Social area analysis in community medicine

ALEX SCOTT-SAMUEL

From Mersey Regional Health Authority, Liverpool

SUMMARY   There is an acknowledged need for better information to guide resource allocation and service planning in the health services. Despite the recognition of the important role of socio-economic factors, difficulties with the appropriate presentation of data have so far proved insuperable. Social area analysis (SAA), which is a generic name for a number of methods employing census and other data to classify small areas into similar socioeconomic groups, is an approach which quantifies data in a useful fashion and has important applications in medical, epidemiological, and health services research. Most previous British exercises in SAA have been in the field of town planning. The potentialities of the approach for community medicine are evaluated, by the use of information from two existing studies. This is shown to discriminate more effectively than does existing health information between hypotheses concerning geographical variations in mortality, and it provides adequate explanations for urban area differentials in infant mortality, the uptake of vaccination, and the incidence of infectious disease. Specific applications of SAA in health planning and research are discussed.

The recent introduction of national systems of health care planning and of resource allocation (Department of Health and Social Security, 1976a and b) has concentrated attention on the effectiveness of existing health information systems and on the relevance of available data. Both leave much to be desired. National norms of provision for service planning are still, in general, based on a crude population count; and although the introduction of mortality data as indicators of need for resource allocation is a step in the right direction, it is widely acknowledged that a considerable degree of refinement will be required. The importance of socio-economic factors as determinants both of health (Registrar General, 1971; Adelstein and White, 1976; Syme and Berkman, 1976) and use of health services (Cartwright and O'Brien, 1976; Forster, 1976) is well recognised, but it has not as yet proved possible to quantify variations in a way useful to health planners. Census information on socioeconomic status by occupation is readily available (Office of Population Censuses and Surveys, 1970, 1975) but problems arise when this is employed to compute rates of morbidity or mortality by social class (Buechley *et al.*, 1956; Office of Population Censuses and Surveys, 1970; Nagi and Stockwell, 1973). In addition, doubts are cast upon the validity of this one-dimensional information as a proxy for the multidimensional socioeconomic aspects of need in health care planning because it imposes the necessity to generalise from the level of the individual to that of the population (the ecological fallacy). Social class is better employed to classify data than to explain differences (Central Statistical Office, 1975). None the less, information on social class, housing conditions, population density, and other relevant variables has in the past been collected by health planners to produce composite socioeconomic valuations of areas (Coulter and Guralnick, 1959), and such exercises can influence the allocation of scarce resources at the local level (Donaldson, 1976).

One approach, as yet untried in health care planning, which appears to permit the reliable use of socioeconomic data at population level, is that of social area analysis (SAA). This has been defined as 'a set of integrated procedures designed to study characteristics of groups or subpopulations of people who live in defined geographic areas' (Struening, 1974). Early studies in this field used census data to derive multivariate classifications which ranked census tracts by social status and urbanisation (Shevky and Williams, 1949). Although the initial emphasis was on describing groups, research workers soon realised the value of relating the classifications to non-census data such as crime and suicide rates (Shevky and Bell, 1955). Applications in the field of health care, however, were slow to develop; psychiatry is the sole health care topic in which SAA has been employed in Britain (Philip and McCulloch, 1966; Giggs, 1973; Skrimshire, 1976a). Possible uses in the analysis of health data have been suggested (Craig, 1975), but actual

applications have been confined to the United States of America, where the use of SAA has demonstrated social area typologies which identify individual variables associated with certain disease states such as total mortality and cause-specific mortality (Nagi and Stockwell, 1973), tuberculosis (Lebowitz and Malcolm, 1964; Guerrin and Borgatta, 1965), mental illness (Struening and Lehmann, 1969), venereal disease (Lebowitz and Malcolm, 1964; Struening and Lehmann, 1969), uptake of antenatal care (Struening and Lehmann, 1969), low birth weight (Struening *et al.,* 1973), and infant mortality (Lebowitz and Malcolm, 1964; Struening *et al.,* 1973).

Methods

SAA is a generic title covering a variety of possible approaches. Convenience and comparability have tended to dictate the analysis of census areas (for example, wards, or enumeration districts) in terms of census variables, but there is no reason why health districts or social service variables should not be employed. Statistical methods suitable for SAA are numerous. Multiple correlation analysis has been used to select study variables and to demonstrate their interrelationships; univariate techniques such as hierarchical multiple regression analysis, and multivariate techniques such as factor analysis, principal component analysis, and cluster analysis have been used to produce classifications. The choice of variables and statistical methods depends to some extent on the aims of the analysis, of which two main types can be distinguished. Purposive classifications are intended to analyse a predetermined dimension or quantity, such as social deprivation or infant mortality, and variables may be chosen with this in mind. Pragmatic classifications are essentially descriptive, distinguishing different types of area, rather than ranking them on a particular dimension. This type of analysis need not preselect variables, as the intention is to characterise areas in terms of the differences that exist between them. Both of these methods of SAA have potential applications in the health field, but purposive analyses may be considered more appropriate to study the distribution of particular diseases or the uptake of particular services, while pragmatic analyses may be expected to demonstrate unanticipated associations and therefore to be of more use in aetiological research.

Evaluation of British studies

Although SAA has shown its potential in the United States, in Britain two existing analyses lend themselves to evaluation in the health services.

LIVERPOOL SOCIAL AREA STUDY

The area typology in this study (Webber, 1975) is derived from census variables, but data were also collected concurrently on 20 non-census variables of which the rates for infant mortality, infectious diseases, and uptake of triple vaccination are of particular interest.

The methodology of the Liverpool study is representative of most British analyses. The data base was the 1971 census small area statistics tables published by the Office of Population Censuses and Surveys, and the level of analysis was the enumeration district (ED), of which there were 1120 in Liverpool in the 1971 census, with a mean ED population of 545. Three hundred variables, expressed in ratio form, were extracted from the ED household and population records. These included age, household composition, ethnic origin, fertility, and housing on a 100% sample basis, and socioeconomic status, educational attainment, residential and occupational mobility, employment structure, and travel to work on a 10% sample basis. Because of the complexity of manipulating 300 pieces of information for 1120 EDs, the EDs were grouped into relatively homogeneous 'basic data areas'. Using multiple correlation analysis, principal component analysis and from previous experience with 1966 census analyses, 14 key variables were selected from the 300 available. Then by cluster analysis, adjacent EDs whose similarity with respect to these 14 variables was greater than a specified threshold were grouped together, reducing the 1120 EDs to 418 basic data areas. The contiguity constraint by which only adjacent EDs could be grouped together was then released, and a wider set of 40 key variables was used to group the basic data areas into 25 'clusters'. The method employed was iterative reallocation, by which the areas are grouped and regrouped into the specified number of clusters until each area is closer to the 'mean' of its cluster than to that of any other cluster. Finally the 25 clusters were aggregated, again using the 40 key variables as criteria, into five groups known as 'Families'. A series of comparisons was made of all possible pairs of clusters, and the most similar pair was fused at each stage, until five 'Families' were formed. The intention at each stage of the grouping process was to reduce the units while retaining as much as possible of the initial variance in the data, and in fact the 25 clusters accounted for over 60% of the variance in 32 of the 40 key variables, while the five Family classification retained over 40% of the variance in 25 of the key variables. The broad characteristics of the five Families were as follows:

Family 1. A high status owner-occupied area type.

Family 2. Areas of subdivided housing with young people, and furnished privately rented accommodation in small units.

Family 3. Inner, older, council estates.

Family 4. Outer, more recent, council estates.

Family 5. Areas of older Victorian terraced housing, mostly privately rented unfurnished, and much of it lacking an inside WC.

Social class characteristics of the Families are shown in Table 1. Social Classes I, II, and IIIN are seen to be represented most strongly in the areas constituting Family 1, and to a lesser extent in Family 2. Infant mortality, triple vaccination, and infectious disease admissions for the Families are shown in Table 2. In each case there are highly significant differences between the Families. Infant mortality is lowest in Family 4, while the rate in Family 5, which is virtually identical with Family 4 in social class structure, is nearly 50% higher. The highest infant mortality is in Family 2. The uptake of triple vaccine is highest in Families 1 and 5; the rate in Family 5 is 11% higher than that in Family 4,

and 25% above that in Family 2. Infectious disease notification is lowest in Families 1 and 4; it is highest in Families 2 and 3, which share the worst performance with regard to each of the variables.

These results would seem to suggest that, while the performance of areas at the extremes of the social class distribution (that is, Families 1 and 3) is broadly predictable, a more detailed analysis of socioeconomic needs is required in the middle range. The poor overall performance of Family 2, for instance, may be related to its highly heterogeneous population and to the poor housing that characterises its constituent areas. These consist primarily of very large Victorian villas which have declined in parallel with the merchant classes for whom they were built; their occupants have become those groups with the least access to housing aid—such as students, immigrants, single parent families, the single elderly, and the transient and rootless population often found in city centres. Families 4 and 5 consist of very different types of area, despite their uniformity in terms of social class; the low infant mortality of the former could be related to its low incidence of marriage and fertility, and its relatively good provision of amenities; while the higher vaccine

Table 1 *Social class distribution between the five Families (social class of head of household as percentage of city total)*

Social Class	Family										Total
	1		2		3		4		5		
I and II	5·8	(28)	1·2	(14)	0·3	(3)	2·0	(6)	1·7	(6)	11·0
IIIN	6·6	(32)	2·2	(25)	1·0	(11)	5·8	(18)	4·4	(16)	20·0
Subtotal I, II, IIIN	12·4	(60)	3·4	(39)	1·3	(14)	7·8	(24)	6·1	(22)	31·0
IIIM	5·4	(26)	2·7	(31)	2·0	(22)	12·6	(38)	10·9	(39)	33·6
IV	2·1	(10)	1·7	(19)	2·5	(28)	7·8	(24)	6·7	(24)	20·8
V	0·7	(3)	1·0	(11)	3·1	(35)	4·8	(15)	4·5	(16)	14·1
Subtotal IIIM, IV, V	8·2	(39)	5·4	(61)	7·6	(85)	25·2	(77)	22·1	(79)	68·5
Total	20·6	(99)	8·8	(100)	8·9	(99)	33·0	(101)	28·2	(101)	99·5

Social class distribution within the Families is shown in parentheses.

Table 2 *Infant mortality, triple vaccination, and infectious diseases within the five Families*

Family	Population[1]	Live births[2]	Infant mortality[3]	Infant mortality rate[4]	Triple vaccine uptake[5]	Vaccine uptake rate %	Infectious disease[6]	Infectious disease rate[7]
1	125 921	289	25	14·4	47	49	84	27
2	52 790	220	36	27·3	17	23	73	55
3	52 600	163	25	25·6	8	15	78	59
4	189 776	433	35	13·5	53	37	140	30
5	158 843	565	68	20·1	90	48	174	44
Total	579 930	1670	189	18·9	215	39	549	38
χ^2 (4 DF)	—	—	13·6	—	31·6	—	49·5	—
P	—	—	<0·01	—	<0·001	—	<0·001	—

1. Population at 1971 census, excluding special EDs.
2. Live births in March and April 1971 (see 8).
3. Infant deaths in the year 1971 (see 8).
4. Per 1000 live births a year, employing the live birth figures for March and April 1971 to calculate notional denominators.
5. Children receiving complete course of triple vaccination (diphtheria, tetanus, and pertussis) within 14 months of birth in a 1 in 3 random sample of the live births in March and April 1971 (see 8).
6. Hospital admissions in Liverpool, October–December 1972, with principal diagnosis of infectious or parasitic disease (*ICD* code nos. 000–136). Source: Hospital Activity Analysis (see 8).
7. Per 10 000 population a year
8. A census gazetteer relating addresses to EDs was used to allocate addresses of live births, infant deaths, children vaccinated, and infectious disease admissions to the appropriate Family.

uptake of the latter might reflect its lower incidences of overcrowding and residential mobility. The lower infectious disease rate in Family 4 is probably due to its advantages in housing stock. Such comparisons as these between Families 4 and 5 stress the importance of quantifying the influences of socioeconomic variables other than social class; this is emphasised by the inverse relationship between infant mortality and infant vaccination rates which has been demonstrated at national level (West and Lowe, 1976), and is evident in Families 1, 2 and 3, but does not appear to be true for Families 4 and 5.

SAA OF BRITISH LOCAL AUTHORITY DISTRICTS
Similar methods to those described above have been employed in an analysis of the local authority districts in England, Scotland, and Wales (Webber and Craig, 1976). Forty socioeconomic variables were used to group the 457 districts into 30 clusters by iterative reallocation, and to aggregate the 30 clusters into six 'Families' by stepwise progression. At the cluster level 75% of the initial variability between the districts was retained, and at the family level 48% was retained. The relative lack of homogeneity of data at district level restricts the applications of SAA on this scale; broad descriptive classifications and comparative exercises of the type illustrated below would appear to be the most appropriate uses.

The social area typology derived in this study was evaluated by a comparison, at the 30 cluster level, with the 113 area health authorities (AHAs) of Great Britain, to show variations in mortality. Geographical variation in mortality (Office of Population Censuses and Surveys, 1976) is as well known as socioeconomic variation; to validate SAA therefore, it is necessary to demonstrate a greater discrimination than is shown by existing methods of classification. The level of AHA would appear to be the most appropriate at which to analyse health service data in the reorganised National Health Service (Jones and Masterman, 1976). The imperfectly defined relationship between the AHAs and the local authority districts in England (Department of Health and Social Security, 1974) can be exactly established if population data are used to assign 'shared' districts to the most appropriate AHA; these difficulties do not exist with regard to Scotland and Wales. Standardised mortality ratios (SMRs) and population data for the 457 districts (Office of Population Censuses and Surveys, 1976; Registrar General, Scotland, 1976) were used to compute SMRs for each of the 30 clusters in the social area typology. The variances between-clusters and within-clusters for SMR were compared by a variance ratio test. The between-cluster variance was the greater and the ratio was highly significant (F = 24·94, DF = 29 428; P <0·001). A further variance ratio test was then performed for SMR between clusters and between AHAs; the ratio of the between-cluster to the between-AHA variance was also highly significant (F = 3·55, DF = 29 112; P = <0·001). Area types derived by SAA are thus sensitive to mortality in their grouping of districts, and are more meaningful in explaining mortality variations than are the geographically defined AHAs.

Discussion

A number of applications of SAA in the field of community medicine can be envisaged. Perhaps the most important is the establishment of an information base for resource allocation and the planning of health services, since the use of geographical SMRs and of national bed norms gives cause for concern about the appropriateness of current practices. After the most suitable variables have been defined by correlational analysis, purposive SAAs could be undertaken to establish comparative resource or service needs for different types of area.

An additional application in the planning field is the use of SAA to highlight the needs for health care in small geographical areas, independently of AHA or health district boundaries, and hence to indicate the necessity for appropriate concentration of resources. In some cases, where such areas traversed administrative boundaries, this could result in the taking of joint action by adjacent AHAs or health districts which might not otherwise have occurred. The existence of the need for such localised action, whether within or across administrative boundaries, has been demonstrated in *ad hoc* studies (Donaldson, 1976; Skrimshire, 1976b, 1977), and the need for further information to elucidate this problem is evident.

SAA could be employed to produce samples of area types in which variations in morbidity, mortality, or use of services could be studied against a background of known socioeconomic comparability. Evaluative studies of innovations in service provision or of sensitivity of health indices to socioeconomic variations would similarly profit from a setting defined in this fashion.

A further role in research techniques is to identify socioeconomic factors in the aetiology of disease. Using SAA it is possible to study associations between cases of disease and socioeconomic characteristics of very small and therefore extremely homogeneous areas, grouped according to their similarity, which separately would not produce sufficient cases for such analysis, even over long periods. As an example, it has been suggested that

ncome is the socioeconomic variable most closely associated with social class differences in mortality, and that its important independent effect is mediated through qualitative variations in diet (Wilkinson, 1976). A purposive SAA which identified similar areas on the basis of socioeconomic group and National Food Survey dietary data could be related to mortality data to clarify this issue, and to provide an appropriate sample of areas for more intensive investigations.

Another important aspect of SAA is the ability to identify 'target groups' of populations with respect to particular variables and its applications to health education, in which there is little valid epidemiological research, can be readily envisaged.

It is intended to undertake SAA of some of the kinds described, with a view to further evaluation in relation to health and health services.

References

Adelstein, A. M., and White, G. C. (1976). Causes of children's deaths analysed by social class (1959-63). In *Child Health: A Collection of Studies*. Office of Population Censuses and Surveys. Studies on Medical and Population Subjects, No. 31, p. 24. HMSO: London.

Buechley, R., Dunn, J. E., Linden, G., and Breslow, L. (1956). Death certificate statement of occupation: Its usefulness in comparing mortalities. *Public Health Reports (Washington)*, 71, 1105-1111.

Cartwright, A., and O'Brien, M. (1976). Social class variations in health care and in the nature of general practitioner consultations. In *The Sociology of the National Health Service*. Sociological Review Monograph 22, pp. 77-98. Edited by M. Stacey. University of Keele.

Central Statistical Office (1975). Social commentary: Social class. In *Social Trends*, No. 6, p. 10. HMSO: London.

Coulter, E. J., and Guralnick, L. (1959). Analysis of vital statistics by census tract. *Journal of the American Statistical Association*, 54, 730-740.

Craig, J. (1975). Multivariate socio-economic area classifications. *Statistical News*, 28, 4-6.

Department of Health and Social Security (1974). *Management Arrangements: Health Districts*. National Health Service reorganisation circular HRC(74)23. DHSS: London.

Department of Health and Social Security (1976a). *The National Health Service Planning System*. DHSS: London.

Department of Health and Social Security (1976b). *Sharing Resources for Health in England. Report of the Resource Allocation Working Party*. HMSO: London.

Donaldson, R. J. (1976). Urban and suburban differentials. In *Equalities and Inequalities in Health*, pp. 127-140. Edited by C. O. Carter and J. Peel. Academic Press: London.

Forster, D. P. (1976). Social class differences in sickness and general practitioner consultations. *Health Trends*, 8, 29-32.

Giggs, J. A. (1973). The distribution of schizophrenics in Nottingham. *Transactions of the Institute of British Geographers*, 59, 55-76.

Guerrin, R. F., and Borgatta, E. F. (1965). Socioeconomic and demographic correlates of tuberculosis incidence. *Milbank Memorial Fund Quarterly*, 43, 269-290.

Jones, D. R., and Masterman, S. (1976). National Health Service resources: Scales of variation. *British Journal of Preventive and Social Medicine*, 30, 244-250.

Lebowitz, M. D., and Malcolm, J. C. (1964). Socioeconomic analysis of the Alameda County Health Department Jurisdiction. *American Journal of Public Health*, 54, 1876-1881.

Nagi, M. H., and Stockwell, E. G. (1973). Socioeconomic differentials in mortality by cause of death. *Health Services Reports (Washington)*, 88, 449-456.

Office of Population Censuses and Surveys (1970). *Classification of Occupations*. HMSO: London.

Office of Population Censuses and Surveys (1975). *Census 1971 England and Wales*. Economic activity county leaflets. HMSO: London.

Office of Population Censuses and Surveys (1976). *Local Authority Vital Statistics: England and Wales 1974*, Series VS, No. 1, Table 1. HMSO: London.

Philip, A. E., and McCulloch, J. W. (1966). Use of social indices in psychiatric epidemiology. *British Journal of Preventive and Social Medicine*, 20, 122-126.

Registrar General (1971). *Decennial Supplement, England and Wales 1961. Occupational Mortality Tables*. HMSO: London.

Registrar General Scotland (1976). *Annual Report 1974. I. Mortality Statistics*, Table C2.29. HMSO: Edinburgh.

Shevky, E., and Bell, W. (1955). *Social Area Analysis*. Stanford University Press: California.

Shevky, E., and Williams, M. (1949). *The Social Areas of Los Angeles: Analysis and Typology*. University of California Press: California.

Skrimshire, A. M. (1976a). A small area analysis of self-poisoning and self-injury in the region of Oxford. *Journal of Biosocial Science*, 8, 85-112.

Skrimshire, A. M. (1976b). *A Deprived Health District in a Prosperous Region: A Study of Reported Sickness in a Part of Newham*. Social Evaluation Unit, University of Oxford.

Skrimshire, A. M. (1977). Health in Newham. *New Society*, 39, 396.

Struening, E. L. (1974). Approaches to evaluation: Social area analysis. *International Journal of Health Services*, 4, 503-514.

Struening, E. L., and Lehmann, S. (1969). A social areas study of the Bronx: Environmental determinants of behavioral deviance and physical pathology. *Research*

Publications Association for Research in Nervous and Mental Disease, **47,** 130-138.

Struening, E. L., Rabkin, J. G., Cohen, P., Raabe, G., Muhlin, G. L., and Cohen, J. (1973). Family, ethnic, and economic indicators of low birth weight and infant mortality: A social area analysis. *Annals of the New York Academy of Sciences,* **218,** 87-107.

Syme, S. L., and Berkman, L. F. (1976). Social class, susceptibility, and sickness. *American Journal of Epidemiology,* **104,** 1-8.

Webber, R. J. (1975). *Liverpool Social Area Study. 1971 Data: Final Report.* Technical Paper 14. Planning Research Applications Group: London.

Webber, R. J., and Craig, J. (1976). Which local authorities are alike? *Population Trends,* **5,** 13-19.

West, R. R., and Lowe, C. R. (1976). Regional variations in need for and provision and use of child health services in England and Wales. *British Medical Journal,* **2,** 843-846. Supplementary Table D.

Wilkinson, R. G. (1976). Socioeconomic differentials in mortality—the importance of diet. M.Med.Sci. thesis, University of Nottingham.