d’Amérique.

Méthodes Les facteurs associés aux taux de mortalité ajustés sur l’âge dans 336 zones métropolitaines et non métropolitaines des Etats-Unis d’Amérique ont été examinés pour la période 1990-1992. Ces taux allaient de 690 à 1108 pour 100 000 habitants (moyenne: 885 ± 78 pour 100 000).


Conclusion Des recherches complémentaires au niveau individuel sont nécessaires pour déterminer si ces associations sont causales, car certains des facteurs présentant les associations les plus fortes, comme le niveau d’études, ont une longue période de latence.

Resumen

Diferencias entre las tasas de mortalidad de subpoblaciones en los Estados Unidos

Objetivo Dar cuenta de las diferentes tasas de mortalidad de subgrupos de población en los Estados Unidos.

Métodos Se examinaron los factores asociados a las tasas de mortalidad ajustadas por edades para 1990-1992 en 366 áreas metropolitanas y no metropolitanas de los Estados Unidos. Las tasas iban de 690 a 1108 defunciones por 100 000 habitantes (media = 885 ± 78 por 100 000).

Resultados El análisis de regresión por el método de los mínimos cuadrados explicó el 71% de esa diferencia. Los factores para los que se halló una mayor asociación positiva independiente fueron el origen afroamericano, la carencia de estudios secundarios, unos gastos elevados para Medicare, y la residencia en regiones del sur y el oeste del país. Los factores con una mayor asociación negativa independiente fueron el...
empieza en la agricultura o silvicultura, el origen hispanoamericano y los ingresos per cápita.

**Conclusión** Es preciso realizar nuevas investigaciones a nivel individual para determinar si esas asociaciones tienen carácter causal, pues algunos de los factores más fuertemente asociados, como la educación, tienen periodos de latencia largos.

**References**