In: Proceedings of the 23rd national meeting of the Public Health Conference on Records and Statistics, July 1991. DHHS Publication No. (PHS) 92-1214. Washington, DC: National Center for Health Statistics, 1992.

BEHAVIORAL, PSYCHOSOCIAL, DEMOGRAPHIC, AND HEALTH FACTORS ASSOCIATED WITH RESIDENCE IN HIGH SMR ZIP CODES

G.A. Kaplan, California Department of Health Services
R.D. Cohen, P. Cirillo, W.B. Davis, J. Bartlett, M. Johnston, W.E. Wright, J. Mendlein

Introduction

Observations of the unequal distribution of disease by geographic location are not new, providing one of the foundations of modern public health and epidemiology. Both before and after John Snow's mid-nineteenth century investigation of the causes of cholera in London, investigators have searched for the equivalent of the "Broad Street Pump" with which to explain non-random occurrences of the location of disease. Interest in this subject remains strong today as evidenced by a recent national conference on clustering of health events devoted to the public demands, methodologic pitfalls, and statistical complications involved in the study of the clustering of health events (National Conference on Clustering of Health Events, 1990).

Since 1965, the Alameda County Study of the California Department of Health Services has been following the health of almost 7,000 persons, investigating the association between behavioral, social, psychological, and demographic factors and health outcomes (Berkman and Breslow, 1983). In the process of conducting this work, we have repeatedly noticed substantial geographic variation in rates of all-cause mortality and other causes of death. For example, in one study the 9-year mortality experience of residents of Oakland, California, the largest city in Alameda County, was examined as a function of whether or not they lived in federally-designated poverty areas in Oakland. Residence in the poverty area was associated with more than a 70% increase in the 9-year risk of death even when there was adjustment for a large number of important covariates (Haan, Kaplan & Camacho, 1987). Similarly, in another study, a number of factor analytically-derived scales based on Alameda County census tract characteristics were found to be independently associated with risk of death, and census tracts which were high on these scales were clustered together (Haan, Kaplan & Syme, 1989).

The present study investigates, in greater detail, factors associated with these high risk areas. The approach is somewhat different from that taken in our previous cohort-based studies. In this study we calculated the standardized mortality ratios (SMR) for all-cause mortality for Alameda County zip codes, conducted a large telephone survey which collected of a wide variety of information including zip code of the respondents, and then brought these two sources of data together. Our analyses were directed at examining the factors associated with residence in high, medium, or low SMR zip codes. Thus, we are not examining factors associated with health status but, instead, with residence in areas which differ in health status as reflected by all-cause SMRs. While this is clearly an ecologic outcome which does not reflect the health status of the respondents, we are interested in understanding variations in the health of areas. Importantly, an informal review suggested that high SMR areas in Alameda County have remained high over relatively long periods of time, regardless of the births, deaths, and relocations which characterize any geographic area over time.

Methods

Estimation of SMRs

Standardized mortality ratios (SMRs) for deaths from all-causes for both sexes combined were calculated for 43 sub-areas of Alameda county defined by postal zip code. The SMR for a zip code was calculated by dividing the observed number of deaths by the number expected if age specific mortality rates for the entire county applied to the population of the zip code.

Records of deaths by census tract for 1984-1988 were obtained from the Alameda County Department of Health. The census tract totals were then allocated to zip codes based on the work of Gould et.al., (1985). These investigators combined information on 1980 census tract boundaries, 1985 proprietary zip code boundary files, and other sources, coupled with manual review to allocate the population of a census tract to one or more zip codes.

To calculate expected deaths for each zip code, population estimates by age and sex for each area for 1984-1988 were required. To simplify the calculation, a mid-period (1986) age and sex-specific estimate was derived for each zip code using a linear interpolation of 1980 and 1989 data obtained from the National Planning Data Corporation. These 1986 estimates were then adjusted so that when summed over zip codes they equaled age and sex-specific estimates for the entire county produced by the California Department of Finance (1990). The estimated age-sex populations in each zip code were then multiplied by age-sex- specific mortality rates for the entire county for 1984-1988 to obtain the expected numbers of deaths in each zip code.

Survey Data

Between April 1988, and June 1990, data were collected from 3,047 adult residents of Alameda County using a Waksberg random digit dialing technique with one adult resident of each household being randomly selected. Data from additional Black and Hispanic residents were obtained in the same way using a stratified sampling technique which sampled more heavily from telephone exchanges which contained higher proportions of Blacks and Hispanics. Analysis weights were calculated by a two step process which involved the calculation of a crude weight for each person in the sample approximately proportional to the reciprocal of that individual's selection probability and post stratification adjustment of the set of crude weights to match the distribution of the county population by age, race and sex.

Information was obtained on a wide variety of variables including: self-reported health conditions and disabilities, behavioral factors, social networks and support, race/ethnicity, socioeconomic position and education, preventive health services use, health care coverage,

socioenvironmental demands and control, and disadvantage.

Statistical Methods

A series of age- and sex-adjusted polytomous logistic models were evaluated in which three outcomes were used: high SMR, medium SMR, and low SMR zip code. The high and low zip codes corresponded to the approximate upper and lower quintiles of the distribution of zip code SMRs. SAS PROC CATMOD was used to fit the multinomial logistic model to all observations in the sample using the WEIGHT option to incorporate the analysis weights (SAS Institute Inc.). The regression coefficients and odds ratios from this fit are those reported. To estimate variances and covariances for the coefficients that reflect the complex design of the sample, a half-sample replication method, suggested by B.V. Shah (personal communication) was used. This method was employed because available software either did not calculate variances for complex survey designs (SAS) or did not accommodate the multinomial logistic model (SUDAAN, Research Triangle Institute).

The half-sample replication method utilizes a series of 50 weighted logistic analyses carried out on 50 simple random samples, without replacement, of one-half the PSUs in the sample. The empirical variances and covariances of the coefficients from the 50 replications, calculated using SAS PROC CORR, provided estimates of the variances of the coefficients obtained from the model fit to the full sample. To test the significance of the logistic regression coefficients, the ratio of the coefficient to its standard error was treated as a Student's t statistic with 49 degrees of freedom (1 less than the number of replications).

Results

Standardized Mortality Ratios

There was considerable variation between the 43 zip code areas, with SMR's ranging from 0.54 to 1.42 (Figure 1). Zip codes were assigned to three categories in order to examine factors associated with residence in high SMR (SMR= 1.13-1.42, N=7), Medium SMR (SMR= 0.75-1.12, N=28), and low SMR areas (SMR= 0.54-0.74, N=8).

Polytomous Logistic Analyses

There was a strong association between health status measures and residence in high SMR zip codes. A one unit increase in an index which measures the number of chronic diseases reported (1-11) was associated with increased odds of residence in high vs. low SMR zip codes (OR= 1.24, 95% C.I.=1.03-1.49) and medium vs low SMR zip codes (OR=1.11, 95% C.I.= 0.95-1.20). The number of activity of daily living (ADL) and instrumental activity of daily living (IADL) problems (1-12) was also associated with area of residence ($OR_{HIGH\ vs.\ LOW}$=1.26, 95% C.I.=1.10-1.45; $OR_{MEDIUM\ vs.\ LOW}$=1.17, 95% C.I.=1.03-1.33), as was reporting fair/poor vs. good/excellent perceived health ($OR_{HIGH\ vs.\ LOW}$=3.00, 95% C.I.=1.78-5.04; $OR_{MEDIUM\ vs.\ LOW}$=1.52, 95% C.I.=0.96-2.42). Higher prevalence of high blood pressure, vision trouble, hearing trouble, back pain, muscle pain, and headache was found in the high vs. low SMR areas as well. Table 1 indicates the strong associations between Race/Ethnicity and socioeconomic variables and area of residence. Blacks are more than nine times more likely than whites to live in a high SMR vs. low SMR area, and Hispanics compared have over three times the odds of whites. Family income and education both show a graded association with residence in high vs. low SMR area. Residence in high SMR zip codes was also associated with respondent's reports of the frequency of having inadequate money for food/month and inadequate money to fill a prescription/year.

Levels of smoking, alcohol consumption, body mass index (BMI)[weight (lbs.)/height^2(inches)], and exercise were also associated with SMR area (Table 2). High SMR areas had increased prevalence of current smoking ($OR_{HIGH\ vs.\ LOW}$=1.58, 95% C.I.=0.98-2.55; $OR_{MEDIUM\ vs.\ LOW}$=1.42, 95% C.I.=0.94-2.14), and past smoking ($OR_{HIGH\ vs.\ LOW}$=1.48, 95% C.I.=0.92-2.36; $OR_{MEDIUM\ vs.\ LOW}$=1.55, 95% C.I.=0.98-2.43). Respondents who live in high SMR areas had increased frequency of reporting no exercise in the last month ($OR_{HIGH\ vs.\ LOW}$=2.34, 95% C.I.=1.69-2.34; $OR_{MEDIUM\ vs.\ LOW}$=1.29, 95% C.I.=0.94-1.76), and were more likely to be in the 5th quintile of BMI ($OR_{HIGH\ vs.\ LOW}$=3.07, 95% C.I.=1.54-6.11; $OR_{MEDIUM\ vs.\ LOW}$=1.81, 95% C.I.=0.95-3.44). Residents of high areas were also more likely to report abstaining from alcohol consumption ($OR_{HIGH\ vs.\ LOW}$=1.69, 95% C.I.=1.06-2.68).

Substantial differences between areas were also seen with respect to medical and preventive care access (Table 3). Lack of health insurance (including Medicare and Medicaid) was elevated in high, but not medium, areas (Table 3) ($OR_{HIGH\ vs.\ LOW}$=2.24, 95% C.I.=1.16-4.33). Not having a check-up within the last 2 years was reported more frequently by those in high SMR areas ($OR_{HIGH\ vs.\ LOW}$=1.41, 95% C.I.=0.90-2.20), as was not having a Pap test (for women) during the last 12 months ($OR_{HIGH\ vs.\ LOW}$=1.83, 95% C.I.=0.91-3.68; $OR_{MEDIUM\ vs.\ LOW}$=1.60, 95% C.I.=0.89-2.85). There was some indication that having "ever" had a cholesterol check was more prevalent in high than in low SMR areas.

Different levels of social network participation and social support were also reported in the high, medium, and low SMR areas (Table 4). Those who reported fewer than three close friends and relatives were three times more likely to live in high vs low areas (95% C.I.=1.38-6.53). Low emotional support was also more prevalent in high risk areas. Those who reported no persons available at least some of the time to listen or to confide in about emotional problems were more likeley to live in high or medium vs. low SMR areas ($OR_{HIGH\ vs.\ LOW}$=4.84, 95% C.I.=1.75-13.34; $OR_{MEDIUM\ vs.\ LOW}$=2.42, 95% C.I.=0.98-5.96).

Table 5 presents the results for socioenvironmental factors. A summative demands index was generated based on three items in which respondents agreed or strongly agreed that their work or usual daily activity, if not working, involved working "hard," "fast," or "repetitively." Those who reported 2 or more demands were 1.75 times more likely to live in a high or medium SMR zip code than those who reported no demands ($OR_{HIGH\ vs.\ LOW}$=1.75, 95% C.I.=1.13-2.70; $OR_{MEDIUM\ vs.\ LOW}$=1.74, 95% C.I.=1.09-2.77). A resources index was based on five questions which asked if respondents agreed or strongly agreed that in their work or usual daily activity, if not working, they were free

to make a lot of decisions, could decide how to do things, were able to be creative, were able to learn new things, and were appreciated. Those who reported fewer than four resources tended to be more likely to reside in high vs. low SMR zip codes ($OR_{HIGH\ vs.\ LOW}$=1.40, 95% C.I.=0.91-2.13). A socioenvironmental strain variable was also associated with zip code residence. The odds of living in high vs. low SMR zip codes was examined in those who reported high demands and low resources vs. those who reported low demands vs. high resources ($OR_{HIGH\ vs.\ LOW}$=1.99, 95% C.I.=1.08-3.67). Variables which reflect safety were also associated with residence in high vs. low SMR zip codes. Respondents who reported that their neighborhood was unsafe or very unsafe from crime were almost three times more likely to live in high vs. low SMR zip codes ($OR_{HIGH\ vs.\ LOW}$=2.72, 95% C.I.=1.52-4.86). Similarly, there was a tendency for reports of having been a victim of a violent crime to be more prevalent in the high SMR areas (($OR_{HIGH\ vs.\ LOW}$=2.24, 95% C.I.=0.85-5.87).

Conclusions

There is considerable variation in all-cause SMR's between zip codes in Alameda County, and a wide variety of sociodemographic, behavioral, social, medical care, and socioenvironmental differences are associated with the variations. It is appropriate to point out a number of methodological and conceptual issues which need to be considered in interpreting these results. The design of this study might be considered a "mixed" ecological analysis, with risk factors measured at the individual level and health outcomes measured at the ecologic level. While this design benefits over the usual ecological analysis in that it collects information from individuals, thereby allowing an examination of confounding, effect modification, and other aspects which cannot be addressed in data which includes risks factors and outcomes which are both measured ecologically, it does have certain limitations. The biggest limitation is that, properly speaking, the analyses do not model a health outcome. Instead they model residence in areas which differ in a summary measure of health, all-cause mortality. While this creates interpretive problems, other observations and analyses (not shown) are somewhat reassuring in that they indicate that the pattern of elevated SMR's has remained relatively constant over several decades, and that the risk factors measured in this study are, in this data set, related cross-sectionally to measures of health status.

The methodologies used in converting from census tracts to zip codes and estimating denominators may be subject to error, but it is not possible to determine, at this time, how large or in what direction. Similarly, the proper analysis of a complex survey design requires population information which is not always available, and we had to rely on proprietary information, the validity of which is unknown. With the continuing release of information from the 1990 census, it should be possible to address many of these data quality issues. Finally, the collection of self-reported information over the telephone must, by necessity, exclude those without telephones, is undoubtedly biased against the inclusion of the very poor or the socially marginal, and suffers from the limitations of self-reported data.

While these are not inconsequential conceptual and methodologic problems, they should not detract from the major conclusion, which is that there are substantial small area variations in health status in Alameda County, and that an extremely wide variety of risk factors seem to be associated with these geographic variations. The implication of this pattern of results is that broad-based interventions involving behavioral, social, socio-environmental, medical, and economic foci may be required in order to reduce geographic inequalities in health.

References

Berkman LF, Breslow L. Health and Ways of Living: The Alameda County Study. New York: Oxford University Press, 1983.

Gould JB, Liebman E, White A, et.al. 1980 Census Tract Composition of California Zip Codes Guide. Improved Perinatal Outcome Data Management Project, Maternal and Child Health Program, School of Public Health, University of California, Berkeley, 1985.

Haan M, Kaplan GA, Camacho T. Poverty and Health: Prospective Evidence from the Alameda County Study. Am J Epidemiol 1987;125(6):989-98.

Haan M, Kaplan GA, Syme SL. Socioeconomic Status and Health: Old Observations and New Thoughts. Background paper prepared for the Henry J. Kaiser Family Foundation Conference on Socioeconomic Status and Health, Menlo Park, CA, 3/87. In: Bunker JP, Gomby DS, Kehrer BH, eds. Pathways to Health: The Role of Social Factors. Menlo Park, CA: Henry J. Kaiser Family Foundation, 1989:76-135.

National Conference on Clustering of Health Events. American Journal of Epidemiology 132 (Supplement), 1990.

Research Traingle Institute, SUDAAN Software for Survey Data Analysis, version 5.40. Research Triangle Park, NC: Research Triangle Institute, 1991.

SAS Institute Inc., SAS/STAT USer's Guide, Release 6.03 Edition. Cary, NC:SAS Institute Inc., 1988.

State of California Department of Finance, Demographic Research Unit, 1990. Unpublished Data (Estimates Consistent County Projections, 1970-1989).

Table 1
Association Between Race/Ethnicity and Socioeconomic Variables and Residence in High/Medium/Low SMR Zip Codes Age and Sex-adjusted: Alameda County California, 1984-1988

Variable	High vs. Low		Medium vs. Low	
	OR	95% CI	OR	95% CI
Race				
Black	9.26	5.10-16.81	1.26	0.77-2.05
Hispanic	3.42	1.95-6.00	1.04	0.69-1.57
White (ref)	1.00		1.00	
Income Quintile				
1	2.33	1.17-4.63	0.70	0.41-1.20
2	2.13	1.11-4.07	0.94	0.59-1.52
3	1.35	0.66-2.75	0.97	0.58-1.61
4	0.86	0.42-1.80	1.12	0.68-1.83
5 (ref)	1.00		1.00	
Education				
< High School	4.29	2.41-7.64	1.58	0.95-2.63
High School/GED	2.81	1.69-4.67	1.76	1.11-2.80
> High School (ref)	1.00		1.00	
Inadequate $ for Food				
$\geq$ 1/month	1.68	0.87-3.26	0.92	0.51-1.65
0/month (ref)	1.00		1.00	
Inadequate $ for R_x				
$\geq$ 1/year	4.03	1.63-9.95	1.55	0.64-3.77
0/year (ref)	1.00		1.00	

Table 2
Association Between Behavioral Variables and Residence in High/Medium/Low SMR Zip Codes Age and Sex-adjusted: Alameda County California, 1984-1988

Variable	High vs. Low		Medium vs. Low	
	OR	95% CI	OR	95% CI
Smoking Status				
Current smoker	1.58	0.98-2.55	1.42	0.94-2.14
Past smoker	1.48	0.92-2.36	1.55	0.98-2.43
Never smoked (ref)	1.00		1.00	
Alcohol Consumption				
Abstainers	1.69	1.06-2.69	0.86	0.58-1.27
>45 drinks/month	1.06	0.59-1.88	1.17	0.68-1.99
1-45 drinks/month (ref)	1.00		1.00	
Body Mass Quintile				
1	0.97	0.50-1.89	1.33	0.73-2.40
2	1.25	0.67-2.34	1.10	0.64-1.88
3 (ref)	1.00		1.00	
4	1.74	0.83-3.65	1.40	0.71-2.77
5	3.07	1.54-6.11	1.81	0.95-3.44
Exercise				
None	2.34	1.69-3.24	1.29	0.94-1.76
$\geq$1/month (ref)	1.00		1.00	

Table 3
Association Between Preventative Care Variables
and Residence in High/Medium/Low SMR Zip Codes
Age and Sex-adjusted: Alameda County California, 1984-1988

Variable	High vs. Low		Medium vs. Low	
	OR	95% CI	OR	95% CI
High Blood Pressure				
Yes	3.04	1.65-5.62	2.18	1.30-3.65
No (ref)	1.00		1.00	
Cholesterol Check				
Yes	1.33	0.84-2.11	1.31	0.87-1.98
No (ref)	1.00		1.00	
Last Pap Smear				
>1 year	0.55	0.27-1.10	0.63	0.35-1.12
≤1 year (ref)	1.00		1.00	
Last General Checkup				
>2 years	0.71	0.45-1.11	0.90	0.60-1.33
≤2 years (ref)	1.00		1.00	

Table 4
Association Between Social Variables
and Residence in High/Medium/Low SMR Zip Codes
Age and Sex-adjusted: Alameda County California, 1984-1988

Variable	High vs. Low		Medium vs. Low	
	OR	95% CI	OR	95% CI
Friends Index				
≤2	3.00	1.38-6.53	1.54	0.75-3.18
≥3 (ref)	1.00		1.00	
Emotional Support Index				
0	4.84	1.75-13.35	2.42	0.98-5.96
1	1.89	0.68-5.24	1.30	0.54-3.16
2 (ref)	1.00		1.00	
Tangible Support Index				
0≤index<2	2.59	0.66-10.13	1.10	0.29-4.22
2≤index<4	1.27	0.77-2.10	0.86	0.53-1.42
4 (ref)	1.00		1.00	

Table 5
Association Between Socio-environmental Variables
and Residence in High/Medium/Low SMR Zip Codes
Age and Sex-adjusted: Alameda County California, 1984-1988

Variable	High vs. Low		Medium vs. Low	
	OR	95% CI	OR	95% CI
Neighborhood				
Unsafe/very unsafe	2.72	1.52-4.86	0.65	0.41-1.05
Safe/very safe (ref)	1.00		1.00	
Victim of Crime				
Yes	2.24	0.85-5.87	1.17	0.47-2.91
No (ref)	1.00		1.00	
Demands Index				
>2	1.75	1.13-2.70	1.74	1.09-2.77
1<index≤2	1.16	0.72-1.85	1.25	0.81-1.93
<=1 (ref)	1.00		1.00	
Resources Index				
≤4	1.40	0.92-2.13	1.10	0.78-1.55
>4 (ref)	1.00		1.00	
Strain Index				
High/Low	1.99	1.08-3.67	1.51	0.89-2.57
High/High	2.14	1.13-4.04	1.80	1.09-2.97
Low/Low	1.64	1.02-2.63	1.17	0.77-1.80
Low/High (ref)	1.00		1.00	

Figure 1

SMRs for All Cause Mortality by Zip code
Alameda County, California: 1984-1988