Inequality, race, and mortality in U.S. cities: A political and econometric review of Deaton and Lubotsky (56:6, 1139–1153, 2003)

Michael Ash a, *, Dean E. Robinson b

a Department of Economics and CPPA, University of Massachusetts Amherst, Amherst, MA 01003, United States
b Department of Political Science and CPPA, University of Massachusetts Amherst, Amherst, MA 01003, United States

ARTICLE INFO

Article history:
Available online 3 April 2009

Keywords:
Inequality
Race
Health inequality
Mortality
Geography
USA

ABSTRACT

Our replication of Deaton and Lubotsky’s (2003) study “Mortality, Inequality and Race in American Cities and States” identifies a coding error in the econometric analysis in the original paper. Correcting the error changes the basic results of the paper with respect to inequality and mortality in a relevant and substantive way. We also propose an alternative interpretation of the other main result of the paper regarding racial composition and mortality.

© 2009 Elsevier Ltd. All rights reserved.

Inequality and health

There is considerable support, e.g., in publications by Babones (2008) and Clarkwest (2008) in this journal, for the view that holding all else equal, including average income and the extent of absolute deprivation, a wider distribution of income causes worse health at the individual level. Proposed mechanisms for the relationship include the effect of inequality on investment in salutary public goods; on exposure to environmental hazards; on trust and other elements of social capital that promote health; directly on biological functioning by inducing stress. Some of the strongest results, and also some of the most contested findings, have come from the analysis of U.S. data (Backlund et al., 2007; Lochner, Pamuk, Makuc, Kennedy, & Kawachi, 2001; Mellor and Milyo, 2001). In this note, we address a paper offered as a refutation of the inequality–health hypothesis, Deaton and Lubotsky (2003), hereafter DL, which considers racial composition of counties and cities in lieu of income inequality as an alternative social–ecological correlate of poor health. Their paper examines inequality in U.S. Metropolitan Statistical Areas (MSAs) in relation to MSA-level age-adjusted mortality rates. Although the relationship studied is ecological, the scale of analysis is at high geographic resolution by the standards of the literature. DL finds a strong positive bivariate relationship between inequality and white-male mortality, but the relationship disappears completely when the multivariate analysis includes controls for percent black, which itself has a strong positive relationship with white-male mortality.

We offer two critiques of DL. First, we show that the final null result for income inequality in DL is unstable with respect to econometric specification. We examine three parts of the econometric analysis in the original paper: weighting of MSAs by population size; measurement of average income; and measurement of inequality with the gini coefficient. The results are generally insensitive both to the definition of income and to the method of constructing the gini, but the results are highly sensitive to weighting. DL made a minor coding error in weighting; alternative approaches, including DL’s intended weighting, highlight the sensitivity.

Second, and more importantly, we argue that posing racial composition as an alternative to inequality misses the social and political meaning of race. We argue that racial composition is itself a form of inequality. Thus, neither inequality nor race (modeled as “fraction black”) is adequately theorized in DL.

Data sources and weighting alternatives

Drs. Deaton and Lubotsky graciously and generously provided their datasets, full documentation, and code.

The source of health outcome data for DL and for our replication is the Compressed Mortality File (CMF) aggregated by the National Center for Health Statistics (NCHS) from individual death	

* Corresponding author. Tel.: +1 413 545 6329; fax: +1 413 545 2921. E-mail address: mash@econs.umass.edu (M. Ash).

0277-9536/$ – see front matter © 2009 Elsevier Ltd. All rights reserved.
doi:10.1016/j.socscimed.2009.02.038
certificates submitted by U.S. counties. The CMF reports counts of
deaths by county, race, sex, age, and cause. The main outcome for DL is the age-adjusted mortality rate for all causes of death for each race–sex group for the geographic area. Age adjustment of mortality rates means applying age-specific crude death rates for each geographic area to the same hypothetical population with a fixed age distribution, e.g., the total U.S. population. The number of deaths that would have occurred in the region in question if it had its actual death rates but the U.S. age structure is used as the numerator instead of the actual number of deaths. Although DL examines states as well as Metropolitan Statistical Areas (MSAs), the unit of analysis in our replication is the MSA. MSAs constitute aggregations of one or more counties in all regions except New England, where we employ the New England County Metropolitan Area (NECMA) instead of MSAs. Because counties are the building blocks of MSAs, aggregation of CMF data from counties to MSAs is straightforward.

In the portion of the paper that we address, DL uses 1990 U.S. Census Public Use Micro Sample (PUMS) data as the basis of the social, economic, and demographic correlates of the mortality outcome. The PUMS is a 5% sample of household data reported on long-form records from the decennial U.S. Census of Households and Population. The long form is submitted by one-sixth of households; hence, the 5% sample uses data from roughly 1/120 of U.S. households. There are advantages and disadvantages to the use of the PUMS for the analysis.

The main advantages of the PUMS data are the data permit the computation of "equivalent household income" adjusted for household size, the measurement of sex- and race-specific distribution of educational attainment, the measurement of sex-specific earnings inequality, the specification of all measurements for non-Hispanic whites and the computation of a "smooth" gini coefficient based on individual data. The main disadvantages of the PUMS are that the PUMS geography, which is based on the Public Use Micro Area (PUMA), does not perfectly map to Metropolitan Statistical Areas but requires a best-estimate matching algorithm (described in detail in the web-published appendix to DL) and that the 5% sample may result in thin data at the tails of the income distribution, which is particularly relevant for precise measurement of inequality.

We use the 1990 U.S. Census Summary Tape File 3C (STF-3C) data which report statistics tabulated at the county level for the full Census. We aggregate counties to the Metropolitan Statistical Area (and use New England County Metropolitan Areas, or NECMAs, in New England) and compute all social, economic, and demographic variables, including the gini coefficient for household income, for each MSA.

Because the STF-3C data report aggregates, our income and inequality measures cannot include "equivalent income" adjustment for household size, and so our inequality measure implicitly (and incorrectly) treats all households as equal-sized units. Our inequality measure for whites includes white Hispanics. For all specifications, we also estimated models that control for percent Hispanic and the results are substantively unaffected.

Our inequality measure is based on lumpy categories and, hence, our gini interpolates within categories. The categorical household income data in the 1990 STF-3C are reported with household counts in 25 categories of annual household income: by $5000 increments from $0 to $10,000, by $2500 increments from $10,000 to $50,000, by $5000 increments from $50,000 to $60,000, and then $60,000–74,999, $75,000–99,999, $100,000–124,999, $125,000–149,999, $150,000 and up. The actual mean income for the top category, not subjected to top-coding, is reported in the Census, which means that the income of the very rich is well accounted in the STF-3C data, while in the PUMS data high incomes are top-coded. Except for the top income category, the computation of the gini from categorical data requires assumption regarding within-category mean incomes, and we tried alternative assumptions for within-category mean incomes (quarter-way between the lower threshold and upper threshold, mid-point of the category). The results were not sensitive to the alternative assumptions regarding within-category mean. Although our gini measure of inequality is inferior with respect to adjustment for household size and use of categorical data, the STF-3C data may include better estimates of the income distribution among the very rich both because the STF-3C data accurately reports high incomes and because they use the full Census (of the one in six long form) rather than a 5% sample.

We also test alternative measures of average metropolitan income. DL uses the mean of log equivalent income by race. Because the STF-3C data are aggregated, we use log of per capita income by race and log of median household income by race. In both DL and our replication, the mean of average income is race specific under the assumption that individual income is self-protective. We hypothesize that the results should be insensitive to the definition of income used.

We find a seemingly trivial coding error in DL with respect to the weighting of regressions in the original publication. Although the text states, "All regressions are weighted by the square root of the relevant population" (p. 1142) the code that DL provided indicates that regressions were in fact weighted by the square root of the square root, or fourth root, of the relevant population. Sampling theory suggests that the standard deviation of the residual will be proportional to the square root of the sample size for each population (metropolitan area) sampled. The implication is that weighting observations by the square root of the sample size addresses heteroskedasticity and thus maximizes the efficiency of estimates. When the originally specified weight is used, the results are substantially more favorable to the inequality–mortality hypothesis.

However, the DL text elsewhere states, "We have also repeated Table 2 without using population weights. This is not our preferred strategy—the MSAs are of very different size, and the regressions should be thought of as on an individual basis not an MSA basis" (p. 1146). This argument implies that, rather than address heteroskedasticity induced by sampling error, the DL weighting scheme is intended to make the regression representative of the average individual's experience of inequality. However, the weighting scheme implied by this intention is absolute, rather than square root, population weights. (We are grateful to an anonymous referee for making this observation.)

To the extent that we have a priori construct, we view the inequality–mortality relationship as a matter of context. People in a city are subjected to the city's inequality and healthfulness, which are parts of the larger societal context in which they live. If more people live in egalitarian and healthful cities, then more people will benefit from its amenities, but the relationship between the qualities pertains to the city, not the person. This interpretation would imply identifying the coefficients from the unweighted, or MSA, specification. The relationship thus identified could then be used as a predictor for all persons based on their cities of residence to determine the "average effect of inequality on health."

Yet our expectation is that the effect will be more apparent in larger geographic aggregates. Reviews of the literature (Hsieh & Pugh, 1993; Subramanian & Kawachi, 2004; Wilkinson & Pickett, 2006) have discussed the tendency to observe more positive findings in large areas compared to small ones, and Wilkinson (1997) has argued that inequality must be understood as deprivation with
respect to the wider society. This interpretation would imply giving additional importance to larger metropolitan areas.

We see the choice of weighting as answering the question “from which observations do we want to learn?” In classically distributed data, the answer is “from all of them equally.” In heteroskedastic data, the answer is “from the more precise ones in proportion to their precision.” If the data differ in importance, the answer may be “from the important ones in proportion to their importance.” In the DL model, the choice of weighting is between listening equally to all cities and listening more to big cities. Weighting all cities equally appeals because there was no prior case that the contextual relationship should apply in some places but not in others (hence, no weights). Increasing the weight of observations from big cities appeals both for the reason of increasing precision in the presence of sampling-induced heteroskedasticity (hence, square-root weights), because that’s where the people are (hence, full weights), or because the effects of inequality are more manifest in more aggregated areas.

In addition to reporting regression results for the corrected code with the square root of population, we also present results both for unweighted regressions, which are hostile to the inequality–mortality hypothesis, and for regressions weighted with absolute population, which strongly support the inequality–mortality hypothesis.

Although DL reports results for all races combined and for white women, both DL and subsequent citations focus largely on white males in the discussion. Following the emphasis in DL, we replicate the results for white-male mortality rates.

Results

In Table 1 we report summary statistics for the relevant variables and in Table 2 we report the bivariate correlations between the variables. Our estimated mortality rates and fraction black were similar to those in DL. The alternative gini (gini 2) was highly correlated with the gini (gini 1) in DL ($r = 0.85$ reported in Table 2). Our gini was on average higher than that in DL, which would be consistent with under-measurement of the very rich in the PUMS data. The mean of gini 1 was 0.386, while the mean of gini 2 was 0.423. The three alternative income measures were also highly correlated (the lowest bivariate correlation is $r = 0.86$). Both gini measurements have modest bivariate correlations with mortality and fraction black has a somewhat stronger correlation with mortality. Fraction black and the two ginis, which measure inequality among white men, are correlated but not overwhelmingly so, with $r$ of roughly 0.35 between each and fracture black. Fraction black is highly correlated with the full population gini, which includes the racial difference in income, but this correlation is not reported here.

Each row in Table 3 reports a regression result from the DL publication or from our replication. In all specifications, the sample is the 287 Metropolitan Statistical Areas in 1990. The dependent variable is the log odds of age-adjusted mortality, which means that coefficients can be interpreted as the percent change in the probability of dying for a unit increase in the dependent variable. The explanatory variables are mean log equivalent income for white men and the gini coefficient, again for white men. For example, the result in row [1] implies that a 10% increase in income (equivalent income measured in natural log) would be associated with a 0.6% decrease in the annual probability of death among white men. Income has its expected healthful effect (the negative sign implies reduced probability of mortality) in all specifications.

The first two rows of Table 3 reprint the results for white males in Table 2 of DL. The coefficient on the Gini in row [1] means that a ten-point increase in the Gini, e.g., from 0.34 to 0.44, would imply a $0.10 \times 0.46 = 0.046$, or 4.6%, increase in the annual probability of mortality. This result implies an important protective effect of equality on health (or dangerous effect of inequality on health), controlling for income.

However, the key point of DL is that the inequality–mortality relationship is not robust to the fault lines of race. When a state’s fraction black is introduced as an explanatory variable in row [2], the coefficient on the gini is small and insignificant (and actually negative in sign). The coefficient on income retains its expected negative sign. The coefficient on fraction black is positive, large, and significant. The finding is that the presence of blacks is associated with worsened health for white men and that the apparent ill effect of inequality works entirely through this racial channel.

In row [3], we directly replicate DL using the DL data. Except for a small difference in the $t$-statistic for the coefficient on fraction black, the replicated results are identical to the original results.

In rows [4] through [13], we report replications of row [2] with alternative weights and variables. In row [4], we directly correct the coding error in DL. We thus weight with the square root of the population, as specified in the DL text, rather than with the fourth root, which was mistakenly applied in the code provided by DL. Because we lack confidence in the basic specification of the model, we also estimate the regressions with several alternative weightings. Thus, we use unweighted data on metropolitan areas, data weighted by the fourth root of MSA population as computed by DL, data weighted by the square root of MSA population as computed by DL, data weighted by the square root of MSA population as reported in the Census, and data weighted by MSA population as reported in the Census. The results differ substantially for the coefficient on the key variable based on the weighting scheme applied.

In all specifications, the results for the direct effect of racial composition are similar to the DL results. Their coefficient of 0.50 and our lowest estimated coefficient of 0.36 imply that a 10%-point increase in the percent black increases the mortality rate by between 4% and 5%. The coefficient is highly statistically significant ($t$-statistic greater than 7) in all specifications. Although the variation in estimated coefficients is fairly small, the coefficient was systematically larger in magnitude, i.e., fraction black appears less healthy for white males, in the less weighted regressions.

The results for the protective effect of income on mortality are similarly stable across replications, with coefficients estimated between $-0.17$ and $-0.13$.

However, the results differ sharply for the effect of inequality on mortality based on the choice of weight. When the weight specified by DL, i.e., the square root of the estimated MSA population for white men, is applied in row [4], the coefficient on the gini is positive, of substantive size, and is statistically significant at the 10% level ($t = 1.64$). In row [6], which applies the square root of the MSA population, the coefficient on the gini is positive, of similar magnitude to and as statistically significant as in row [1]. When no weight is applied in row [5], the coefficient on the gini is large, negative, and close to statistical significance ($t = 1.92$). When the absolute-population weight is applied in row [7], the coefficient on
the gini is positive and almost 50% larger than in row [1], and the t-statistic exceeds that of the coefficient on income.

Although the coefficient on income is reasonably stable across specifications, the specification of the income variable affects the estimated effect of inequality. In row [12], which uses the corrected DL weight and log per capita income, the coefficient on inequality rises to 0.58 (t = 4.63).

The alternative gini variable with the original, incorrect weighting yields an insignificant point estimate of 0.10 (t = 0.51) for its coefficient (row [10]), although the swing from the DL point estimate of −0.09 (t = 0.6) is reasonably large. When the alternative gini is combined with the correct weighting (row [14]), the coefficient on the alternative gini is large, 0.54, and highly significant (t = 3.50).

Discussion

Although the alternative specifications of variables generate some difference among the results, the weighting scheme is clearly central. We take the extraordinary sensitivity of the gini–mortality results to the weighting system as an indication of fundamental misspecification or likely heterogeneity. Because the inequality–mortality relationship is strongest in the more weighted specifications and weakest in the less- or non-weighted specifications, the implication is that the relationship obtains between larger metropolitan areas that vary in the dimension of inequality but not between small metropolitan areas. In any case, the empirical results strongly suggest that we are not ready to close the book on the inequality–mortality hypothesis in its own right. The empirical evidence from the alternative weights suggests that the relationship between inequality and health may vary, for example, between large and small metropolitan areas. Sensitivity to population weighting, to the sample, and potentially to the unit of analysis suggests directions for further empirical research, but it is hard to consider the case closed.

Leaving the comparison of alternative weighting schemes, we believe that DL has left an important question unresolved. DL demonstrates that once percent black is included in the white-male log odds mortality-rate regression, the effect of inequality, either overall or among whites disappears. The result is taken as evidence against the inequality–mortality hypothesis. The authors do not further address the importance of percent black, either as itself a significant variable or as a confounder of inequality among the determinants of white-male mortality. In the discussion, DL conclude, “Although we do not know what causes this result, we note that the mechanisms that are emphasized in the literature on inequality and health, that work through the psychosocial environment, particularly stress and trust, are equally plausible as mechanisms through which race affects health” [emphasis in original] (DL, p. 1140).

We make two observations. First, the effect of percentage black on the health of white men is a striking ecological result. The gini for one’s own race and sex could, in principle, be a manifestation of the reflection problem (Manski, 1995), but the presence of people of another ethnicity is clearly a contextual matter. Death by human agency is rare; the increased black presence is not “causing” white deaths in any direct way. The presence of blacks in a metropolitan area reflects the presence of social determinants that are detrimental to population health, and it is these factors that warrant identification. (Black mortality as well as white mortality responds unfavorably to MSA percentage black.)

Second, American racial difference is inequality. Racial composition is inequality. The variable “fraction black” captures more than the ratio of blacks to whites in a given locality, it captures something about the dynamics of political and economic power in that locality. Commenting from the vantage of Jim Crow, Du Bois (1903) famously characterized the problem of the 20th century as the problem of the color line. A key feature of this was the sublimation or displacement of class division in the region, especially those areas dominated by plantation economies, and the corresponding conservative effect on politics and public policy (Du Bois, 1903, 1935; Key, 1949). Following the civil rights movement and the end of de jure apartheid in the South and Border States, the fact of Southern conservatism remains and is still driven by white Southern resistance to left-of-center social policy. Outside the South, areas with high percentages of blacks are places defined by pronounced racial segregation, and the predictable disadvantages that go with that: worse schools, worse neighborhoods, worse job prospects and so forth. These conditions are the direct and indirect result of white political attitudes.

As we noted in the Results section, fraction black and inequality among white males are correlated but not overwhelmingly so. One possibility is that inequality among white males has multiple facets, and that fraction black is correlated with the component of income inequality that is most detrimental to health.

Thus, far from interpreting the strength of the percent black variable as counter-evidence to the inequality–health hypothesis, we interpret the relationship between percent black and mortality as underlining the concept that pronounced racial difference is itself a form of inequality, and that inequality is unhealthy. Combining the ambiguous and specification-sensitive results on the relationship between income equality and mortality with the
strong relationship between racial composition and mortality, we conclude that unequal environments are likely unhealthful.

References


