HUMAN ECOLOGY: TOWARD A HOLISTIC METHOD

EDWARD S. ROGERS

AND

HARLEY B. MESSINGER

The idea of the human environment as a complex interacting web has had general acceptance in the biological and social sciences since the time of Darwin. The subordinate theory of holism, as developed by Smuts[1] and Bews,[2] holds that a consequence of the environmental web is the necessity of studying any natural phenomenon in the context of its total environment if the occurrence is to be thoroughly understood. That theory is also widely, but by no means universally, accepted, and in studies involving relatively simple biological ecosystems holism is often approximated. However, in studies of human ecology the immense complexity of the human environment appears to have discouraged any major attempts to apply, or even to develop, holistic methods. Yet it seems important that this be done.

The need may be illustrated by an example from public health. In the search for the causes of illness, especially the chronic diseases, the spectrum of environmental relationships under investigation are being widened with rewarding results. Identifying the proximate cause is no longer sufficient, and may even lead to dangerous half-truths and the prescribing of inadequate or inappropriate remedies and controls. Moreover, control measures themselves may have complex consequences which should be anticipated. For example, those responsible should have been able to foresee that rehousing, *per* *se,* would do little to improve the health, or other major problems, of slum populations—as has now been amply demonstrated[3]—and more appropriate measures might have been undertaken earlier had the whole environment of the slum been properly assessed. As another illustration, the attempt to stem the rising incidence of lung cancer by control of cigarette smoking has raised economic questions not only in the states where tobacco is the main industry, but also in the country as a whole where the tax revenues and advertising income are important. No effective control program is likely until these and other problems, many of which should be foreseeable, are resolved.

Thus, the ability to discover causative sequences and the wise choice of measures for dealing with them seem to require comprehensive information about the universe of relationships of which both are part. The epidemiology of cause cannot be separated meaningfully from the epidemiology of consequences, and the levels of validity attained in any study will depend upon the completeness of the model used—i.e., upon how holistic it is.

Although, in recent years, epidemiology has moved toward holism in recognizing and attempting to deal with problems of multiple causation, it has done so by a process of expanding the environment recognized as being of probable relevance. It is a question of how long such a gradual approach would require to reach the level of completeness that holism demands. Both, new sources and types of information descriptive of man's total environment and environmental associations— physical, biological, social and mental—and a new analytical technique capable of dealing with such data holistically, must be sought. To be ecological the analytical technique should reveal the patterns of association between man and his environment. To be holistic the technique must treat man and his environment as an interacting whole, and be open-ended as far as the nature and number of environmental variables are concerned.

This study explores the hypothesis that, as an analytical technique, factor analysis can fulfill these criteria when coupled with the potential capacity of computers to handle large volumes of data. To test the validity of this hypothesis in a preliminary fashion, factor analysis is applied to the demographic and vital data for 1960 from the continental United States. The factors thus obtained can be viewed as a set of axes which provide an ecological framework. In this framework all variables are related to these axes by their factor coefficients. Existing, and possibly unsuspected, systems of association among variables may be revealed, especially when the results are displayed by spherical analysis (illustrated later). The technique is holistic in that the matrix of correlations among pairs of variables is factored in its entirety without casting any one variable or group of variables in a dependent role, as is the case in multiple regression. Also, the technique is open-ended and can, in principle, handle any number of variables.

Of course, neither factor analysis nor any other multivariate technique can generate new data and the results are no more valid than the sum total of the information supplied. However, with the prospect of a useable, holistic method of ecological analysis in sight, constructive attention can be given to the development of data of the kinds and in the quantity necessary to complete the process.

STUDY METHOD

Of the several kinds of factor and cluster analysis available, two—principal factor analysis with varimax rotation[4] and Tryon's cluster analysis[5]—have been used most extensively in these studies and, with the data being used, have been found to yield quite similar results. However, the findings in this report are only those obtained by the factor analysis.

The original data pool consisted of 372 items.[6] These included demographic variables dealing with age, sex and race; foreign origin, school completion and enrollment; proportions in the labor force or unemployed; detailed occupational and industry breakdowns by sex, family income and income of employed persons; marital status by race; fertility and family size and structure; age and condition of housing, years of occupancy and an index of crowding and other characteristics of housing. Mortality variables included measures of fetal mortality and mortality by specific age and sex groups; death rates and proportions of deaths for 44 specific cause groups; crude

TABLE I. FACTOR COEFFICIENTS* OF 87 VARIABLES, 48 STATES

Variable Number†	Variable Name	Factor 1	Factor 2	Factor 3	Factor 4	Communality (h^2)
1	Median age	2855	−0473	(4434)	(−8308)	9706
2	Under 1 year	−0461	0578	−1314	(8788)	7950
8	45–54 years	0418	−0245	(4556)	(−7845)	8253
14	Age 65 and over	(4696)	−1041	−1790	(−7974)	8992
17	Males in total population	3052	(5089)	−0290	(4435)	5497
23	White	(8770)	0769	2314	−1453	8497
28	Foreign born	2498	−1853	(7474)	−2709	7287
40	0–4 years elementary school	(−8636)	−1462	−2800	1625	8720
42	8 years elementary school	(6649)	−2467	−1619	(−3986)	6880
50	College enroll 18–24 years	(7322)	−1225	1972	0329	5911
66	Males-laboring	(−4130)	2682	(−6645)	2236	7341
71	Females-crafts occupation	2878	(−3954)	3570	−3439	4849
73	Females-private household occupation	(−7327)	0315	(−5234)	1731	8417
80	Income-$4,000–5,999	(7042)	−0844	−0274	−0825	5105
84	Income-$15,000+	1190	0568	(9067)	−0534	8423
93	Separated, white	(−5896)	0752	(6283)	−0930	7566
94	Widowed, white	1179	(−4416)	−0928	(−7189)	7343
95	Divorced, white	0308	(8529)	2343	0059	7833
100	Divorced, nonwhite	(7537)	2216	−0077	−2087	6608
101	In 1955 residence	1590	(−8796)	−1862	−2909	9183
113	Dilapidated housing	(−7215)	0329	(−6191)	0900	9131
115	Up to 0.50 occupants/room	(5251)	−1916	−0132	(−6767)	7706
122	In two-unit housing	0912	(−5014)	(4991)	−2972	5972
128	Crude fertility rate	2607	−0357	−3950	(6839)	6930
131	Males-mining industries	0123	2506	−1030	3262	1800
132	Males-construction	−3780	(6860)	0347	3634	7468
133	Males-manufacturing	−0851	(−5080)	2756	−3132	4393
138	Males-machinery manufacturing	2833	(−4820)	3409	−2678	5005
139	Males-electric machinery mfg.	0795	(−4244)	(5163)	−3388	5677
143	Males-other durable goods	0081	(−4129)	(4586)	−3802	5255
147	Males-printing manufacturing	2833	−3428	(6854)	−3535	7925
150	Males-nondurable goods	−0826	−1334	−0877	−2728	1067
154	Males-communications	1347	(5027)	(7508)	−0378	8360
158	Males-general merchandise	−2689	−1285	3774	−1235	2465
159	Males-eating places	0404	2572	(8043)	−1332	7325
164	Males-hotels, lodging	−1869	(5778)	3173	−0277	4703
165	Males-other personal services	(−7871)	1612	1415	−1345	6837
168	Males-government education	(5122)	(4679)	−0488	(5135)	7474
169	Males-private education	2372	(−4277)	3193	−1841	3750
171	Males-other professions	−0037	1378	(7872)	1444	6595
196	Females-trucking	(5822)	2988	−0976	1214	4525
207	Females-repair services	3081	(7813)	−0625	1065	7206
209	Females-other personal services	(−7205)	1274	(−5286)	1827	8481
211	Females-medical	(8201)	−0341	0141	−0858	6814
212	Females-government education	2108	2911	(−5157)	(4116)	5646
213	Females-private education	(5497)	(−4532)	−0122	−0639	5119
214	Females-welfare	(8187)	−1792	0359	−0522	7063
222	Average family size	(−4130)	−3063	(−4225)	(6948)	9255

TABLE I. CONTINUED

Variable		Factor				Communality
Number †	*Name*	*1*	*2*	*3*	*4*	*(h^2)*
225	Median age-employed males	3396	−1384	0671	(−8224)	8154
227	Median earnings, male experienced labor	(5030)	1315	(7661)	0158	8574
229	Median earnings, female experienced labor	2140	−0211	(8686)	−1949	8387
231	Males with income	(7132)	3168	3276	−1067	7277
232	Females with income	0022	1189	2574	−3365	1937
235	Units with telephone	(7800)	−1598	(4657)	−1432	8714
236	Autos per occupied unit	(5832)	(5410)	0924	(4509)	8447
245	Fetal death rate	(−7707)	−2106	−0880	−0959	6553
250	Deaths under one year	(−8129)	1910	−2308	2758	8267
278	Crude death rate	0438	−2845	−0568	(−8836)	8669
280	CDR, nonwhite	−2954	−0617	−3317	−0724	2063
293	Cancer-digestive organs	3517	(−4180)	3584	(−6839)	8945
294	Cancer-respiratory system, not secondary	0041	−1253	(4126)	(−7523)	7518
296	Cancer-genitals	1697	−3202	−2782	(−7568)	7813
297	Cancer-urinary organs	(5161)	−1573	(3985)	(−5776)	7835
306	Rheumatic fever and heart	(6086)	−0698	(4543)	−0631	5857
307	Arteriosclerotic heart	3168	−3330	2607	(−7835)	8931
312	General arteriosclerosis	(4645)	−1924	−0874	(−5539)	5672
316	Ulcer stomach + duod.	(4907)	0276	(4361)	(−4378)	6233
335	Other infections + parasitic	(−5400)	1490	−1758	(5738)	6740
354	Hypertensive heart disease	(−6404)	−3889	−1991	−0490	6034
358	Chronic nephritis	(−5932)	−0702	(−4651)	0342	5743
363	Cirrhosis of liver	0543	1848	(9140)	−0338	8736
365	Acute nephritis + other kidney	(−7355)	−0191	−1712	0725	5759
367	Maternal mortality	(−8153)	0931	−1257	(3990)	8484
368	Congenital malformation	2308	0160	−0589	(8828)	8364
370	Infections of newborn	(−4367)	2788	−0829	(7201)	7939
372	Other disease and tuberculosis	0577	(6181)	0784	(5610)	7062
373	Vehicle accidents	−0826	(5789)	−1430	(6563)	7932
375	Suicide	1599	(7291)	2368	3152	7125
376	Homicide	(−8101)	3177	−1139	3367	8835
377	Age-adjusted death rate	(−8669)	−1728	1542	0342	8063
378	Female net reproduction rate	3350	−0084	−2268	(5814)	5018
379	High-school dropouts	(−8419)	0416	−1587	0609	7394
380	Life expectancy at 45 years	(7427)	2719	−3059	0640	7232
381	Number living at 25 years	(6752)	−3449	3120	−3841	8197
382	Life expectancy at 75 years	−2695	(6449)	−2377	2799	6234
383	Life expectancy at birth	(9173)	−0263	−0243	−1709	8720
384	Female survival to 45 years	(8877)	−2348	1495	−2433	9248
	Percent of converged communality	30.18	13.39	17.34	20.88	81.79

* Principal factor analysis with varimax rotation of the 4 factors with eigenvalues greater than unity; communalities are those for first 4 factors only. The standard error of a factor coefficient is about 0.10 (see Reference 4, pages 439, 441). Loadings ≥ 0.4 are in parentheses.

† Numbers greater than 372, the total number of variables being used, do occur in the lists.

and age-adjusted death rates; and several life table functions derived from a separate analysis.

The number of variables was systematically reduced to a level which could be handled in a single factor analysis on the available computer. First, by a series of correlation analyses dealing with related blocks of variables many of the variables which intercorrelated at levels of 0.8 or above were removed. Next, a somewhat arbitrary selection was made from the remaining blocks of closely correlated variables. In this selection, demographic variables which had high correlations with mortality variables were favored because of the initial interest in identifying clusters relating to health levels. (This was a regrettable step in procedure and has not been continued in further studies.[7]) In this manner, a group of 87 variables was derived which was generally representative of the types of data contained in the original pool (Table 1).

FINDINGS FROM FACTOR ANALYSIS

Using these 87 variables and principal factor analysis with varimax rotation, nine essentially independent factors are derived.[8] The first four factors exhaust 82 per cent of the common variance and only the results from these are presented in Table 1,[9] though the others are not without interest.

As an aid in visualizing the patterns of association among the variables that are significantly related, three-dimensional models of various sets of clusters may be projected onto a two-dimensional surface by Tryon's spherical analysis technique.[10] Figure 1 does this for Dimensions 1, 3 and 4 and Figure 2 for Dimensions 2, 3 and 4. The Z, X and Y axes are orthogonal and represent complete independence from each other. The signs of the factor loadings are important only within each factor. On the charts an "R" means that the location of the variable is reflected from the opposite side of the sphere and that it carries the opposite sign from that of the variables that are projected directly and are designated by an "O." To simplify the discussion of these variables, and their factor loadings, which represent relative, rather than absolute, levels of the proportions or values concerned, they will be referred to as being either "high" or "low." Thus Variable A, with a negative factor loading of, say, – 0.8636 (designated by "R") might appear in close proximity to Variable B with a positive factor loading of + 0.8770 (designated by "0"). This relationship would be described as a low proportion or value of Variable A in association with a high proportion or value of Variable B.

In the two sets of three-dimensional projections shown, which display the variables in terms of three factors at a time, the variables are grouped in distinct associational systems or clusters. Some of these clusters appear to be "linked" by variables which are loaded about equally on the axes of two or more clusters. For example, in Figure 1, Factors 1 and 3 share Variables 306, 113, 358, 209 and 73.

At this stage in the study the presence of associational systems can be more satisfactorily recognized than can the exact nature of such systems. Productive study of the associations within a given system is not likely unless the data being analyzed are extensive enough to yield a reasonably complete picture. Thus, among the nine factors indicated in the analysis of 87 variables, Factor 1 (Table 1 and Figure 1) appears to contain the most detail and evidence of completeness justifying inference as to its probable nature. Whatever else it may eventually reveal, this factor shows strong associations among conditions having to do with survival experience. Low levels of mortality, as indicated by several variables (such as Variables 245, 250, 377, 380, 383, 384) are associated with such conditions as a high proportion of white population (Variable 23), high levels of education (Variables 42 and 50) and good income (Variables 80 and 231). Since Factor 1 is quite a large cluster containing more than 40 variables with very significant loadings, it is possible to factor analyze just this cluster to better define its internal relationships.[11] Six distinct subfactors can be obtained of which the first three exhaust 78 per cent of the common variance (Table 2). These three subfactors have been plotted by Tryon's technique (Figure 3). The associational subsystems thus revealed suggest that the variables in Factor 1 form subgroups in terms of such population characteristics as low economic status, nonwhite, poorly educated, poorly housed and high infant mortality (Subfactor 1–1); low infant and maternal mortality, high proportion of older people and high risks from malignancy and cardiovascular diseases (Subfactor 1–2); high educational attainment, high nonwhite divorce rate and favorable mortality experience and life expectancy (Subfactor 1–3).

Factor 2 (Table 1 and Figure 2) groups residential instability, i.e., a low proportion of the population with permanency of residence since 1955 (Variable 101), together with other variables suggestive of social instability, i.e., high prevalence of divorce (Variable 95), and high level of mortality from suicide (Variable 375). Also, this cluster shows a high proportion of females in the repair service industry (Variable 207). This is an association which cannot be rationalized at this juncture.

Factor 3 (Table 1 and Figures 1 and 2) strongly associates high levels of income (Variables 84, 227 and 229) and high mortality from cirrhosis of the liver (Variable 363). This intriguing association probably is not capable of being fully explained on the basis of the variables included in this cluster as they are both qualitatively and quantitatively insufficient to present a complete picture. For example, no variable among the 87 finally used adequately represents the "skid-row" type of population known to have a high rate of cirrhosis and which might very well be found in the same state populations as are high income families. Additionally, Factor 3 links high proportions of foreign-born (Variable 28), separated, white population (Variable 93) and four industrial categories (Variables 147, 154, 159 and 171) with low proportions of males in laboring occupations (Variable 66) and dilapidated housing (Variable 113).

Factor 4 (Table 1 and Figures 1 and 2) shows high proportions in the following: under 1 year of age (Variable 2), average family size (Variable 222), and crude fertility rate (Variable 128), coupled with low values in the following: median age (Variable 1), median age of employed males (Variable 225), proportion of widowed white population (Variable 94), and proportion 65 years old or over (Variable 14). These variables, indicative of a young population with high fertility, are associated, as might be expected, with low proportions of deaths from cancer (Variables 68, 75, 76), low crude

TABLE 2. FACTOR COEFFICIENTS* FOR 44 VARIABLES IN ORIGINAL DIMENSION ONE, 48 STATES

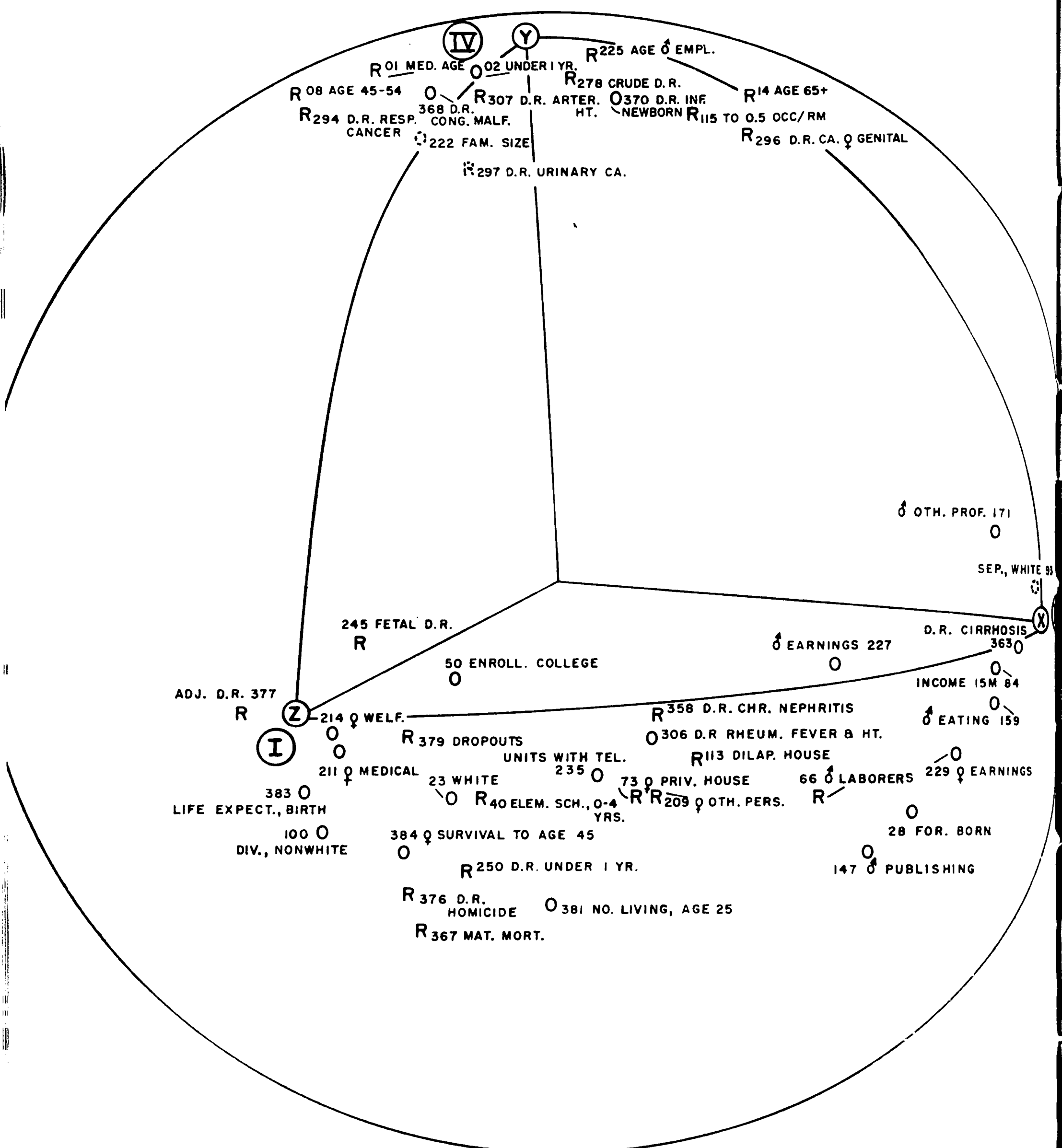
Variable Number†	Variable Name	Factor 1	Factor 2	Factor 3	Communality (h^2)
14	Age 65 and over	0085**	(7508)	(4335)	7517
23	White	(−7484)	1606	(4301)	7708
40	0–4 years elementary school	(8239)	−2539	(−3896)	8952
42	Eight years elementary school	−2376	(3885)	(3482)	3287
50	College enroll 18–24 years	(−4495)	1429	(5402)	5142
71	Female-crafts occupation	−2802	(6279)	0578	4761
73	Female-private household occupation	(8338)	(−3090)	−1951	8287
80	Income $4,000–5,999	(−5500)	1782	(3192)	4362
100	Nonwhite divorced	(−4197)	2735	(6187)	6337
113	Dilapidated housing	(8622)	(−3034)	−1761	8664
115	Up to 0.50 occupants per room	−2057	(7987)	2874	7628
122	In two-unit housing	−2975	(4964)	−2787	4126
131	Males-mining industries	−0187	(−5628)	0526	3199
138	Males-machinery manufacturing	−2751	(5712)	−0619	4057
150	Males-other nondurable goods	0448	2885	−1361	1038
165	Males-other personal services	(5072)	−0037	(−3705)	3946
169	Males-private education	−2211	(3810)	0246	1947
196	Females-trucking	(−4179)	−1028	(4901)	4254
209	Females-other personal services	(8330)	(−3556)	−1339	8382
211	Female-medical industry	(−4361)	1198	(5062)	4608
213	Female-private education	−1609	2474	(3499)	2096
214	Female-welfare industry	(−3628)	1910	(5741)	4977
231	Males with income	(−7686)	2345	(3764)	7874
235	Units with telephone	(−7396)	(4201)	2855	8050
245	Fetal death rate	(6053)	0494	(−4732)	5927
250	Deaths under one year	(6109)	(−5086)	(−4174)	8061
293	Cancer-digestive organs	−2706	(7988)	−0319	7124
297	Cancer-urinary organs	(−4873)	(6628)	0964	6860
306	Rheumatic fever and heart	(−7507)	1966	1180	6162
307	Arteriosclerotic heart	−2515	(8482)	−0358	7839
312	General arteriosclerosis	−2083	(6075)	1780	4441
316	Ulcer stomach + duodenum	(−6089)	(5086)	1001	6395
335	Other infections + parasitic	(5223)	(−6941)	−1013	7648
358	Chronic nephritis	(6505)	−1009	−1750	4640
365	Acute nephritis, other kidney	(6146)	−1132	(−3095)	4863
367	Maternal mortality	(5638)	(−4797)	(−4151)	7203
370	Infections of newborn	2156	(−8486)	−1981	8058
376	Homicide	(5209)	(−4951)	(−3749)	6571
377	Age-adjusted death rate	3058	−0803	(−9053)	9195
379	High-school dropouts	(5872)	(−3096)	(−5079)	6986
380	Life expectancy at 45 years	−1386	−0940	(9167)	8684
381	Number living at 25 years	(−5088)	(6754)	2913	7998
383	Life expectancy at birth	(−3975)	(3233)	(8121)	9220
384	Female survival to age 45	(−5576)	(4664)	(5224)	8014
	Percent of converged communality	32.58	26.02	19.80	78.40

* Principle factor analysis with varimax rotation of the 6 factors with eigenvalues greater than unity; communalities are those for the first 3 factors only. The standard error of a factor coefficient is about 0.14 (see Reference 4, pages 439, 441). Loadings ≥ 0.4 are in parentheses.

† Numbers greater than 372, the total number of variables being used, do occur in the lists.

** Decimal points omitted.

FIGURE I. SPHERICAL ANALYSIS OF VARIMAX FACTORS I, 3, AND 4

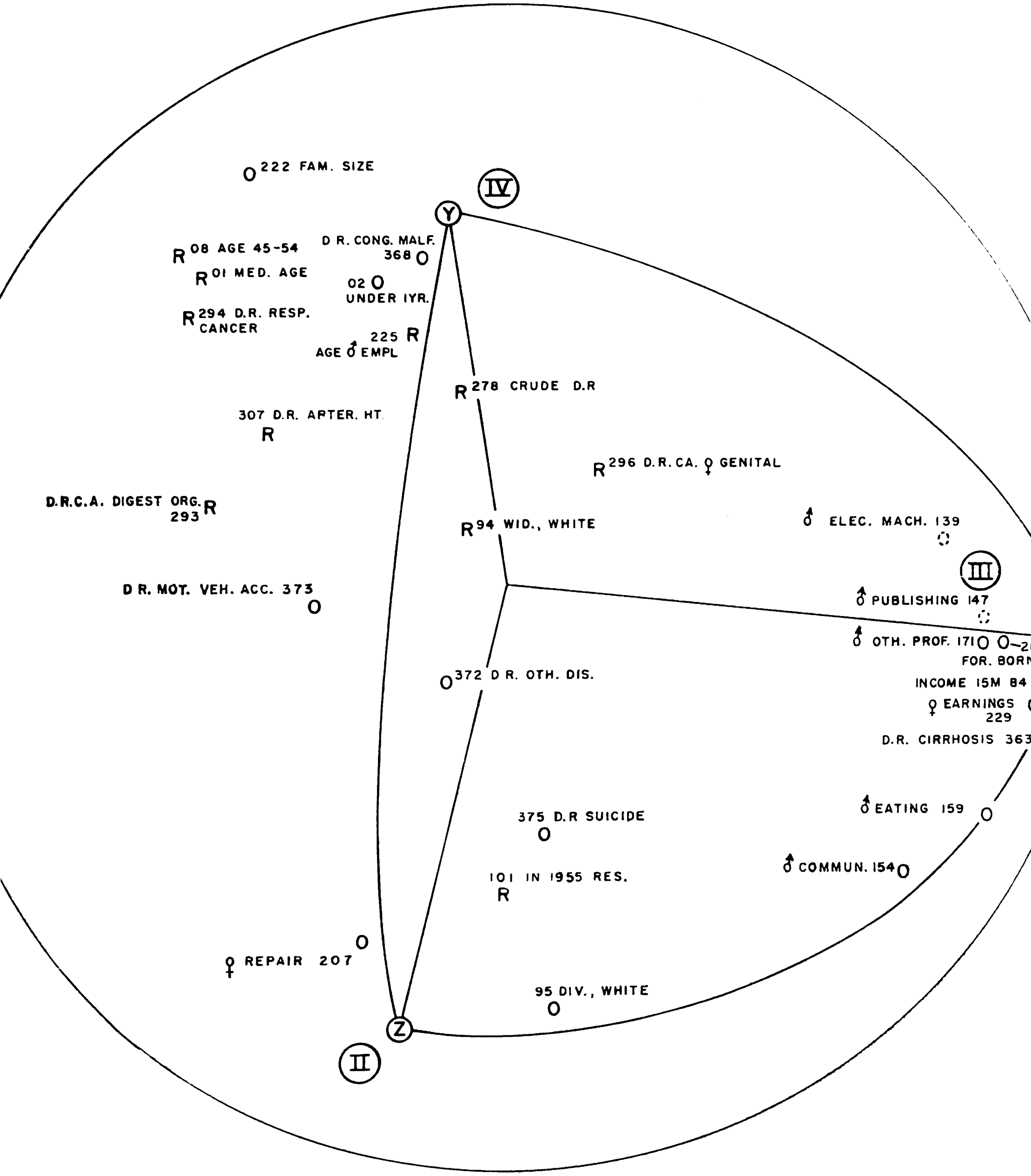


Note: Variables may be identified by reference to Table 1. All variables with at least 75% of their communality accounted for by factors 1, 3, and 4 are shown.

O locates variables with + sign on front of sphere (broken O on back of sphere).

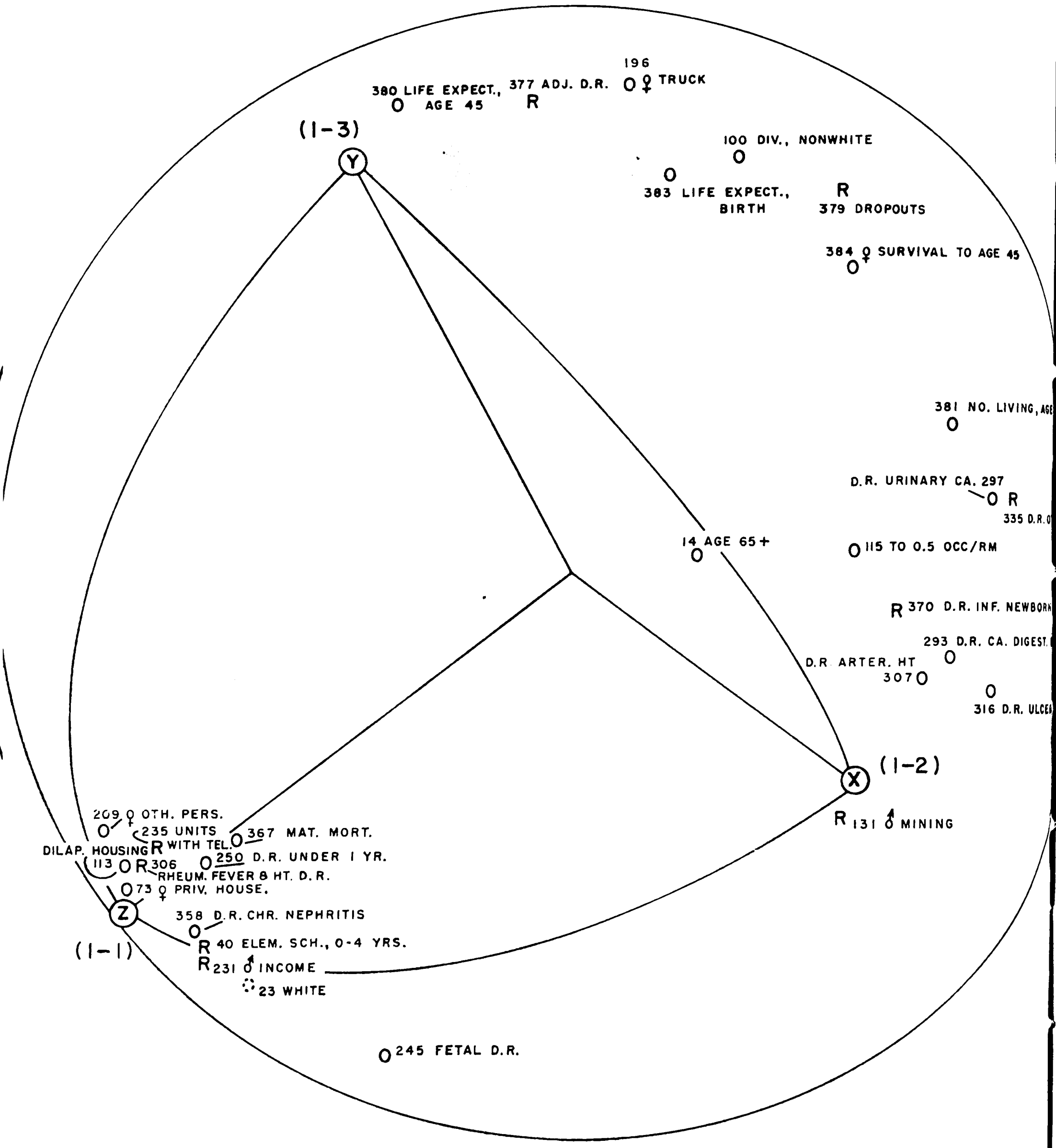
R locates variables with − sign on front of sphere (broken R on back of sphere).

FIGURE 2. SPHERICAL ANALYSIS OF VARIMAX FACTORS 2, 3, AND 4



: Variables may be identified by reference to Table 1. All variables with at least 75% of their communality accoun
rs 2, 3, and 4 are shown.
cates variables with + sign on front of sphere (broken O on back of sphere).
cates variables with − sign on front of sphere (broken R on back of sphere).

FIGURE 3. SPHERICAL ANALYSIS OF VARIMAX SUB-FACTORS I–I, I–2, AND I–3. OF FACTOR



Note: Variables may be identified by reference to Table 2. All variables with at least 75% of their communality accou[nted for] by factors 1-1, 1-2 and 1-3 are shown.

O locates variables with + sign on front of sphere (broken O on back of sphere).

R locates variables with – sign.

death rate (Variable 278), high proportions of deaths from infections of the newborn (Variable 370), and deaths from congenital malformations (Variable 368). Other variables associated on this factor are a low proportion of the population in units of under 0.5 persons per room (Variable 115) and a high proportion of deaths from motor vehicle accidents (Variable 373).

The remaining five factors indicated by the factor analysis are not without interest but they are not complete enough to add much to the present consideration of the utility of this method of ecological study and discussion of them is omitted.

DISCUSSION

The ideas and methods proposed here are importantly related to the work of others. As noted, the holistic emphasis goes back to Smuts and to Bews although the concept of environment as outlined in this paper appears to be somewhat more extensive. The holistic philosophy is consistent with present day interest in the study of systems.

Though similar in some respects, the use of factor analysis in this paper should not be confused with its use in social area analysis to obtain dimensions for typology of geographic areas. The focus is significantly different here in that variables are identified in terms of associational systems and the relations among these systems.[12] In consequence, every variable that is not completely redundant assumes a place of potential importance.

The approach as outlined here is numerical and objective in the same sense as is that of Sokal and Sneath in their work on numerical taxonomy,[13] but, again, it has a different focus. In numerical taxonomy, a collection of objects (individuals) is grouped into "natural groups" solely on the basis of an objective assessment of the number of shared attributes. Each attribute is given equal weight. The focus in this paper, however, is not on the taxonomic classification of objects, but rather on the identification of the associations of attributes (or variables) regardless of whether they have membership in one or many systems. The classification of attributes does not deny the possibility of taxonomic classification of the objects (in this study, states rather than individual organisms), but that is not its present purpose. Thus, an associational system should not be confused with either a biological "natural group," such as a species, or an "artificial group," such as "marine animals."[14] Nevertheless, it is a natural group in a literal sense and if attributes were to be substituted for objects they could be classified, in the main, by the associational system or systems to which they belong.

In factoring attributes of a defined set of subjects, account is taken of the amount of relevance each variable has to the set as a whole, i.e., its communality. On the other hand, Gouldner and Petersen,[15] in a study of the systems of attributes in 71 primitive or pre-industrial societies, assume a communality of one for each variable, i.e., that all of the variance of every variable is factorable. Otherwise their ideas and methods are very similar to those set forth here, and, apart from specific findings which reflect the contrasting types of societies studied, differ principally in matters of emphasis. Their "stratified system model" is a valuable concept and their method deserves more attention than it appears to have received.

That the analysis points to the existence of a survival-centered associational system, such as Factor 1, is not surprising since the census and vital data used owe their existence, in part, to their utility for the health and welfare services. Furthermore, screening the initial 372 variables involved choices in which interest in retaining health indexes resulted in further weighting the sample finally used. Even had the introduction of such bias been resisted, the proposed process presents a certain inescapable circularity. Selection bias may be taken for granted in any likely choice of variables to be studied and this, in turn, makes probable the finding of the associations which generated the bias in the beginning. However, this need not disqualify the procedure. In a cumulative series of such analyses, each cycle should add new variables and inclusiveness, and the total process would represent an increasingly close approximation to a holistic model of man's environmental relations. Admittedly, the complex data system necessary to deal with the measurable components of man's total environment presents an awesome prospect. However, it should occasion no greater alarm than that already being felt in confronting the knowledge explosion and the necessity of turning to the computer for data storage and retrieval at levels of volume and complexity far beyond the capacity of man's unaided brain. This paper suggests that a share of attention in this new effort be given to acquiring the data necessary for ecological studies.

The authors' experience is not yet sufficient to have tested the proposed method as it ought to be tested and results should be considered only as illustrative, but encouraging. Also, a few technical problems inherent in the existing methods of factor and cluster analysis merit further study.

Above all else, further testing of the method will require much broader coverage of descriptive information. A bold and imaginative approach is needed to develop new categories of man's environmental relations; for functional indexes such as Hawley's M.P.O. ratio $\left(\frac{\text{managers, proprietors and officials}}{\text{employed labor force}}\right)$;[16] and for study of existing indexes to make them more useful. For example, occupational classification based on current employment may serve certain economic and sociological study purposes, but is quite insufficient for health studies where accumulated exposure to certain types of occupational environments is important and requires an index with a time dimension. The comparatively large number of occupational-industrial variables retained in the study (25 out of 87) has resulted in their appearing frequently on most of the factors. At this juncture, very little interpreting of them can be accomplished. The probability is that many of these variables are composites of qualitative characteristics. Indeed, their frequent appearance as variables shared among two or more associational systems may be misleading. An ideal linking variable would associate a single characteristic in two or more systems. In any case, further study of these variables should be worthwhile and might provide clues to new descriptive categories of a more unitary nature. As noted, the prospect of a method for handling such ecological information analytically will hopefully serve as a stimulus to its acquisition.

The difficulty that may arise from excluding intercorrelating variables, provided they are not completely redunant, already has been noted (see also reference 7). Furthermore, the variables linking the systems and, therefore, not heavily loaded on any single factor are exceedingly important if the whole structure or associations among systems is to be shown.

Wherever the clusters of variables appear well enough defined, attempts to further study the nature of the associations among these variables would seem justifiable. A first approach to such study is illustrated in the subfactoring of Factor 1. Such separation of an identifiable associational system from the whole of which it is a part does not violate the idea of holism as it might at first appear to do. The segment removed has been identified by objective means as essentially self-contained, occupies known relations with adjacent systems, and can always be related back to the whole.

As presently developed, the method deals strictly with frequency associations among variables. The fact, alone, of such associations among variables in a cluster does not indicate the nature of these associations. It does not distinguish causative relations from any of a variety of other possibilities. Here, the application of knowledge about the characteristics of individual variables derived through sociological studies, in particular, should find useful focus. The statistical analysis itself may possibly be augmented by adapting time series techniques, cybernetic and other systems modes, and the like. Even simple application of the techniques used here to the same population at different points in time, or to different populations (cross-cultural) at the same point in time, would seem almost certain to yield more sensitive insights.[17]

Something should be said about the problems of "spurious" and "ecological" correlations. The former arises in the correlation of rates of occurrence for many different conditions in the same population groups. For example, in this study many cause-specific death rates have been computed for each state. If the absolute number of deaths for two cause groups and the absolute population sizes for the states were all mutually uncorrelated, one could still obtain a nonzero correlation of the rates as pointed out by Yule and Kendall.[18] However, in this study the indexes (rates) seem more fundamental than the absolute numbers and no adjustment has been attempted for this possible source of error.

The "ecological" correlation problem, posed by Robinson,[19] relates to the fact that the correlations based on marginal population distributions in a four-fold table cannot be equated with correlations based on individuals; the former in general tend to be larger than the latter in proportion to the size of the areas from which they were determined. The answers, not original with this paper, that appear relevant to the methods used, are: 1. inferences are to be made about an ecological system and not about individuals,[20] 2. some variables have meaning only as group descriptors and must necessarily be "ecological,"[21] 3. as with the "spurious" correlations, one must ask what is the primary object of study and beware of "adjusting out" effects of importance.[21]

Another general problem of importance is the question as to what geographic areas and sizes of populations are the optimum unit of study.[21] Certainly a state is not an ecosystem in the sense that biologists apply the term. But the usefulness of the method does not depend upon the assumption that it is. Homogeneity with respect to a large proportion of the population characteristics within the units certainly would yield different levels of intercorrelations (probably lower) but it might have less to say about significant systems. In their studies, thus far, the authors have worked at two levels, the census tract[22] and the state.

Considered from an epidemiological viewpoint the analysis provides no startlingly new information nor should it be expected to. However, the manner in which the findings restate relations already known to exist establishes confidence in the validity of such new associations as may be identified. The findings are sufficient to suggest the possible value of the method as a means of accomplishing a much needed holistic epidemiological overview of man's relations with his environment. By bringing to light clusters of demographic variables associated with mortality or morbidity in general, or from a given specific cause, situations which merit further study may be identified.[23] Another epidemiologic approach is to look for different mortality variables with the same profile of factor loadings. Variables 297 (deaths from cancer of urinary organs) and 316 (deaths from ulcers of the stomach and duodenum) are concordant on the four factors of Table 1 and the three of Table 2. This is certainly not a finding that would have been expected in advance. Similarly, the location of the variable representing proportion of divorced nonwhite population (Variable 100) with a high loading on Factor 1 and of that for divorced white population (Variable 95) with a high loading on Factor 2 might have some heuristic appeal to behavioral scientists.

CONCLUSION

The findings of this study appear to have both methodological and theoretical significance. In terms of method, a holistic technique for ecological analysis is demonstrated. Given sufficiently comprehensive data, factor analysis can reveal the relationships among ecological variables as interrelated associational systems which, when viewed as a whole, should approximate a model of man's relations with his total environment. In terms of ecological theory, the potential utility of such a model lies not only in its possible contributions to the development of more systematic concepts, but also in its role as a frame of reference, or common ground, among the several disciplines concerned with studies of man and his environment.

REFERENCES

[1] Smuts, Jan Christian, Holism and Evolution, New York, Viking Press, 1961. (Published originally in 1926).

[2] Bews, J. W., Human Ecology, Oxford University Press, 1935.

[3] This conclusion now appears to be gaining acceptance as the result of several decades of experimentation in rehousing of slum populations. However, few definitive studies have been done. The most thorough one in the health area is Wilner, Daniel M., *et al.*, The Housing Environment and Family Life, Baltimore, Johns Hopkins Press, 1962.

[4] Harman, Harry H., MODERN FACTOR ANALYSIS, Chicago, University of Chicago Press, 1960, Chapters 9 and 14.

[5] Tryon, Robert C. and Bailey, Daniel E., The BC TRY Computer System of Cluster and Factor Analysis, *Multivariate Behavioral Research*, 1, 95–111, January, 1966.

[6] The data included 234 demographic variables derived from the 1960 censuses of population and housing. United States Bureau of the Census, UNITED STATES CENSUS OF POPULATION: 1960, DETAILED CHARACTERISTICS, UNITED STATES SUMMARY, Washington, United States Government Printing Office, 1963. See also ———, UNITED STATES CENSUS OF POPULATION: 1960, DETAILED CHARACTERISTICS, Final Report PC(1)-2D through PC(1)-53D, United States Government Printing Office, Washington, 1963; ———, UNITED STATES CENSUS OF HOUSING: 1960, Volume I, STATES AND SMALL AREAS, Washington, United States Government Printing Office, 1963. The data also included 138 mortality and morbidity variables based on vital statistics data for the same year. National Vital Statistics Division, VITAL STATISTICS OF THE UNITED STATES: 1960, Volume I, NATALITY, and Volume II-B, MORTALITY, Washington, United States Government Printing Office, 1963. Population densities were from Long, Luman H. (Editor), THE WORLD ALMANAC, New York, New York World-Telegram Corporation, 1962, pp. 258, 281–297. The several life-table functions were computed in the course of a separate study: Dimensions of Geographic Mortality, *Biometrics*, June, 1966. The details of raw data preparation and checking and the FORTRAN programs for computation of the demographic and mortality variables were the subject of a separate report: Final Report, Exploratory Study of Demographic Determinants of Community Health Status, Contract No. 616, March, 1964. Microfilm or Xerox copies will be supplied on request. These data will be shared with any interested investigator at the cost of copying and shipping.

The 234 demographic variables computed for the 48 states as of 1960 included (with number of levels or categories in parentheses):

Median age
Proportions—people by age (15)
    males in total population and by age (5)
    white, negro and other races
    foreign-born in total population and in foreign stock
    foreign stock by country (8)
Median years school completed
Proportions—school completion, by level (7)
    school enrollment, by level (4)
    adults in labor force, by sex (2)
    civilian employment, by sex (2)
    blue collar workers, by sex (2)
    employed, by sex (2) and occupation (9)
Median income, families and unrelated individuals
Proportions—family income, by levels (7)
    primary family households
    married couples with own household
    children under 18 with parents
    married couples by age of children (2)
    marital status categories (5), by color (2)
    persons 5 and over still in 1955 residence
Average occupants per room
Proportions—units occupied by present occupants, by period of time (4)
Average age of housing units

Proportions—housing units built, by period of time (3)
housing units sound, deteriorating, and dilapidated
Average occupants per room
Proportions—housing units, by persons-per-room (4 levels)
owner occupied housing units
housing by number of units in structure (4)
Median value of owner occupied units
Median gross rent
Crude fertility rate
Population density
Proportions—employed by sex (2) and industry (44)
urban, rural (farm and nonfarm)
born in state
Average family size
Marriage fertility (2 measures)
Median age of employed, by sex (2)
Median earnings of experienced civilian labor force, by sex (2)
Proportions—working 50–52 weeks, by sex (2)
persons 14 and over with income, by sex (2)
median income of persons 14 and over with income, by sex (2)
occupied units with telephone
Average number autos per occupied unit
Proportions—married mothers in labor force, children under 18, husband working
maternal heads of family with own children under 18 as only household member in labor force
married mothers in labor force, children under 6, husband present
The 138 mortality variables included:
neonatal mortality rates—total, and by color and urban status
crude fetal death rates—total and by color and urban status
death rates by sex (2) and age (11 groups)
crude death rates, total and by color, sex and urban status, and by sex within color
death rates from specific causes (44 cause groups)—also proportions
age-and sex-adjusted death rates
female net reproduction rate
high-school dropout rate
life expectancies at ages 0, 45 and 75 years
number living at age 25 (life-table value)
female survival probability (from birth to age 45)

[7] In later studies, alternatives to this reduction process have been found; one can, in particular take random or judgmental samples of variables, factor them, and compare the factor structures. The study reported in reference 17 used judgmental samples successfully. See also comments on redundant variables in the Discussion.

[8] The number of dimensions was set equal to the number of eigenvalues greater than 1.0 in the principal axis solution. The communalities were iteratively adjusted to fit this dimensionality before rotation of the solution.

[9] The complete tabulation will be supplied upon request.

[10] This discussion of Tryon's spherical analysis technique is based on Tryon, Robert C., Theory of SPAN and GYRO: Cluster Structure from the Spherical Configuration, *in* Theory of the BC TRY System: Statistical Theory, 1964, an unpublished work which will be part of the forthcoming book, Tryon, Robert C. and Bailey, D. E., CLUSTER AND FACTOR ANALYSIS.

A set of M independent factors can be thought of as a set of orthogonal axes with the factor coefficients serving as coordinates permitting the variables to be plotted in three-dimensional subspaces. If they are plotted as points on a generalized unit sphere when at least 80 per cent of their communality is explained by the particular trio of factors, then very little augmentation of the coefficients is required (to get the variables to the surface of the sphere). The relation of the two variables to each other on this sphere depends solely on their correlation and cannot be distorted by the particular method of factoring used.

[11] The 44 variables assigned to Factor 1 in a forced unifactor solution (Tryon's cluster structure analysis) were analyzed separately by the varimax method. Again, factors with eigenvalues greater than unity in the principal axis solution were retained for rotation.

[12] Among studies of social area analysis, a few place more emphasis on the nature of the dimensions found than on the typology of the geographic areas. Sweetser takes this point of view in his series of papers. *See* Sweetser, Frank L., Factorial Ecology, *Demography,* 2, 372–385, 1965. This paper also refers to the Tryon studies and those of Shevky and Bell, which have become classics of social area analysis.

[13] Sokal, Robert R. and Sneath, Peter H. A., PRINCIPLES OF NUMERICAL TAXONOMY, San Francisco, W. H. Freeman & Co., Publishers, 1963.

[14] Gilmour, J. S. L., Taxonomy and Philosophy, *in* Huxley, Julian (Editor), THE NEW SYSTEMATICS, Oxford, Clarendon Press, 1940, p. 466.

[15] Subsequent to the initial presentation of this paper, at the AAAS meetings in Berkeley in December, 1965, attention was called to this highly relevant study. This paper has been revised to include this important reference in proper perspective. Gouldner, Alvin W. and Petersen, Richard A., TECHNOLOGY AND THE MORAL ORDER, Indianapolis, The Bobbs-Merrill Co., Inc., 1962. Since the correlations in their data tend to be modest (very few over 0.5), the use of communalities of one could make for decisive differences in the results obtained. Such a solution is sometimes called "principal components" and is of greatest value for data reduction. Where interpretation is the goal, factoring only the common variance is a more conservative procedure. A discussion of this matter may be found in Cattell, R. B., Factor Analysis: An Introduction to Essentials, *Biometrics,* 21, 190–215 and 405–435, 1965.

[16] Hawley, Amos H., Community Power and Urban Renewal Success, *The American Journal of Sociology,* 68, 422–431, January, 1963.

[17] For purposes of intercultural comparisons the study of state data is being replicated using data from the 46 prefectures of Japan. Rogers, Edward S., Yamamoto, Mikio and Messinger, Harley B., Ecological Associations of Mortality in Japan and the United States: A Factor Analytic Study, Proceedings, Eleventh Pacific Science Congress, Volume 11, Tokyo, 1966.

[18] Actually, what is more likely to happen in this instance is that even though a pair of rates may be uncorrelated the absolute numbers involved would be positively correlated. Yule and Kendall note, ". . . the answer to the question whether the correlation between indices or that between absolute measures is misleading depends on the further question whether the indices or the absolute measures are the quantities directly determined by the causes under investigation." Yule, G. V. and Kendall, M. G., AN INTRODUCTION TO THE THEORY OF STATISTICS, 12th Edition, J. B. Lippincott Co., 1940, pp. 300–301.

[19] Robinson, W. S., Ecological Correlations and the Behavior of Individuals, *American Sociological Review,* 351–357, June, 1950.

[20] Duncan, O. D., Human Ecology and Population Studies, *in* Hauser, P. M. and Duncan, O. D. (Editors), The Study of Population, Chicago, The University of Chicago Press, 1959, pp. 678-716.

[21] Duncan, O. D., Cuzzort, R. P. and Duncan, B., Statistical Geography, Glencoe, Illinois, The Free Press, 1961, pp. 26–28, 62–68.

[22] Messinger, Harley B., *et al.,* Use of Factor Analysis and Multiple Regression in Relating Mortality to Demographic Data, Berkeley, University of California, 1964, (Unpublished.)

[23] Guerrin, Robert F. and Borgatta, Edgar F., Socio-economic and Demographic Correlates of Tuberculosis Incidence, *Milbank Memorial Fund Quarterly,* 43, 269–290, July, 1965. In this study, tuberculosis morbidity was associated with cultural deprivation, as indicated by completion of less than five years of school, even after socioeconomic status had been taken into account. The heuristic value of this might be, for example, that an attack on poverty, *per se,* should not be expected to solve the tuberculosis problem; a long-term effort to remedy what the authors call "functional illiteracy" would seem to be needed as well.

ACKNOWLEDGMENTS

This study has been supported through several sources. The assistance of the California State Department of Public Health, Bureau of Chronic Diseases, and the Computer Center of the University of California, Berkeley, is gratefully acknowledged. This project was supported by General Research Support Grant I-SO1-FR-05441 from the National Institutes of Health, U. S. Department of Health, Education, and Welfare. We are also indebted to Miss Irene Lehr for her valued assistance in various phases of this project.