12-2003

On the Importance of Age-Adjustment Methods in Ecological Studies of Social Determinants of Mortality

Jeffrey Milyo

Jennifer M. Mellor
William & Mary, jmmell@wm.edu

Recommended Citation
Milyo, Jeffrey and Mellor, Jennifer M., On the Importance of Age-Adjustment Methods in Ecological Studies of Social Determinants of Mortality (2003). *Health Services Research*, 38(6p2), 1781-1790. https://doi.org/10.1111/j.1475-6773.2003.00202.x

On the Importance of Age-Adjustment Methods in Ecological Studies of Social Determinants of Mortality

Jeffrey Milyo and Jennifer M. Mellor

Objective. To illustrate the potential sensitivity of ecological associations between mortality and certain socioeconomic factors to different methods of age-adjustment.
Data Sources. Secondary analysis employing state-level data from several publicly available sources. Crude and age-adjusted mortality rates for 1990 are obtained from the U.S. Centers for Disease Control. The Gini coefficient for family income and percent of persons below the federal poverty line are from the U.S. Bureau of Labor Statistics. Putnam's (2000) Social Capital Index was downloaded from http://www.bowlingalone.com; the Social Mistrust Index was calculated from responses to the General Social Survey, following the method described in Kawachi et al. (1997). All other covariates are obtained from the U.S. Census Bureau.
Study Design. We use least squares regression to estimate the effect of several state-level socioeconomic factors on mortality rates. We examine whether these statistical associations are sensitive to the use of alternative methods of accounting for the different age composition of state populations. Following several previous studies, we present results for the case when only mortality rates are age-adjusted. We contrast these results with those obtained from regressions of crude mortality on age variables.
Principal Findings. Different age-adjustment methods can cause a change in the sign or statistical significance of the association between mortality and various socio-economic factors. When age variables are included as regressors, we find no significant association between mortality and either income inequality, minority racial concentration, or social capital.
Conclusions. Ecological associations between certain socioeconomic factors and mortality may be extremely sensitive to different age-adjustment methods.

Key Words. Income inequality, social capital, race, age-adjustment, ecological bias

Numerous ecological studies report that mortality in the United States is significantly associated with area-level socioeconomic factors, including income inequality (e.g., Kaplan et al. 1996; Kennedy et al. 1996; Kawachi and Kennedy 1997b; Lynch et al. 1998; Muller 2002; and Ross et al. 2000), minority racial concentration (e.g., Cooper et al. 2001; Singh and Hoyert 2000; Fang et al. 1998; McLaughlin and Stokes 2002), and various measures of social capital (Kawachi and Kennedy 1997a; Kawachi et al. 1997; Kawachi and Berkman 2000; Putnam 2000). The results of these and similar studies are often based on multivariate regression analyses of mortality rates, in which the dependent variable is directly age-adjusted to account for the higher mortality rates of older populations, but the explanatory variables are not similarly adjusted. However, standardization of only the dependent variable in a regression (whether for age, race, time, or some other confounder or set of confounders) has long been understood to be problematic (e.g., Rosenbaum and Rubin 1984; Greenland and Morgenstern 1989, 1991; Greenland 1992). This is more than just an esoteric concern; in this study, we demonstrate that state-level associations between mortality rates and several socio-economic variables may be extremely sensitive to different age-adjustment methods.

BACKGROUND

There have been several cogent discussions of the advantages and disadvantages of ecological studies in epidemiology; recent examples include Greenland and Robins (1994) and Morgenstern (1995). While it is apparent that convenience and simplicity are often the motivations for ecological analyses, it is also the case that relevant exposures may be difficult to measure at the individual level, or that available data on individuals contain insufficient variation. These latter advantages are particularly apparent for the study of social determinants of mortality. Of course, the primary disadvantage is ecological bias; this bias may be reduced in either of two ways: (1) by including confounders as control variables in the regression analysis, or (2) by rate standardization for these confounders applied to all variables in the regression analysis. Rosenbaum and Rubin (1984) describe the sufficient conditions for which ecological bias is reduced by either of these two methods, although they also note that the latter method is often impractical, since standardized measures of many variables of interest are not readily available to researchers (e.g., social capital). In contrast, Rosenbaum and Rubin find few conditions for which standardization of only the dependent variable reduces such bias; in their words: "The point is: if we adjust mortality for age, we must adjust the covariates for age as well" (p. 438). This lesson is reiterated by

Address correspondence to Jeffrey Milyo, Ph.D., Assistant Professor, Harris Graduate School of Public Policy Studies, University of Chicago, Chicago, IL 60637. Jennifer M. Mellor, Ph.D., is an Assistant Professor, Department of Economics, The College of William and Mary, Williamsburg, VA 23187-8795.

Greenland and Robins (1994) and Morgenstern (1995), as well as in a recent epidemiological textbook by Rothman and Greenland (1998).

Nevertheless, Kawachi and Blakely (2001, 2002) and Subramanian et al. (2003) describe standardization of only the dependent variable in mortality studies as common practice among social epidemiologists. In contrast, in a recent ecological study of income inequality and mortality, Mellor and Milyo (2001) control for age (and other confounders) by simply including these as control variables in their regressions. However, this difference in methods is of little concern if either method yields substantively similar results. Therefore, we compare the estimated associations between state mortality rates and several different socioeconomic factors using either age-adjustment method. Consistent with the evidence reported in several previous studies, we find that standardization of only the dependent variable can yield strong associations between mortality and some measures of state socioeconomic factors. In contrast, when we include control variables for age in the regression analysis, we find no significant associations between these socioeconomic factors and mortality.

A Stylized Example

In this section we apply the analysis of Rosenbaum and Rubin (1984) to our case of interest: recent ecological studies of area-level socioeconomic determinants of mortality.

Assume that individual mortality is related to age category, poverty status, and some state level socioeconomic factor (X_s) in the following manner:

$$m_i = \beta_1 X_s + \beta_2 * poverty_i + \Sigma_k(\gamma_k age_{ki}) + \varepsilon_i. \quad (1)$$

Given this model, unbiased estimates of β_1 and β_2 may be obtained from a weighted least squares regression of state mortality rates:

$$M_s = \beta_1 X_s + \beta_2 * POVERTY_s + \Sigma_k(\gamma_k AGE_{ki}) + u_s \quad (2)$$

where the weight is state population, M_s is the crude state mortality rate, and $POVERTY_s$ and AGE_K measure the fraction of the state population in each category. Rosenbaum and Rubin (1984) also note that in place of the categorical age variables in model (2), sufficient moments of the age distribution may be substituted. For example, in their analysis of the association between inequality and mortality, Deaton and Lubotsky (2002) control for mean age in some specifications.

In contrast to model (2), researchers often estimate a model of the form:

$$M_s = \alpha_1 X_s + \alpha_2 * POVERTY_s + u_s, \quad (3)$$

where M_s is the *age-adjusted* state mortality rate. Again, Rosenbaum and Rubin (1984) note that this popular technique does not yield unbiased estimates of the parameters in model (1) unless: (i) state variables are constant within states and (ii) covariates representing aggregations of individual responses are not included in the model. If these conditions are not met, then Rosenbaum and Rubin caution that this technique should be avoided.

METHODS AND DATA

In our illustrative example, model (2) produces unbiased estimates of β_1 and β_2; now, we demonstrate that regression estimates of the parameters of interest in model (3) may diverge substantially from those obtained by estimating model (2).

We use least squares regressions to analyze the ecological association between mortality rates in U.S. states and each of three socioeconomic factors: inequality, minority racial concentration, and social capital. We compare the results obtained from estimating the models described in (2) and (3), above. These specifications are broadly consistent with at least some model specifications employed in several earlier studies, but we have not sought to replicate all of the myriad specifications employed in the previous literature. As such, our empirical exercise is intended only as an illustration of the potential importance of different methods of age-adjustment in ecological regressions.

We compare the estimates obtained from each model for the cases where X is either income inequality, minority racial concentration, or one of two measures of social capital. All regressions include a constant term, so estimates of model (2) omit one age category.[1] Descriptive statistics for all variables are listed Table A1 in the appendix.

Data on the age distribution within states are obtained from the 1990 U.S. Census. Mortality rates are defined as deaths per 100,000 in 1990 (source: U.S. Centers for Disease Control). Poverty is defined as the percent of individuals in a state with incomes below the federal poverty level in 1990, and income inequality is measured by the Gini coefficient for household income in 1990 (source: U.S. Census Bureau). Minority racial concentration is measured by the percent black in a state in 1990 (source: U.S. Census Bureau). Social capital is measured by Robert Putnam's index of state social capital (source: http://www.bowlingalone.com) and by an index of social mistrust derived from responses to the General Social Survey (source: National Opinion

Research Center) and based on the method described in Kawachi et al. (1997).[2] However, the social capital variables are available for only 48 and 39 states, respectively.

RESULTS

In Table 1, we report the results of our regression analyses when X is income inequality. In the first two columns we present weighted and unweighted estimates of model (3). Because this model is ad hoc, it is not clear whether it is appropriate to weight by state population or some other variable. Without weighting, inequality is significantly and positively related to state age-adjusted mortality (as reported in many previous studies); however, weighting by state population leads to a change in both the sign and significance of this estimate. In all subsequent regressions we use weighted least squares to facilitate the comparison of estimates obtained under different age-adjustment methods.

Table 1: Ecological Associations between State Mortality Rates and Inequality: Sensitivity of Regression Estimates to Age-Adjustment Methods

	Age-Adjusted Mortality Rates		*Crude Mortality Rates*	
	(1) OLS	*(2) WLS*	*(3) WLS*	*(4) WLS*
Income inequality	1228.4*	– 45.8	231.6	– 500.1
	(2.20)	(0.10)	(0.59)	(1.15)
Poverty	3.6	9.3**	3.5	11.3**
	(1.19)	(3.18)	(0.60)	(3.96)
Mean age				72.7**
				(14.85)
Ages 0–18			8.4	
			(1.55)	
Ages 19–24			28.2**	
			(2.80)	
Ages 45–64			51.7**	
			(5.88)	
Ages 65+			35.9**	
			(8.33)	
Adjusted R^2	.30	.26	.88	.81

Notes: **($p<.01$) and *($p<.05$); absolute values of T-statistics in parentheses. All models include a constant (ages 25–44 is the omitted age category). Number of observations is 50 states for all regressions. OLS refers to ordinary least squares estimation; WLS refers to weighted least squares (weight = state population).

In the last two columns of Table 1, we report our estimates of model (2). In one case we control for age by including age variables as controls, in the other we include only mean age (following Deaton and Lubotsky 2002). While the coefficient estimates on inequality vary under either method (from +231 to − 500), the standard errors are so large that we cannot reject the hypothesis that the coefficients on inequality are the same in either specification. In fact, in this example, the estimated association between inequality and state mortality is significant and positive only for the unweighted regression of the ad hoc model (3). However, the same is not true for the estimated coefficients on poverty.[3]

Controlling for several age categories (see column three of Table 1) versus mean age (see column four) changes the coefficient on poverty significantly. Unfortunately, we can not distinguish which is the better estimate of β_2, since our specification of model (1) did not dictate the appropriate number and type of age categories. If we instead include only a control for only age 65+ (not shown), then the estimated coefficients on inequality and poverty are much closer to those found using the full set of age controls (column three), but if we add age 0–18 as a control (not shown), then the results are more in line with those found by using mean age (column four).

In Table 2, we compare weighted least squares estimates of models (2) and (3), where X is either minority racial concentration, social capital, or social mistrust. In the first three columns, we present estimates of model (3), where the dependent variable is age-adjusted mortality. In each case, the state socioeconomic factor of interest is significantly associated with mortality rates. Again, these findings are broadly consistent with what has been reported in previous ecological studies. However, this pattern is not repeated in the last three columns of Table 2; estimates of model (2) do not yield a significant association between either state socioeconomic factor and mortality. Further, we reject the null hypotheses that the estimated coefficients reported in the last three columns are identical to the corresponding estimates in the first three columns. These results are in stark contrast to previous findings. Finally, the estimated coefficients on poverty in these specifications appear less sensitive to these age-adjustment methods; we cannot reject the null hypotheses that the estimated coefficients for poverty in columns (1–3) are identical to the corresponding estimate in (4–6).

DISCUSSION

We demonstrate that statistical associations between at least some focal area-level socioeconomic factors and mortality rates are sensitive to different

Table 2: Ecological Associations between State Mortality Rates and Other Social Factors: Sensitivity of Weighted Regression Estimates to Age-Adjustment Methods

	Age-Adjusted Mortality Rate			*Crude Mortality Rate*		
	(1)	(2)	(3)	(4)	(5)	(6)
Minority racial concentration	4.9**			1.0		
	(6.46)			(1.03)		
Social capital		– 69.5**			6.6	
		(5.93)			(0.57)	
Mistrust			2.8**			– 0.5
			(2.82)			(0.72)
Poverty	3.3	2.2	6.1*	3.6	5.4**	6.0**
	(1.89)	(1.16)	(2.53)	(1.68)	(2.93)	(3.58)
Ages 0–18				7.1	7.0	8.0**
				(1.48)	(1.88)	(2.12)
Ages 19–24				25.2*	35.0**	29.1**
				(2.44)	(4.23)	(3.56)
Ages 45–64				46.5**	52.6**	55.5**
				(5.06)	(7.47)	(8.04)
Ages 65+				36.6**	37.3**	35.6**
				(8.41)	(10.86)	(10.27)
Adjusted R^2	.60	.58	.44	.89	.93	.94
Observations	50	48	39	50	48	39

Notes: **($p < .01$) and *($p < .05$); absolute values of T-statistics in parentheses. All models include a constant term (ages 25–44 is the omitted age category). All estimates are obtained from weighted least squares regressions (weight = state population).

methods of age-adjustment. However, our purpose here is only to remind producers and consumers of ecological studies about the pitfalls of standardizing only the dependent variable in multivariate regressions. We do not take these results as evidence that there is no link between state-level socioeconomic factors and health outcomes; however, this analysis does suggest that previous conclusions based on findings of strong ecological associations between these particular state-level socioeconomic factors and state mortality rates may need to be revisited. The potential sensitivity of results to different age-adjustment methods, and even the number of age categories used as controls, is one more reason to be cautious about interpreting ecological studies. As such, our findings offer additional support for the use of contextual or multilevel analyses as a valid approach for investigating area-level effects on individual health (e.g., Mellor and Milyo 2002, 2003).

NOTES

1. The age variables are defined as the percent of state population that is younger than 19 years old, between 19 and 24 years old, between 45 and 64, and 65 or older; the omitted category is 25–44 years old.
2. Social mistrust is one of four similar social capital measures developed in Kawachi et al. (1997); these measures have been used in a number of subsequent studies.
3. Rosenbaum and Rubin (1984) also note that under certain restrictive conditions, weighted least squares estimation of model (2) with age-adjusted mortality substituted as the dependent variable will yield unbiased estimates of β_1 and β_2. This exercise results in a coefficient on inequality of 61.1 ($t = 0.20$) and on poverty of 7.5 ($t = 3.50$).

APPENDIX

Table A1: Means and Standard Deviations for All Variables

Variable	*50 States*	*48 States*	*39 States*
Mortality per 100,000	853.4 (112.2)	861.3 (106.0)	867.9 (110.3)
Age-adjusted mortality per 100,000 (base 1990)	852.1 (69.3)	855.0 (66.7)	860.7 (67.6)
Poverty (% of population)	13.1 (4.2)	13.2 (4.2)	13.3 (4.3)
Income inequality (Gini for household income)	.398 (.022)		
Minority racial concentration (% black)	9.5 (9.3)		
Social Capital Index (Putnam 2000)		.020 (.781)	
Social mistrust			65.8 (10.6)
Ages 0–18 (% of population)	26.2 (2.5)	26.1 (2.4)	25.9 (2.5)
Ages 19–24	10.6 (0.7)	10.6 (0.8)	10.7 (0.7)
Ages 45–64	18.5 (1.2)	18.6 (1.1)	18.6 (1.1)
Ages 65+	12.5 (2.1)	12.7 (1.8)	12.7 (1.9)

REFERENCES

Cooper, R. S., J. F. Fennelly, R. Durazo-Arvizo, H. Oh, G. Kaplan, and J. Lynch. 2001. "Relationship between Premature Mortality and Socioeconomic Factors in

Black and White Populations of U.S. Metropolitan Areas." *Public Health Reports* 116 (5): 464–73.

Deaton, A., and D. Lubotsky. 2002. "Mortality, Inequality, and Race in American Cities and States." *Social Science and Medicine* 56 (6): 1139–53.

Fang, J., S. Madhavan, W. Bosworth, and M. H. Alderman. 1998. "Residential Segregation and Mortality in New York City." *Social Science and Medicine* 47 (4): 469–76.

Greenland, S. 1992. "Divergent Biases in Ecological and Individual-Level Studies." *Statistics in Medicine* 11 (9): 1209–23.

Greenland, S., and H. Morgenstern. 1989. "Ecological Bias, Confounding and Effect Modification." *International Journal of Epidemiology* 18 (1): 269–74.

———. 1991. "Neither within-region Nor Cross-Regional Independence of Exposure and Covariates Prevents Ecological Bias." *International Journal of Epidemiology* 20 (3): 816–7.

Greenland, S., and J. Robins. 1994. "Ecological Studies: Biases, Misconceptions and Counterexamples." *American Journal of Epidemiology* 139 (8): 747–60.

Kaplan, G. A., E. R. Pamuk, J. M. Lynch, R. D. Cohen, and J. L. Balfour. 1996. "Inequality in Income and Mortality in the United States: Analysis of Mortality and Potential Pathways." *British Medical Journal* 312 (7037): 999–1003.

Kawachi, I., and L Berkman. 2000. "Social Cohesion, Social Capital and Health." In *Social Epidemiology*, edited by L. Berkman, and I. Kawachi, pp. 174–90. Oxford: Oxford University Press.

Kawachi, I., and T. A. Blakely. 2001. "When Economists and Epidemiologists Disagree." *Journal of Health Politics, Policy and Law* 26 (3): 533–41.

——— 2002. "Response." *Journal of Health Politics, Policy and Law* 27 (2): 299–302.

Kawachi, I., and B. P. Kennedy. 1997a. "Health and Social Cohesion: Why Care about Income Inequality?" *British Medical Journal* 314 (7086): 1037–40.

——— 1997b. "The Relationship of Income Inequality to Mortality: Does the Choice of Indicator Matter?" *Social Science and Medicine* 45 (7): 1121–7.

Kawachi, I., B. P. Kennedy, K. Lochner, and D. Prothrow-Stith. 1997. "Social Capital, Income Inequality, and Mortality." *American Journal of Public Health* 87 (9): 1491–8.

Kennedy, B. P., I. Kawachi, and D. Prothrow-Stith. 1996. "Income Distribution and Mortality: Cross-Sectional Ecological Study of the Robin-Hood Index in the United States." *British Medical Journal* 312 (7037): 1004–7.

Lynch, J. W., G. A. Kaplan, E. R. Pamuk, R. D. Cohen, K. E. Heck, J. L. Balfour, and I. H. Yen. 1998. "Income Inequality and Mortality in Metropolitan Areas of the United States." *American Journal of Public Health* 88 (7): 1074–80.

McLaughlin, D. K., and C. S. Stokes. 2002. "Income Inequality and Mortality in U.S. Counties: Does Minority Racial Concentration Matter?" *American Journal of Public Health* 92 (1): 99–104.

Mellor, J. M., and J. Milyo. 2001. "Re-Examining the Evidence of an Ecological Association between Income Inequality and Health." *Journal of Health Politics, Policy and Law* 26 (3): 487–522.

——— 2002. "Income Inequality and Individual Health: Evidence from the Current Population Survey." *Journal of Human Resources* 37 (3): 510–39.

——— 2003. "Is Exposure to Income Inequality a Public Health Concern? Lagged Effects of Income Inequality on Individual and Population Health." *Health Services Research* 38 (1, part 1): 137–51.

Morgenstern, H. 1995. "Ecological Studies in Epidemiology: Concepts, Principles, and Methods." *Annual Review of Public Health* 16: 61–81.

Muller, A. 2002. "Education, Income Inequality and Mortality: A Multiple Regression Analysis." *British Medical Journal* 324 (7328): 23–5.

National Opinion Research Center, http://www.norc.org, [last accessed on August 15, 2002].

Putnam, R. D. 2000. *Bowling Alone: The Collapse and Revival of American Community.* Available at http://www.bowlingalone.com, [last accessed on August 15, 2002]. New York: Simon and Schuster.

Rosenbaum, P. R., and D. B. Rubin. 1984. "Difficulties with Regression Analyses of Age-Adjusted Rates." *Biometrics* 40 (2): 437–43.

Ross, N. A., M. C. Wolfson, J. R. Dunn, J. M. Berthelot, G. A. Kaplan, and J. W. Lynch. 2000. "Relationship between Income Inequality and Mortality in Canada and in the United States: Cross-Sectional Assessment Using Census Data and Vital Statistics." *British Medical Journal* 320 (7239): 898–902.

Rothman, K. J., and S. Greenland. 1998. *Modern Epidemiology.* Philadelphia: Lippincott-Raven.

Singh, G. K., and D. L. Hoyert. 2000. "Social Epidemiology of Chronic Liver Disease and Cirrhosis Mortality in the United States, 1935–1997: Trends and Differentials by Ethnicity, Socioeconomic Status, and Alcohol Consumption." *Human Biology* 72 (5): 801–20.

Subramanian, S. V., T. Blakely, and I. Kawachi. 2003. "Income Inequality as a Public Health Concern: Where Do We Stand?" *Health Services Research* 38 (1): 153–67.

United States Census Bureau, http://www.census.gov, [last accessed on August 15, 2002].

United States Centers for Disease Control, http://wonder.cdc.gov, [last accessed on August 15, 2002].