SOCIO-ECONOMIC AND DEMOGRAPHIC CORRELATES OF TUBERCULOSIS INCIDENCE

ROBERT F. GUERRIN

AND

EDGAR F. BORGATTA

PREVIOUS RESEARCH

Notions that environment is involved in the epidemiology and etiology of tuberculosis have probably existed longer than has our knowledge of the exact causative agent of the disease, the tubercle bacillus. Within the relatively short time since Koch's discovery, however, such notions have persisted and, indeed, derive credibility from two sources of information. One of these is the observation and program statistics associated with the tuberculosis control activities in local health department jurisdictions and compilations of such statistics as prepared, for example, by the Public Health Service.[1] These basic age, sex, and race tabulations of mortality, morbidity, and prevalence reports have long indicated an overrepresentation of non-whites and of males, particularly in older age categories. The second source is the more direct research considerations of the relationship of environment to tuberculosis, and many of these are supported by data.

Historically, a characteristic of such studies has been that the questions posed for examination were often more sophisticated than the data and methodology available for their analysis. For example, most investigators have confined themselves to the use of registration

data, and the variable taken to represent tuberculosis has generally been a mortality figure. Mortality may have been a better measure than it is now, since the death data did tend to represent the totality of the problem before the advent of drug therapy for tuberculosis. However, the situation has changed.

A second characteristic common to the early studies was the use of a single economic indicator as the independent variable. Using data from Hamburg, Germany, for a four-year period ending in 1900, Collins demonstrated the conformity of the mean annual mortality rate from tuberculosis to a distribution of family income.[2] Using 1930 census data, Green arranged census tracts in the Cleveland five-city area into 14 economic areas based on median family income.[3] The adjusted tuberculosis mortality rates for the white population followed the same order, and the rate for the two lowest economic areas was somewhat disproportionately high.

Sydenstricker suggested that high tuberculosis rates might be associated primarily with the low end of an economic scale.[4] In one of the few studies that did not use mortality figures, he demonstrated the rate of tuberculosis discovered in the screening of a sample of garment workers in 1914 to be 12 times higher among those earning less than $700 a year than among those earning $700 and over. This assertion was supported by data from the National Health Survey of 1935 which showed that an index of disability days from tuberculosis corresponded to a scale of annual income, but was more than twice as high among those on relief as it was among those not on relief whose income was under $1,000.[5] Sydenstricker also asserted that environmental concomitants of tuberculosis were ". . . not 'conditions of life' only, but also particular conditions constituting occupational environment."[6] This consideration was demonstrated by Britten who showed the distribution of adjusted mortality rates by occupational category using 1930 data.[7] Using mortality rates for males 15–64 years of age in 10 states, Britten pointed out the correspondence of such rates to a six-point scale of occupations ranging from professional men to unskilled workers.

Among the most fascinating of these early studies—one apparently overlooked by all subsequent researchers with the exception of

Sydenstricker—was the work of Gray.[8] Using 1920 Census data she computed Spearman correlation coefficients between six variables and tuberculosis mortality for the 10 original registration states. Although her correlations were based on state data for only 10 states, her selection of variables was sophisticated and her findings were interesting. Strong positive correlations were presented between tuberculosis mortality and the following variables: per cent urban population (.85); per cent illiterates (.76); per cent paying income tax (.74); mean per capita income (.52); per cent children employed (.71); per cent females employed (.56). With the exception of the income tax variables, these findings suggest directions for research which have been generally ignored. The positive relationships between the income tax variables and tuberculosis mortality deserve comment because they defied explanation by the author and were unexpected in the sense that these variables were taken to represent economic status. However, it is likely that, for 1920 state data, income tax variables were more a measure of urbanization than of economic status.

It was not until 1948 that serious thought was given to the idea that single economic indicators may represent a range of underlying conditions of importance in discussing the social etiology of tuberculosis. Terris [9] made the point: "It is clear that several causes may be involved—overcrowding, bad housing, poor nutrition, overwork, increased exposure, etc. It is extremely difficult to separate these factors. . . ."[10] Using 1940 data for Buffalo, New York, Terris constructed an economic status index for census tracts by combining four indicators: median monthly rent; per cent of houses with central heating; per cent of houses with mechanical refrigeration; and median school years completed. He then assigned the census tracts to quartiles based on this index and attempted to investigate resident tuberculosis deaths for each quartile for a three-year period, classified by age, sex, and color. Since only 3 per cent of the total population was non-white, the color classification by tract quartiles was not possible. Although no comment is made on the point, the mortality rates for nonwhites based on total population were about five times greater than those for whites. While Terris' data did not

constitute a direct test of his original hypothesis, he was able to demonstrate that, despite the clear relationship between his mortality data and the economic quartiles, consistent age and sex differences occurred within the quartiles, indicating the insufficiency of the economic indicator by itself to account for differences in mortality rates.

While the use of mortality data in these studies can be excused on the grounds that they were the only generally available data, and while a somewhat tenable case can be made for their representation of tuberculosis prior to drug therapy, it was not until 1960 that an important methodological consideration in the use of such data was proposed. Terris and Monk hypothesized that if place of residence at death was used as the basis for computation of a mortality rate, correlations of such data with economic indicators might indicate the effects, rather than an etiological component, of the disease.[11] In a partial replication of the previous study, Terris and Monk examined mortality data from 1949 to 1951 for Buffalo and traced residence locations in an attempt to measure socio-economic drift. Although their findings failed to support the hypothesis in any substantial way, a reasonable question had been raised as to the use of raw mortality data.

The advent of drug therapy for tuberculosis caused a significant change in mortality rates, further distinguishing mortality data from alternative measures of tuberculosis. The development of tuberculosis case registers to implement tuberculosis control programs made the use of alternative measures of tuberculosis more feasible. In his analysis of 1950 New York City data, Lowell was able to use a prevalence rate.[12] Using a limited number of variables such as median family income, condition of housing, overcrowding, unemployment, juvenile delinquency, and race, Lowell demonstrated that all of these were to some degree associated with his prevalence rate. A limited application of correlation analysis, however, also demonstrated the interrelationship at significant levels of his three independent variables: housing, income, and race.

A recent analysis by Lebowitz and Malcolm is of interest, although tuberculosis was not the focal variable.[13] In this study 15 population

and health variables were factor-analyzed in an attempt to produce an index of socio-economic status for Alameda County, California. Factor analysis is a numerical procedure for getting at "underlying" variables. Tuberculosis appears at a meaningful level on two of the four orthogonally rotated factors. One of these is taken by the authors to represent the socio-economic index, but the principal association of the tuberculosis variable is with another factor for which it constitutes the primary definition. It is clear, therefore, that a great deal of the variance associated with their measure of tuberculosis morbidity is unaccounted for by the variables comprising the socio-economic status index.

If epidemiological investigations of the relationship between tuberculosis and population characteristics have tended to confuse the issue rather than to clarify it, investigations of the context in which such relationships occur—the description of urban areas—have not entirely satisfied demographers and sociologists, as a recent dialogue indicates.[14] The purpose of this investigation, therefore, is to present an adequate general description of an urban area based on census tract units, and to examine tuberculosis morbidity as the focal variable within this context.

METHOD

Of major importance is the development of a measure of tuberculosis in the community. Clearly, mortality data have become less relevant in the investigation of social etiological factors. Both prevalence and incidence registration data are subject to the speculation that not all of the existing cases in the community are known. By means of a massive x-ray survey, Anderson, Enterline, and Turner attempted to estimate the number of undetected cases according to economic areas in Cleveland.[15] They reported only a small difference between their estimate of undetected cases and cases known to the health department within the economic arrangement of census tracts. The reported differences, in fact, do not satisfy a test of statistical significance, although the authors themselves did not apply such a test. There appears to be little real evidence to support

a contention—and some evidence to the contrary—that detected cases in a community do not accurately represent the real situation. This is probably true in spite of the fact that prevalence data are subject to the problem of socio-economic drift and that they may also reflect differential response to treatment, since this is ultimately the criterion for removal from case registers. However, it is chiefly because of the interest in the problems of social etiology that a measure of incidence rather than prevalence was chosen for this research.

Like other registration data, the sources of data on morbidity are often limited. In this case we viewed the determination of tuberculosis morbidity data as a major task of the research. In order to minimize any effects of case-detection activities on these data, several criteria for the inclusion of cases in the research were established. Because of the use of 1960 Census data, we were interested in tuberculosis cases which coincided with these data. The time span for which cases were considered was broadened to include all newly reported cases from January 1959 through September 1961. The total population of the District of Columbia at the time of the census was 764,000 and the three-year annual average tuberculosis case rate was 70.5 (per 100,000). All cases were screened on the basis of diagnosis at the time of the original report. Excluded from the analysis were all diagnoses of inactive tuberculosis. In addition, cases of primary tuberculosis were excluded because separate criteria are used to classify activity status of such cases and these criteria involve an arbitrary age cut-off.

Although the use of old, rather than current, morbidity data involves some search problems, it has the advantage of providing clarifications of original reports. Thus we were able to apply the criteria for inclusion to all cases originally reported with undetermined or unknown activity status. The search was aided by the availability of the original report documents. These documents, from physicians, hospitals, health department clinics, and other sources, are the basis for official tabulations of new cases and inclusion of cases in the tuberculosis case register. This source was cross-checked with the official listing of new cases for the period,

which was provided by the Biostatistics Division of the Health Department. All cases identified by this procedure were then searched for in the case register so that diagnostic information could be verified and residence at time of original report could be verified and coded by census tract. A total of 1,480 cases satisfied the criteria and were included in the analysis.

In the analysis, a choice had to be made between adjusting the measure of tuberculosis morbidity to take into account the size of the census tract, and using the raw count of cases. Since census tracts tend to be defined by size, the adjustment for size of census tract would tend to be meaningless in cases where no—or only one or two—cases of tuberculosis were reported, and this is the most common situation. From the point of view of tuberculosis morbidity, the distinction between tracts tends to be one of no occurrence or of occurrence in some degree. Since the reported values of occurrence are discrete, the correction was anticipated to have a trivial effect. For these reasons, the raw data, uncorrected for size of tract, were used directly as the measure of tuberculosis morbidity in the analysis, and it was decided to verify the accuracy of this assumption afterward.

A second registration variable, juvenile delinquency, is included in the census tract analysis for the District of Columbia. These data were coded from juvenile court records and made available to the analysis by Charles Willie, Ph.D., who was then associated with the Washington Action for Youth Project.

The determination of registration data by tract is of importance because such data for small geographical units are usually not available. On the other hand, the data apply only to the District of Columbia, and not to the larger metropolitan area. In a metropolitan area such as this one, the difference between the urban center and the suburban ring may be large. Of particular note has been the apparent movement of the white population to suburban areas and the corresponding increase of the proportion of nonwhites within the District.

The District of Columbia contains 125 census tracts. Three of them are entirely occupied by institutions and another two are en-

tirely composed of Federal park lands. These five census tracts were excluded from the analysis, leaving a total of 120 census tracts for the District of Columbia. From the available tract data, a selected number of variables were chosen for analysis in conjunction with the variable of particular interest, tuberculosis morbidity. The selection was guided in part by the data available, but also by the information provided from a general study of tract data in process (1964) at the University of Wisconsin. At the time the current analysis was undertaken, a relatively final list of variables that would encompass most of the relationships among the entire panel of variables provided for tract data by Census publications had been compiled at Wisconsin. Our current analysis includes 28 of the 31 variables. Thus our analysis substantially includes the same variables as those used in the University of Wisconsin studies, and this makes feasible a comparison of the District of Columbia with the Standard Metropolitan Statistical Area (SMSA) of Washington. We note, of course, that the two registration variables used—tuberculosis morbidity and delinquency rate—are not available for the SMSA analysis. An additional unusable variable was the distinction between the politically defined city and its ring of suburbs. Data of the two analyses are presented in parallel in order to permit the comparison of the structure of relationships. The variables for the District of Columbia analysis were as follows:

District of Columbia Census Tract Variables

1. Per cent Negro
2. Per cent foreign born
3. Per cent other races
4. Male separation rate
5. Male divorce rate
6. Male widower rate
7. Median age
8. Per cent females under 5 years
9. Fertility ratio
10. Per capita in high school
11. Median school years completed

12. Per cent completed less than 5 years school
13. Per cent college graduates
14. Per cent males 14+ in labor force
15. Per cent males in labor force unemployed
16. Per cent females 14+ in labor force
17. Per capita males prof. & managerial
18. Median income of families
19. Per cent families $10,000 income +
20. Per cent in same house 1960 as 1955
21. Population per household
22. Per cent owner-occupied units
23. Per cent units sound
24. Per cent units built 1950 or later
25. Median value of owner-occupied units
26. Median monthly rent
27. Per cent single dwelling units
28. Per cent moved into units 1958–60
29. Tuberculosis morbidity
30. Delinquency rate

The basic method of analysis used here is correlational. "Image" factor analysis, retaining eight factors for varimax rotation, is used, but reference to the generating correlation matrix is seen as useful. The matrix of correlations is presented as Table 1. The number of factors used in rotation was determined by the experience at Wisconsin with sets of tract data, and in this case the first eight factors included 96 per cent of the total variance in the correlation matrix. In our analysis, interpretation is generally restricted to loadings of .30 or greater. While correlations of substantially smaller magnitude would satisfy ordinary statistical hypothesis tests, our concern here is to emphasize relationships of sufficient magnitude to be interpretively meaningful and, hopefully, that might recur in other cities.

FINDINGS

Factor I. Socio-economic Status

The first factor is here called *socio-economic status,* and involves a substantial number of the variables in the analysis. The loadings

TABLE I. INTERCORRELATION MATRIX, DISTRICT OF COLUMBIA CENSUS TRACT VARIABLES

N = 120 n = 30

Variable	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30
1	90	−76	−30	57	−24	17	−60	61	65	18	−73	66	−66	−21	52	−10	−78	−66	−67	03	64	−03	−33	−19	−63	−48	19	−18	52	32
2		86	44	−41	25	01	73	−73	−64	−13	68	−50	70	13	−45	05	80	73	75	20	−59	17	23	05	70	56	−07	−07	−33	−23
3			77	08	41	16	20	−24	−18	−23	04	08	06	10	12	−35	11	06	07	−11	−29	−17	−30	−07	17	−05	−17	20	11	21
4				90	42	70	−10	09	21	−40	−68	77	−49	−20	66	14	−53	−65	−60	−11	−01	−45	−81	−46	−33	−57	−24	06	68	75
5					88	62	45	−50	−45	−66	08	05	15	08	21	38	12	−13	−08	−14	−65	−47	−45	−29	34	−24	−48	21	11	46
6						88	37	−35	−10	−44	−33	52	−06	−32	44	18	−11	−25	−17	11	−32	−33	−67	−45	09	−27	−24	−10	45	71
7							95	−91	−71	−27	55	−35	68	03	−33	39	71	59	06	42	−78	08	05	−19	70	42	−20	−27	−15	01
8								96	78	28	−59	38	−72	10	36	−40	−74	−61	−65	−42	76	−13	−03	24	−71	−44	09	30	20	01
9									92	48	−67	54	−62	−15	47	−56	−64	−46	−45	−07	86	17	−21	12	−57	−24	41	−05	29	04
10										80	−07	−06	−07	−20	−12	−48	−04	20	18	44	62	68	35	19	−09	26	66	−45	−20	−51
11											96	−86	88	19	−69	29	84	77	78	13	−52	24	62	24	69	67	−02	−01	−64	−47
12												91	−64	−32	72	−28	−64	−62	−60	−09	35	−24	−73	−32	−46	−53	01	−01	75	58
13													97	−04	−55	21	92	82	87	26	−55	25	39	17	83	76	02	−11	−48	−34
14														71	−12	23	09	03	−01	−23	−17	−22	19	25	−01	−08	−30	37	−10	−07
15															73	−18	−58	−03	−57	−18	25	−36	−64	−21	−37	−48	−16	13	52	48
16																86	18	−03	−02	03	−54	−23	07	−21	16	−03	−41	−02	−04	19
17																	96	86	90	29	−57	27	37	24	83	75	−01	−11	−48	−38
18																		97	96	43	−28	54	50	17	75	83	30	−33	−50	−53
19																			98	42	−35	49	45	19	80	84	24	−28	−47	−47
20																				85	04	62	10	−23	30	33	44	−86	−05	−21
21																					96	40	06	10	−57	−12	64	−20	13	−21
22																						92	40	01	18	54	85	−69	−26	−53
23																							83	41	18	48	23	−12	−63	−64
24																								60	09	28	−05	33	−35	−31
25																									88	62	−09	−14	−34	−22
26																										84	36	−26	−44	−49
27																											93	−54	−12	−41
28																												90	−01	24
29																													69	62
30																														80

Note: Decimal points are omitted. The square of the multiple regression coefficient is shown in the main diagonal. Values of r ≥ .18 will satisfy a symmetric test of significance at the .05 level with 118 d.f.

also tend to be relatively large. This is so in spite of the reported prior screening of variables in order to minimize duplication. That this is the case, however, is not surprising. Communities *are* determined in large part, with regard to character, by the general economic status of families and all that this implies. While there is no assertion that tract areas are homogeneous, the delineation of tracts is supposed to follow a principle of grouping that places similar units in the same tracts.

The principal defining variables are 19, "Per cent families $10,000 income +," 13, "Per cent college graduates," 25, "Median value of owner-occupied units," and 18, "Median income of families." Variable 17, "Per capita males prof. & managerial," could be listed with this group, but since the magnitude in the two analyses differs, it is noted separately. The professional and managerial categories are apparently more closely associated with the high end of the socio-economic continuum within the District of Columbia than in the entire SMSA. Possibly, this is because the District of Columbia tends to include more distinct and extreme cases in the contiuum, while in the additional suburban areas there are more examples of intermediate tracts. Other possible explanations are noted below in considering other variables.

Variable 7, "Median age," has a higher loading for the data in the District of Columbia than in the SMSA. This corresponds to a notion that there are more distinct areas with high proportions of professional and managerial persons, and these would generally be older in the sense that maturity is involved in reaching the higher levels of status. More "completed families" would be the implication. Additional differential loadings add to the plausibility of this interpretation. Variable 2, "Per cent foreign born," has a higher loading within the District of Columbia and this could be the association of older age with the possibility of being an immigrant. Generally, younger white families moving to the suburbs are likely to include older and second-generation Americans, while the older, completed families are less likely to move. The latter may, in fact, concentrate in better housing in the center of the urban area. In these data, rental value (variable 26, "Median monthly rent") is

FACTOR TABLE I

Variable		*Loadings* D.C.	SMSA
19	Per cent families $10,000 income +	91	90
17	Per capita males prof. & managerial	91	80
13	Per cent college graduates	90	93
25	Median value of owner-occupied units	86	84
18	Median income of families	84	84
26	Median monthly rent	78	69
2	Per cent foreign born	76	43
7	Median age	72	47
11	Median school years completed	71	75
8	Per cent females under 5 years	−71	−46
1	Per cent Negro	−65	−40
21	Population per household	−51	−19
9	Fertility ratio	−50	−17
12	Per cent completed less than 5 years school	−43	−44
15	Per cent males in labor force unemployed	−40	−35
4	Male separation rate	−35	−31
29	Tuberculosis morbidity	−30	NI
23	Per cent units sound	20	36
10	Per capita in high school	−07	34
X	Per cent families under $3,000 income	NI	−40

NI = Not included in the analysis
Decimal points omitted

highly correlated with the factor, and the association is somewhat stronger within the District of Columbia than in the whole of the SMSA. The interpretation of completed families is aided by the negative correlation of variable 8, "Per cent females under 5 years," which is more strongly associated with the factor in the District of Columbia than in the SMSA data.

The negative loading of variable 1, "Per cent Negro," is expected, of course, but the fact that the loading is higher within the District of Columbia than in the SMSA data suggests the more distinct comparison of white, high socio-economic status neighborhoods in contrast to low status, colored neighborhoods. The additional variables in the factor provide more descriptive material for the concomitants of the factor, but essentially they add little to the description since the loadings are progressively smaller.

Of some interest is the loading of the focal variable for this study, 29, "Tuberculosis morbidity." The loading is −.30, indicating a small negative association between tracts of high socio-economic status and tuberculosis morbidity, or a small positive association between tracts of low socio-economic status and tuberculosis morbidity. Examining the direct relationship to tuberculosis morbidity of a pivotal variable for the factor, such as 19, "Per cent families $10,000 income +," the magnitude of the relationship is seen to be higher, −.47. The meaning of this will become apparent as we discuss the second factor, which is sometimes related to notions of socio-economic status.

Factor II. Cultural Deprivation

The second factor is here provisionally called *cultural deprivation.* That this is not the usual socio-economic status factor is clear, since such a factor has already been accounted for in the analysis. In the SMSA data, the variable "Per cent families under $3,000 income" was included, and it occurs in the second factor. Economic considerations thus are involved in the second factor, but in the sense of the percentage of poor or very poor families in the community rather than those with average income or the percentage of persons with high incomes. Income in this sense appears to involve connotations of not being even in the usual range of consideration of socio-economic status, but being almost qualitatively separate. In the parallel analyses, the variable that appears to have relatively the highest loading in both analyses is 12, "Per cent completed less than 5 years school," usually taken as the best indicator of "functional illiteracy" among the variables available in published Census data. Additional variables with high correlations for the parallel analyses are 4, "Male separation rate," 23, "Per cent units sound," and 6, "Male widower rate." These variables suggest, possibly, substandard homes and deprived living conditions or "social disorganization." Variable 30, "Delinquency rate," which was available for District of Columbia data, adds to this descriptive base. However, the notion that deterioration is involved is not suggested here. Possibly physical plant and neighborhood housing are deteriorated,

FACTOR TABLE II

Variable		*Loadings* D.C.	SMSA
4	Male separation rate	85	77
23	Per cent units sound	−82	−73
6	Male widower rate	80	73
12	Per cent completed less than 5 years school	79	80
30	Delinquency rate	76	NI
29	Tuberculosis morbidity	73	NI
15	Per cent males in labor force unemployed	63	64
11	Median school years completed	−57	−59
1	Per cent Negro	41	70
5	Male divorce rate	41	31
24	Per cent units built 1950 or later	−39	−45
18	Median income of families	−38	−38
10	Per capita in high school	−36	−13
22	Per cent owner occupied units	−33	−31
26	Median monthly rent	−32	−36
19	Per cent families $10,000 income +	−30	−26
14	Per cent males 14+ in labor force	−22	−55
X	Per cent families under $3,000 income	NI	82
X	Per capita married males	NI	−49
X	Central city vs. urban ring	NI	−45

NI = Not included in the analysis
Decimal points omitted

but there is no implication that the families involved are moving up or down the socio-economic or cultural ladder. Current illiteracy in the adult population is hardly the result of deterioration. Other variables indicate cultural deprivation and, possibly, inability to enter the general socio-economic system. These include variables 15, "Per cent males in labor force unemployed," 11, "Median school years completed," and others.

It is appropriate at this point to focus on variable 29, "Tuberculosis morbidity." This variable has a large loading in the factor. In considering the *socio-economic status* factor we suggested that some explanation was required to clarify the relationship of tuberculosis morbidity to the type of tract involved. Here, let us look specifically at the relationships involved. We shall use the primary defining variables in the first two factors; in particular, we shall use variable 19, "Per cent families $10,000 income +," to represent the

socio-economic status factor, and variable 12, "Per cent completed less than 5 years school," to represent the *cultural deprivation* factor. These are the items with the most generally high loadings in the two analyses, and they also are the more conceptually independent variables. For example, variable 13 could be used instead of 19 in this analysis, but the educational implication might be viewed as overlapping that of variable 12 directly. Actually, however, the substitution would in no way affect the subsequent analysis here. One conceptual difference between income and education as indicators of socio-economic status should be noted. While current income may reflect prior conditions, a statement of education completed is more or less directly a statement of prior conditions. In this sense, indicators of adult educational status may reflect prior history better than those of income.

The correlation between variables 19 and 29 (high socio-economic status and tuberculosis morbidity) within the District of Columbia was found to be – .47. Such a negative correlation is commonly expected to occur and has been indicated by others.[2, 3, 5, 6, 9, 10, 12] The relationship between variables 12 and 29 (high illiteracy and tuberculosis morbidity) was found to be .75. But we know from the factor analysis that socio-economic status and illiteracy tend to occur in different factors. The actual correlation in these data is still substantial (– .60). Thus the question is raised of what happens to the relationship of each of these variables to tuberculosis morbidity when the other is taken into account. This is a matter of taking simple partial correlations. The correlation between high socio-economic status and tuberculosis morbidity decreases to – .04, a trivial relationship, when illiteracy is taken into account. The correlation between illiteracy and tuberculosis morbidity remains relatively high (.66) when high socio-economic status is taken into account, suggesting the relative importance of the *cultural deprivation* factor, independently of *socio-economic status,* as a correlate of tuberculosis morbidity. That functional illiteracy in a community is a good indicator is quite consistent with the notion that tuberculosis is not an immediately contagious disease, but one that arises out of a history of poor conditions and exposure.

Since this is the important finding of this report, it is appropriate to remind the reader that our measure of tuberculosis morbidity has been the raw data report. At this point it becomes mandatory to test the assumption that the size of population in the tract would have little effect on the analysis. Thus, the data were further "partialled" to take into account not only socio-economic status but also population size, and the result was similar (.69).

The association between Negro population and tuberculosis morbidity is also well documented.[1,9,12] In the present study, however, the loading in the second factor is seen to be relatively small. It was noted that the loading of variable 1, "Per cent Negro," was higher in the District of Columbia data than in the parallel data for the SMSA in the first factor, and the reverse is true in the second. The correlation between the proportion Negro and tuberculosis morbidity is .56. When illiteracy is taken into account, the correlation between per cent Negro and tuberculosis morbidity becomes .05, a trivial relationship. On the other hand, the relationship between illiteracy and tuberculosis morbidity remains substantial (.63) when the proportion of Negroes in the tract is taken into account.

Additional Parallel Factors

The additional parallel factors are presented here, but are not discussed in detail. They are of some interest because they indicate some of the kinds of characteristics of tracts that are *not* involved as correlates of tuberculosis morbidity. Factor III is called *age-sex family structure*. It will be noted that in the SMSA data there is a positive loading for the variable "central city vs. urban ring." In the factor there is some association of recent construction and single dwelling units with indicators of fertility, females present in the household, larger families in the household, and, generally, younger children and younger age.

Factor IV is provisionally called the *mobility* factor and is defined largely by the variables indicating the amount of residence in the same household. Associated with this is ownership of home and, concomitantly, dwelling in single family units.

The fifth factor tends to be relatively weak in definition, and ap-

FACTOR TABLE III

Variable		*Loadings* D.C.	*SMSA*
16	Per cent females 14+ in labor force	−77	−80
9	Fertility ratio	76	82
21	Population per household	72	88
10	Per capita in high school	60	40
8	Per cent females under 5 years	58	68
5	Male divorce rate	−57	−56
7	Median age	−52	−76
27	Per cent single dwelling units	46	68
24	Per cent units built 1950 or later	41	58
22	Per cent owner-occupied units	32	56
6	Male widower rate	−31	−44
2	Per cent foreign born	27	−52
29	Median monthly rent	20	30
X	Per cent females 65 years +	NI	−76
X	Central city vs. urban ring	NI	49

NI = Not included in the analysis
Decimal points omitted

FACTOR TABLE IV

Variable		*Loadings* D.C.	*SMSA*
28	Per cent moved into units 1958–60	−92	−92
20	Per cent in same house 1960 as 1955	87	90
22	Per cent owner-occupied units	70	64
27	Per cent single dwelling units	56	61
10	Per capita in high school	52	57
24	Per cent units built 1950 or later	−34	−27
14	Per cent males 14+ in labor force	−23	−37
21	Population per household	24	30

Decimal points omitted

FACTOR TABLE V

Variable		*Loadings* D.C.	*SMSA*
3	Per cent other races	80	76
16	Per cent females 14+ in labor force	−47	−24
2	Per cent foreign born	42	50
1	Per cent Negro	−35	−18
5	Male divorce rate	26	32

Decimal points omitted

pears to indicate that members of "other races" and the foreign born tend to be found in tracts where there is a relative absence of Negroes. There are no socio-economic implications to this factor except, possibly, the relative absence of females in the labor force.

The additional nonparallel factors are not discussed here, but it is noted that in examining the relationships that occur at a relatively low level, one was uncovered which may warrant additional examination in other sets of data. In particular, the relationship between divorce and tuberculosis morbidity in tracts is noted, but at a level of relationship that would normally not satisfy statistical hypothesis test requirements (.11). However, subsequent examination with partial correlations indicates that it is relatively impervious to explanation on the basis of the illiteracy status of the tract.

SUMMARY AND CONCLUSIONS

Tuberculosis morbidity data were compiled for the District of Columbia according to census tracts. A large mass of additional tract data, available from United States Census publications, was examined and reduced to a set of representative and relatively independent measures of census tract characteristics. A factor analytic study was carried out to determine the factors underlying the distribution of tracts, and the relationship of the characteristics of the tracts to tuberculosis morbidity. Data from another study (at the University of Wisconsin) of tracts from the entire Standard Metropolitan Statistical Area of Washington were available for parallel analysis. Five factors were retained for rotation and analysis. Two of these, called *socio-economic status* and *cultural deprivation,* involve the focal variable of tuberculosis morbidity. Examination of the first order correlation matrix and subsequent computation of partial correlations suggests that the crucial concomitant of tuberculosis morbidity is *cultural deprivation,* and most particularly is it identified with the literacy level of the tract. Virtually all of the association of tuberculosis morbidity with economic indicators is accounted for when literacy level is controlled, but the reverse is not true. This finding

warrants speculation that involves a notion that the conditions which underlie functional illiteracy are also related to tuberculosis.

Examination of the additional factors is suggestive of only one additional association that might be worth further exploration as a correlate of tuberculosis morbidity. A very small relationship between divorce and tuberculosis morbidity is not explained by the illiteracy rate of the tract. The additional three parallel factors involve concepts of mobility, family status, and distribution of racial characteristics.

These findings suggest implications for programs of tuberculosis control and certainly for the more recent notion of eradication of tuberculosis. While correlations exist between tuberculosis morbidity and ordinary notions of socio-economic status, separate and more stable relationships were shown to underlie these commonly expected correlations. The variables involved tend to indicate the relative importance of prior history, rather than current conditions, and, thus, would imply the necessity for public health tuberculosis programs to be concerned with those conditions which may be affecting future case rates.

The need for further research of several kinds is indicated by this analysis. First, Washington, D. C., may present a relatively unique situation. The changing complexion of the city is immediately evident to any who are familiar with it. Within the District there has been construction of many housing developments and slum removal. Shifts of population have been noted, primarily involving the outmigration of whites to the suburbs, leaving high proportions of Negroes within the District. Every major city, in a sense, is relatively unique. New York City has not only concentrations of Negroes, but also the immigration of Puerto Ricans. Los Angeles has large numbers of Spanish-speaking peoples and is characterized by recent and rapid growth. Therefore, any study of urban problems must be viewed as essentially a case description—no matter how objective the measures used—and additional cases must be sought.

Secondly, important hypotheses relative to social factors may be advanced. Studies like the one reported here cannot provide final answers; these will have to come from more complex longitudinal

studies of individuals, that are as yet virtually unknown to our science. It is important to emphasize that ecological correlations may be suggestive of individual correlations, but the two must not be mistaken for each other. An important substitute, in the meantime, is the attempt to abstract from ecological correlations hypotheses relevant to etiological considerations and, further, the examination of alternate sets of data to examine them. The investigation of additional health conditions is important. The infant mortality rate, for example, may be viewed as an indicator of a present state of poor conditions with a dual time reference: one dealing with the past conditions that made for poor health of the mother, and the second dealing with social and health conditions that may more immediately affect the probability of survival for the infant. The venereal disease rate—as an indicator of other, current types of social disorganization, with emphasis on their impact on young adults—would be a similarly interesting variable. Further exploration, therefore, appears warranted in drawing and analyzing additional morbidity and mortality data by tracts.

The third indication for further research is methodological and deals with the appropriateness of the tract as a unit of analysis. In some communities, for example, larger units are used for analysis of data. In New York City, which keeps excellent records, the health area is the unit. Health areas are generally composed of several adjacent census tracts, picked for homogeneity, and with a population of about 25,000. The District of Columbia is beginning to report data according to 18 similarly constructed statistical areas. If judging homogeneity of tracts is difficult, the criticism is more serious for the assumption in larger units. Hence exploration should be carried out on the impact of using larger and smaller units of reporting. Obviously, the use of smaller units also presents difficulties. On the one hand, the ultimate objective might be reduction to the single family as a unit. This would be a different type of study from one of ecological characteristics, as reported here, but the potential for moving in this direction should not be ignored. For example, if the Census Bureau develops more trust in the integrity of scientists dealing with social data, or if it makes possible the use of such data by

scientists concurrently with the passage of laws that would make it a crime to reveal information, the data might be co-ordinated with other reported and registration data in such a way as to lead to tremendous possibilities for new research.

In a lesser but equally important way, enumeration districts, which are subdivisions of tracts, may be used. The limitation is that there are few cases for an entity that has a low probability in the population. Thus the data are subject to sampling errors associated with small sizes, the most obvious of which is the fact that a few cases make a lot of difference in the relative statistic for the enumeration district. In spite of this, the analysis appears warranted if only to test the relevant questions of the limitations of homogeneity in the tract in relation to the limitations of size of the smaller units.

REFERENCES

[1] Tuberculosis Program, Tuberculosis in the United States: Status of the Disease in the Early Sixties, United States Public Health Service, Publication No. 1036, May, 1963.

[2] Collins, Selwyn D., Economic Status and Health, Public Health Bulletin 165, Washington, D.C., United States Public Health Service, 1926.

[3] Green, Howard W., TUBERCULOSIS AND ECONOMIC STRATA: CLEVELAND'S FIVE-CITY AREA, 1928–31, Anti-Tuberculosis League of Cleveland, 1932.

[4] Sydenstricker, Edgar, HEALTH AND ENVIRONMENT, New York, McGraw-Hill Book Co., Inc., 1933.

[5] National Health Survey, 1935–36, Disability from Specific Causes in Relation to Economic Status, Preliminary Reports, Sickness and Medical Care Series, Bulletin 9, Washington, D.C., United States Public Health Service, National Institutes of Health, 1938.

[6] Sydenstricker, Edgar, *op. cit.*, p. 145.

[7] Britten, Rollo H., Mortality Rates by Occupational Class in the United States, *Public Health Reports,* 49, 1101–1111, September 21, 1934.

[8] Gray, Cora E., Tuberculosis Mortality in the Original Death-Registration States: A Statistical Study of the Death Rates from 1900 to 1924 and of the Influence of Certain Factors upon Them, *American Review of Tuberculosis,* 18, 687–719, November, 1928.

[9] Terris, Milton, Relation of Economic Status to Tuberculosis Mortality by Age and Sex, *American Journal of Public Health,* 38, 1061–1070, August, 1948.

[10] *Ibid.,* p. 1062.

[11] Terris, Milton, and Monk, Mary A., The Validity of Socio-economic Differentials in Tuberculosis Mortality, *American Review of Respiratory Diseases,* 81, 513–517, April, 1960.

[12] Lowell, Anthony M., SOCIO-ECONOMIC CONDITIONS AND TUBERCULOSIS PREVALENCE: NEW YORK CITY, New York, New York Tuberculosis and Health Association, 1956.

[13] Lebowitz, Michael D., and Malcolm, James C., Socioeconomic Analysis of the Alameda County Health Department Jurisdiction, *American Journal of Public Health,* 54, 1876–1881, November, 1964.

[14] Duncan, Beverly, Devolution of an Empirical Generalization, *American Sociological Review,* 29, 855–862, December, 1964. *See also* Goldstein, Sidney and Mayer, Kurt, Comment on Duncan, *ibid.,* p. 863; Schmid, Calvin F., Van Arsdol, Maurice, and MacCannell, Earle H., Comment on Duncan, *ibid.,* p. 866.

[15] Anderson, Robert J., Enterline, Philip E., and Turner, Otis D., Undetected Tuberculosis in Various Economic Groups, *American Review of Tuberculosis,* 70, 593–600, October, 1954.

ACKNOWLEDGMENT

This investigation was supported by research grant R&S-7a-9 from the National Tuberculosis Association. We also appreciate the co-operation of the District of Columbia Department of Public Health.