The need to identify geographical areas of highest priority for health and social services to families led to a study of the availability of health and social indexes in San Francisco, and a comparison of two methods (factor analysis and map plotting) for determining their usefulness. The most useful health and socioeconomic indexes are discussed. Means of supplementing indexes derived from the decennial census are also dealt with.

AVAILABILITY AND USEFULNESS OF SELECTED HEALTH AND SOCIOECONOMIC DATA FOR COMMUNITY PLANNING

Helen M. Wallace, M.D., F.A.P.H.A.; Victor Eisner, M.D., F.A.P.H.A.; and Samuel Dooley, M.D., F.A.P.H.A.

THE continuing in-migration to the large metropolitan centers of large numbers of low-income families in need of health and social services means that agencies serving the needs of these families must be able to identify geographical areas of highest priority for such services. Identification of these highest priority areas is essential to effective planning in order to bring necessary health and social services to families with the greatest health and social needs. The concept of "high risk women or families" is inherent in the development of new Maternity and Infant Care Projects, and of new comprehensive services for children of school and preschool age. Both of these programs accept the principle of concentrating services in "high risk areas" in which there are large concentrations of "high risk families."

The need for rapid identification of high risk areas raises several questions, centered around the use of census tract units in urban areas as the focus of study and planning.

1. What health and social indexes are readily available to local health departments in large metropolitan centers, as a first step in planning health services for mothers and children?
2. Are there health and social indexes among those usually available that are most useful and practicable for identifying needs and planning services?
3. What methodology is useful in the analysis and use of available indexes for planning purposes?

This paper will attempt to provide answers to these questions, using data available from the city and county of San Francisco for the years 1958-1962. San Francisco as the twelfth largest city in the United States had a total population of 740,316 in 1960.

Methods

A group of arbitrarily selected indexes were chosen for study. These com-

prised a large variety of measures of health and of socioeconomic status which in the opinion of the authors had been useful in the past or might prove to be useful in the future for program planning purposes. An attempt was made to obtain data for calculating these indexes for each of the 94 census tracts of the city and county of San Francisco which had a population of at least 500 under 18 years of age in 1960.

Each index was then placed into one of four groups:

a. "Available." The necessary data could be obtained from existing records.

b. "Available, numbers too small." The necessary data could be obtained from existing records, but the number of events was too small to allow calculation of indexes for more than an occasional census tract.

c. "Not available by census tract." The necessary data could be obtained from existing records, but was neither coded by census tract nor identified by street address.

d. "Not available." The necessary data could not be obtained from existing records.

All of the available indexes were then computed. Definitions of these indexes are given in Appendix A. Each index was derived in a manner allowing fairly rapid computation: thus, proportions, e.g., of families with incomes below $3,000 per year, were used instead of medians. In regard to venereal disease, an index was computed to show the proportion of nonprimary syphilis cases (secondary and tertiary) out of the total reported cases.

For statistical reasons, when the census tract data were secured, they were arranged in rank order; correlation coefficients were then computed on the rank orders.

The available indexes were then judged for "usefulness" by the following arbitrary criteria:

a. The existence of a high index in a census tract was generally associated with other high indexes in the same tract. It was assumed that the existence of several high indexes pointed to a need for medical or social services.

b. The statistics appeared to be reliable, from what was known of the original data collection and from the size of the numbers used to derive the rates in each census tract.

Two methods were used to judge the first criterion:

a. *Census Tract Map Method*—This consisted of plotting the indexes by quartiles on census tract maps, and studying the maps. Five independent map studies were made by the three authors, and an index was judged "useful" by this criterion if it was so judged on at least four of the five studies. This is, of course, an extension of the method traditionally available to most health departments for identification and location of problem areas in a community.

b. *Factor Analysis*—This is a mathematical treatment of correlation coefficients which results in grouping the indexes into a number of "factors." Each factor accounts for a certain percentage of the variance between the indexes, and is composed of all of the indexes, with varying weights assigned to each index. For the purpose of this study, it may be assumed that a weight or loading of 0.5 or more for an index in a given factor means that that index is an important component of the factor. It may further be assumed that the factor with high loadings for the largest number of health and social indexes represents a factor for "high risk" in the sense we are using it in this paper. The "high risk" factor in this study accounted for 43.5 per cent of the total variance of all of the indexes. No other factor accounted for more than 13 per cent.

Source of Data Used

Census tract information[1] was used for social indexes (with the exception of school age illegitimacy and juvenile delinquency) and for establishing population bases for health indexes. Data were obtained from birth and death certificates through the courtesy of the Bureau of Statistics of the San Francisco City and County Health Department. Information on venereal disease and tuberculosis was obtained from this bureau also, with the cooperation of the appropriate divisions of the health department. Material on juvenile delinquency was obtained from the Research and Statistics Subcommittee of the San

Table 1—Availability of selected health and socioeconomic indexes

Health Indexes	Available	Available no. too small	Not available by census tract	Not available
Maternal mortality		X		
Inadequate prenatal care	X			
*Fetal mortality	X			
*Neonatal mortality	X			
*Postneonatal mortality	X			
*Childhood mortality	X			
*Incidence of prematurity	X			
Use of well child conference			X	
Immunization levels				X
Pertussis incidence				X
Rheumatic fever incidence		X		
Diphtheria incidence		X		
Tetanus incidence		X		
Poliomyelitis incidence		X		
Typhoid fever incidence		X		
Tuberculosis incidence	X			
*Venereal disease incidence	X			
Incidence of battered children				X
School lunch usage			X	
Selective service rejection			X	
Summary	8	6	3	3

Socioeconomic Indexes	Available	Available no. too small	Not available by census tract	Not available
Unemployment	X			
Low income	X			
Inadequate education	X			
Overcrowding	X			
Parental composition	X			
School-age illegitimacy	X			
AFDC rates			X	
Juvenile delinquency	X			
Adult crime				X
	7	0	1	1

* The number of events for these indexes in one year was too small, but information could be obtained by grouping data for several years.

Francisco Committee on Youth. This material had originally been prepared for the subcommittee from original records of the San Francisco Juvenile Court and Police Department.

The cooperation of staff members of the San Francisco Health Department in making it possible to secure and use the data cited in this report is gratefully acknowledged.

Results

A total of 20 health indexes and nine socioeconomic indexes were selected by

the authors for study and analysis. These indexes are listed in Table 1.

Of these 29 indexes, only eight of the health indexes and seven of the socioeconomic indexes proved to be available. The available health data consisted of mortality (maternal, fetal, neonatal, postneonatal, childhood). In addition, information was also available for adequacy of prenatal care, incidence of prematurity, reported cases of venereal disease, and incidence of tuberculosis. Nonavailable health data consisted first of some service statistics (use of well-child conferences and school lunches), and some information on the level of health or morbidity (immunization levels, pertussis incidence, selective service rejections, and incidence of battered children). There was also a group of six indexes which were unavailable because the conditions measured are now uncommon. This group included maternal mortality and the incidences of five communicable diseases: rheumatic fever, diphtheria, poliomyelitis, tetanus, and typhoid fever.

The available socioeconomic data consisted of that which traditionally describe the sociologic environment in which families live and children are reared (unemployment, income, education, overcrowding, and parental composition) plus two reported symptoms of asocial behavior of youth (juvenile delinquency and school age pregnancy). Data were not fully available on Aid to Families with Dependent Children rates or for crimes committed by adults.

Table 1 summarizes the availability of these indexes.

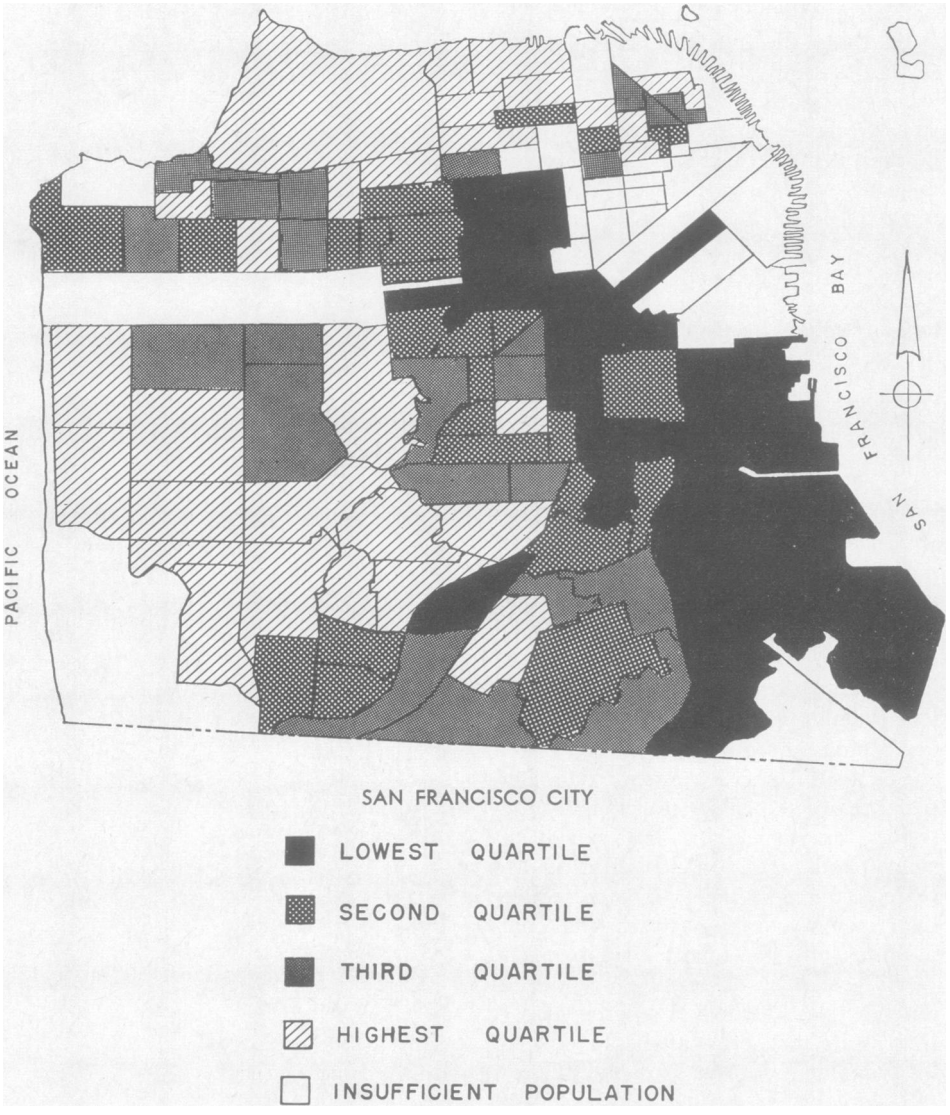
Table 2 summarizes the two methods used to judge the usefulness of the 15 available indexes. The results of the factor analysis are shown in Appendix B. Three of the 15 maps used are reproduced in Figures 1, 2, and 3.

There was agreement on the usefulness of 13 of the 15 indexes. One exception was fetal mortality which appeared useful by the map method but which had a factor loading of 0.46, slightly below the arbitrary figure of 0.5, which was the lower limit of "usefulness" for the factor analysis. The other was juvenile delinquency which, despite a factor loading of 0.79, was judged as "useful" in only three out of the five independent determinations. These two indexes will be included among the useful indexes in the subsequent discussion. Thus substantial agreement was found between the two meth-

Table 2—Comparison of charting and inspection of maps with factor analysis

I. Indexes Judged "Useful" by Both Methods
Health Indexes
inadequate prenatal care
incidence of prematurity
tuberculosis incidence
Social Indexes
low income
inadequate education
unemployment
overcrowding
parental composition
school-age illegitimacy
II. Indexes Where Methods Did Not Agree
Health Indexes
fetal mortality ("Useful" by map method, "not useful" by factor analysis.)
Social Indexes
juvenile delinquency ("Useful" by factor analysis, "not useful" by map method.)
III. Indexes Judged "Not Useful" By Both Methods
Health Indexes
postneonatal mortality
neonatal mortality
venereal disease ratio
childhood mortality
Social Indexes
(None)

Figure 1—Prenatal care index.



ods and, in the opinion of the authors, either method can be used successfully.

The reliability of 13 of the 15 indexes was judged to be satisfactory. The venereal disease ratio was judged unreliable because of the question of the adequacy of reporting. The childhood mortality index was considered unreliable because, although there were sufficient events recorded to enable an index to be computed, the number of childhood deaths in all but the largest census tracts was too small for confidence in the stability of the resulting death rates. Neither of these indexes had been judged "useful" by either the map method or factor analysis.

Thus, of the 29 indexes studied, a

total of 11 indexes proved to be available and useful in identifying high risk census tracts. These 11 indexes included four health indexes (inadequate prenatal care, fetal mortality, incidence of prematurity, and tuberculosis incidence) and seven social indexes (low income, inadequate education, unemployment, overcrowding, parental composition, school-age illegitimacy, and juvenile delinquency).

Discussion

It is evident that the traditional "map method" of analyzing the indexes reported in this paper can be used, and in fact is used by the public health

Figure 2—Overcrowding index.

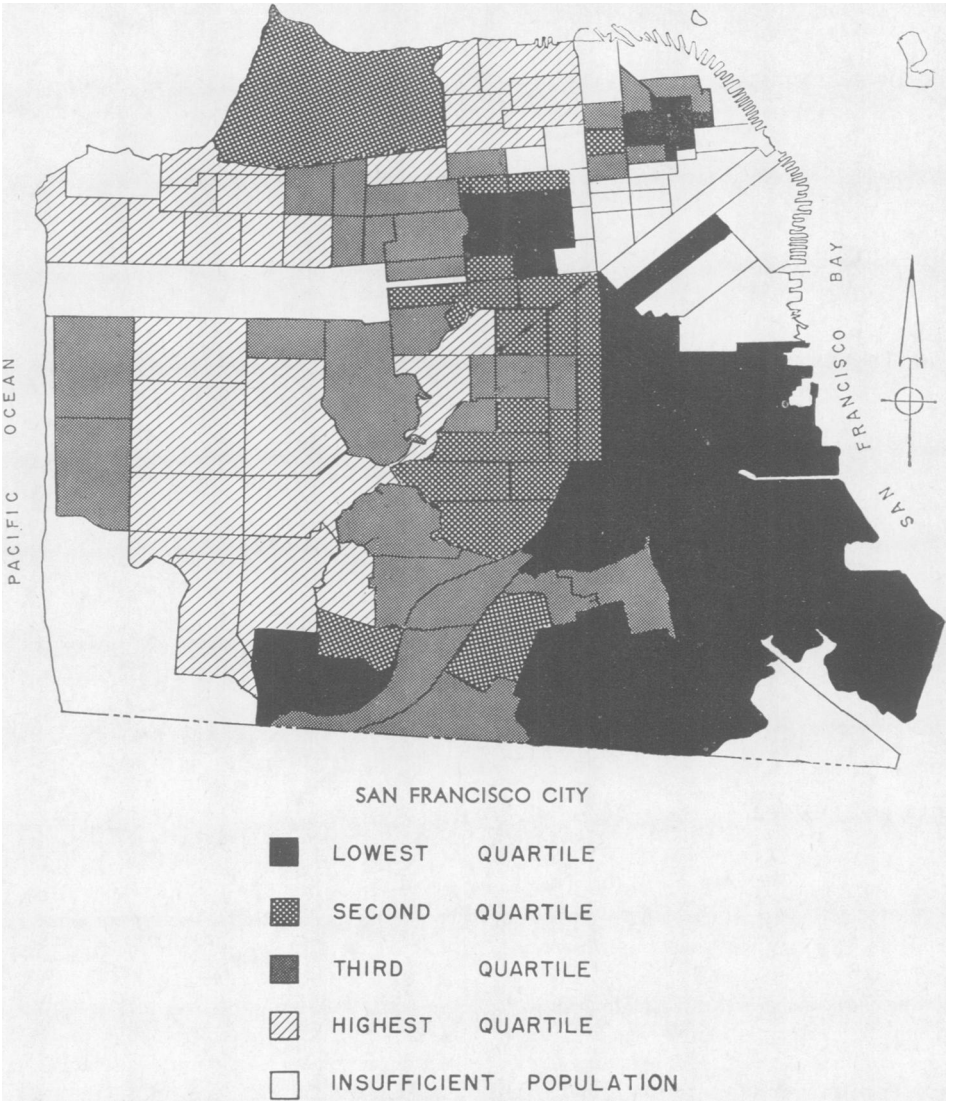
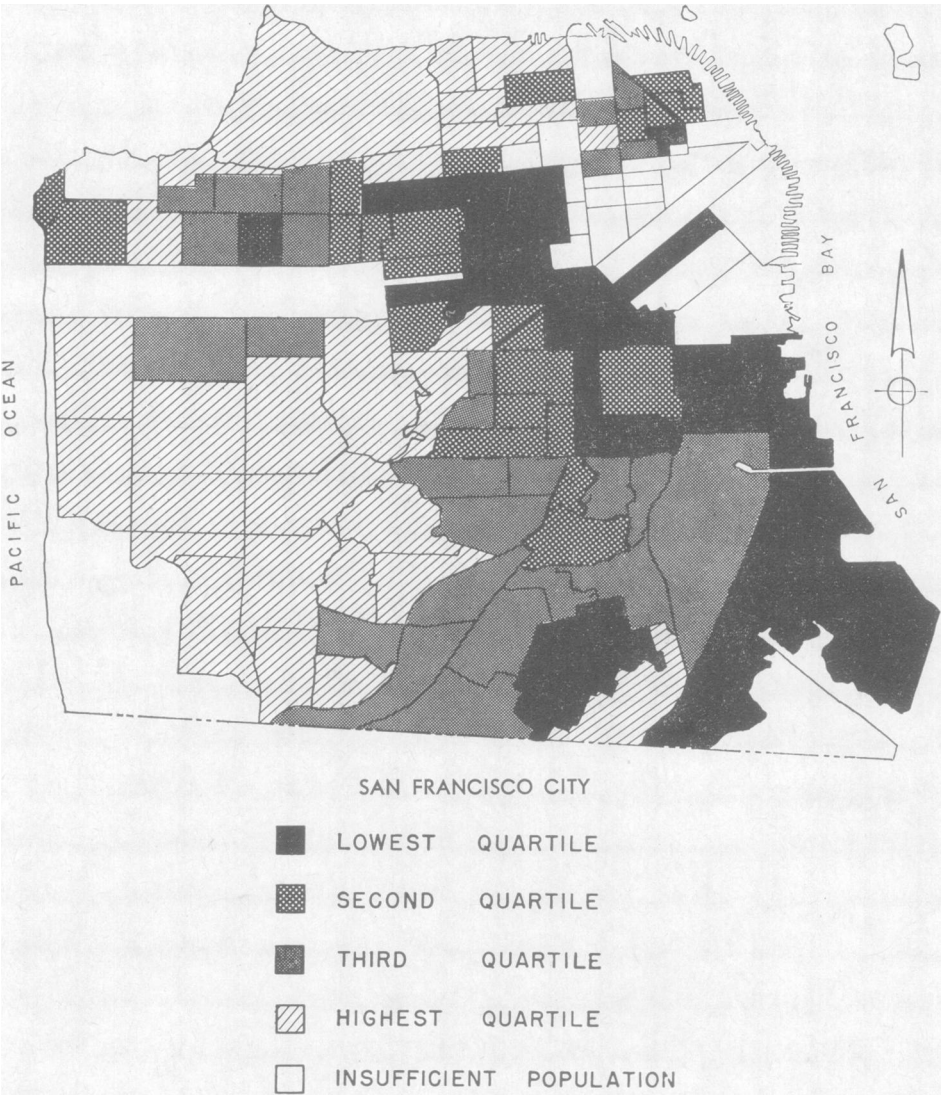


Figure 3—Income index.



administrator for identifying problem geographic areas in the community under study. To our knowledge, a factor analysis has not been used for this purpose, although it has been applied,[2] as we have done, to the identification of groups of indexes which can be used for specific purposes. The agreement between the two methods in our study suggests that the map method is adequate for the identification of "high risk areas" by public health administrators.

It is not known to what extent the indexes used in this study would prove to be available and useful in other cities. It would be anticipated that all of the census information, and much of the morbidity and mortality informa-

tion, would be available. Certain indexes, however, are less likely to be generally available. For example, our information on school-age illegitimacy can only be duplicated in communities where either the legitimacy of the child or the name of the father is recorded on birth certificates. (In San Francisco, the latter information is recorded, and our data on illegitimacy in reality represent children whose parents' surnames are different.) Our data on juvenile delinquency came from a special study, and this information had not previously been available.

Other useful data may also be available in other communities. Cities with a battered child registry may find this information useful, and certainly rates for AFDC, if they can be obtained by census tract, would indicate the location of areas with increased social or economic problems. Cities with higher incidences of communicable diseases than San Francisco may also find communicable disease indexes useful.

It is curious that certain health indexes which have been traditionally used as indicators of high risk areas did not appear as useful as other indexes in this study. For example, inadequate prenatal care and the incidence of prematurity are found to be more useful health indexes than neonatal and postneonatal mortality. This is unexpected in view of the usual high value placed upon the infant mortality rate as an index of community health.

In this study, more socioeconomic indexes were found to be consistently useful than health indexes. Seven of the original nine socioeconomic indexes but only four of the original 20 health indexes were found to be available and useful. This finding has several interpretations: (1) that we selected better socioeconomic indexes than health indexes to study; (2) that more data were available for the socioeconomic indexes, at least from the one community under study; (3) that socioeconomic indexes are generally more useful than health indexes, in identifying high risk geographic areas.

It is noteworthy that seven of the 11 indexes, which we found to be available and useful require census data either for the entire index or for denominators. Since census tract data are collected at ten-year intervals, this seriously limits the use of these indexes in areas which are undergoing rapid population changes. Pending further studies of the four useful indexes derived entirely from annual birth and death registrations (inadequate prenatal care, fetal mortality, incidence of prematurity, and school-age illegitimacy) it appears most practical to identify high risk areas by a consideration of indexes derived from decennial censuses, supplemented by an examination of more recent changes in those indexes which can be obtained in intercensal years. Indexes for which it is advisable to pool data for more than one year might be computed at three-year intervals, or alternatively, as annual averages of the three-year rate.

A word might be added on the practicality of using geographical divisions other than census tracts for computing indexes. Census tract divisions are in many areas too small to permit calculation of reliable indexes, and in addition they vary tremendously in population. (In San Francisco, the 126 census tracts have a population range from 8 to 21,931, and an area range of 9.43 to 1,621.30 acres.) Many agencies have administrative divisions which include groups of census tracts. If the boundaries of these divisions coincide with census tract boundaries, indexes can easily be developed for the larger divisions, and these indexes will be more reliable than ones based on census tracts. However, when these boundaries do not coincide with census tract boundaries, it will often be an insuperable task to apply census data to the larger divisions.

Summary

This paper describes the availability of health and social indexes in San Francisco, and compares two methods (plotting the data on maps and factor analysis) of determining their usefulness in the identification of high risk census tracts. Either method can be used to determine the usefulness of indexes and the findings from one method reinforce those of the other. In this report the most useful health indexes of those available and studied were inadequate prenatal care, fetal mortality, and the incidence of prematurity, and the incidence of tuberculosis. The most useful socioeconomic indexes were low income, inadequate education, unemployment, overcrowding, parental composition, school-age illegitimacy and juvenile delinquency. Since four of these indexes are available in intercensal years, it is suggested that these four indexes can be used in postcensal years in areas of rapid population change to supplement indexes derived at the time of the decennial census.

REFERENCES

1. U. S. Bureau of the Census. U. S. Censuses of Population and Housing, 1960. Census Tracts, Final Report, PHC (1)-137. Washington, D. C.: Gov. Ptg. Office, 1962.
2. Shevsky, E., and Bell, W. Social Area Analysis. Stanford, Calif.: Stanford University Press, 1955.

The authors are affiliated with the Division of Maternal and Child Health, University of California School of Public Health (Earl Warren Hall), Berkeley.

This investigation was partially supported by PHS Research Grant No. 1-GS-142 from the University of California School of Public Health. The paper was submitted for publication in May, 1966.

APPENDIX A

Definitions of Health and Social Indexes

Health Indexes

- inadequate prenatal care: Live births with no prenatal care, or prenatal care only in third trimester per 1,000 live births.
- fetal mortality: Infants over 400 grams born dead per 1,000 live births, 1959-1961.
- incidence of prematurity: Infants born alive weighing 2,500 grams or less at birth per 1,000 live births, 1959-1961.
- neonatal mortality: Deaths age 0 to 28 days per 1,000 live births 1959-1961.
- postneonatal mortality: Deaths from 28 days to the first birthday per 1,000 live births, 1959-1961.
- childhood mortality: Deaths 1-19 years per 100,000 population 1-19 years, 1959-1961.
- VD index: Reported nonprimary syphilis cases (excluding congenital) per 1,000 reported cases of syphilis, 1958-1962.
- tuberculosis incidence: Reported new cases per 10,000 population, 1960.

Social Indexes

- unemployment: Unemployed males in civilian labor force per 1,000 males in civilian labor force, 1960.
- low income: Families with annual incomes under $3,000 per 1,000 families, 1960.
- inadequate education: Adults with 8th grade education or less, per 1,000 adult population, 1960.
- overcrowding: Housing units with more than 1.0 persons per room per 1,000 housing units.
- parental composition: Children under 18 not living with 2 parents, per 1,000 population under 18 years, 1960.
- school-age illegitimacy: Illegitimate live births to mothers aged 15-19 per 1,000 live births, 1960.
- juvenile delinquency: Boys 8-17 charged with a non-traffic offense by police or juvenile court per 1,000 male population 8-17, 1960.

All indexes were recomputed to rank orders.

APPENDIX B

Results of Factor Analysis of Data

Factor	Sum sq.	Cumulative per cent of variance
1	6.5183	43.45
2	1.8894	56.05
3	1.2000	65.05
4	0.9844	70.61

Index	Factor 1	2	3	4
1. Low income	0.85702	0.23161	0.16115	0.02059
2. Inadequate education	0.69214	0.58300	0.13064	0.10287
3. Parental composition	0.77614	0.33972	0.03778	0.03177
4. Juvenile delinquency	0.78941	0.10861	0.20358	0.02809
5. School-age illegitimacy	0.66582	0.42715	0.12846	0.02528
6. Overcrowding	0.76767	0.46278	0.00469	0.05911
7. Unemployment	0.84022	0.25200	0.21246	0.11641
8. Incidence of prematurity	0.68092	0.46206	0.30668	0.06319
9. Neonatal mortality	0.45737	0.39591	0.28514	0.25171
10. Inadequate prenatal care	0.86255	0.02952	0.00182	0.12873
11. Tuberculosis incidence	0.67427	0.19480	0.22496	0.16415
12. Fetal mortality	0.46215	0.31737	0.27362	0.01601
13. Childhood mortality	0.20222	0.57111	0.39680	0.13021
14. VD index	0.35598	0.22873	0.76908	0.16890
15. Postneonatal mortality	0.29633	0.14229	0.05117	0.89291