THREE TESTS FOR SPACE-TIME INTERACTION: A COMPARATIVE EVALUATION

PIERO MUSTACCHI,*
UNIVERSITY OF CALIFORNIA MEDICAL CENTER, SAN FRANCISCO
F. N. DAVID,**
UNIVERSITY COLLEGE, LONDON and UNIVERSITY OF CALIFORNIA, BERKELEY
and EVELYN FIX**
UNIVERSITY OF CALIFORNIA, BERKELEY

1. Introduction

The study of low order epidemics has required developing methods suitable for the analysis of the temporospatial relationship between cases. This acquires particular importance in conditions of unclear etiology where the direction of research can be greatly influenced by an indication that cases tend to occur in clusters. Mantel [1] has critically reviewed a number of methods applicable also in the absence of knowledge of the basic population. This characteristic is quite intriguing, since epidemiologic methodology has traditionally insisted on adequate definition of the denominator population.

The merit of two specific approaches, namely, the ridit method of Bross [2] and the sum of empirical clusters device of Ederer, Myers and Mantel [3], is immediately evident from their ability of demonstrating randomness where this would be anticipated, for example, addresses of traffic fatalities [4], or of revealing the expected clustering of cases in diseases of known infectious origin [3].

It is, however, difficult to evaluate adequately their respective usefulness as investigative tools in the study of the temporospatial distribution of diseases of undetermined etiology, for example, childhood leukemia. In fact, several considerations may restrict the direct comparability of the two methods as well as of the resulting conclusions.

Methodologically, one approach [2] is based on the general premise that within all possible $n(n-1)/2$ pairs, temporospatial clustering is reflected in an overrepresentation of pairs with short spatial and temporal distances. The other [3] considers the maximum number m_1 of cases occurring in any one of

* With support of the National Institutes of Health, USPHS Research Grant 9751.
** With the partial support of the National Institutes of Health, USPHS Grant GM-10525.

the equal time subintervals into which the study period had been divided. Given the total number of cases observed during the entire study period, combinatorial analysis gives the expectation and the variance of m_1. This information is used for the calculation of a one degree, continuity corrected chi square.

The first method may suffer from not taking into account the fact that population density is not uniform throughout the city, whereas, within relatively broad limits, population distribution may be safely ignored in the other. Geographically too, the two studies are not entirely comparable. Aside from the possibility that the conflicting conclusions on leukemia clustering [3], [4] could reflect true variability in the pattern of its occurrence, it should be noted that one study [4] analyzed the paired distance of childhood leukemia cases in the area of Buffalo, New York, while the other [3] considered the empirical clusters of childhood leukemia observed during two five year periods in each of the 169 towns comprising the state of Connecticut. Conceivably, in the Connecticut study, two biologically related cases could escape being identified as a pair if, by occurring in adjoining towns, they were separated by a political or administrative boundary. This potential source of difficulty appears unlikely in the paired distance study encompassing the Buffalo area.

Thus, methodologic and geographic considerations preclude a satisfactory reconciliation of the evidence for clustering of leukemia cases reported for Buffalo with the one negating it in Connecticut. It seemed, therefore, desirable to determine the direction of the results each method would yield when applied to the *same* study population, since in this way one could expect a reduction, or hopefully even the deletion, of the difficulties inherent in the comparison of two different populations with two differing methods.

To our knowledge, there is no published report of such a study. Nor are we aware of any comparative evaluation, at the operational level, of these two methods with traditional approaches requiring information on the population at risk.

For these reasons, this note presents and compares the results obtained when three different methods proposed for the analysis of the temporospatial distribution of rarely occurring events are applied to the same study group. Of the three, one [5] predicates the definition of the basic population while the others are applicable in the absence of such knowledge.

2. Material

The study group consisted of 80 poliomyelitis deaths that occurred among the three quarter million San Francisco residents during the prevaccine decade 1946–55. (There was one additional decedent for whom the exact home address could not be identified; he was excluded from the study.) This condition was selected because, in its sporadic occurrence as well as in its fatal outcome, it seemed to bear a closer biologic resemblance to childhood leukemia than did the entire spectrum of clinical poliomyelitis.

During 1946–55, 1180 resident cases of poliomyelitis were reported to the San Francisco Department of Health; the home address was available for all but three. In contrast with all the clinical cases of poliomyelitis where a clustering effect was immediately apparent by simple inspection of the morbidity reports, no pattern of temporospatial aggregation was discernible for the addresses and dates of onset of the polio fatalities, when they were plotted on a topographic map of the city. On the one hand, this did not seem surprising, since the low probability of such an occurrence considerably lowered the likelihood that any two polio deaths be closely associated. On the other hand, it could be argued that since in San Francisco certain variables which are not randomly distributed (for example, age, sex pregnancy, and crowding) influence the risk of contracting [6] and presumably also of dying from polio, some clustering effect should be present.

Age of decedents was expressed in number of completed years at the time of the clinical onset of the fatal disease. Address was the home address as reported to the San Francisco Department of Health. Population data obtained at the time of the U. S. census of 1950 were used wherever indicated. Census tract boundaries were those adopted in 1950.

3. Methods

The temporospatial distribution of the San Francisco poliomyelitis deaths was studied with the ridit analysis of Bross [2], the sum of empirical clusters device of Ederer, Myers and Mantel [3], and the persisting high rates approach of Mustacchi [5].

3.1. *Ridit analysis.* In essence, pairs of decedents whose homes were separated by a distance of one mile or more were considered independent from each other. There were 2889 pairs out of all possible 3160 which, by this definition, were "not related"; they constituted the relatively identified distribution. Again by definition, the average ridit of any spatial distance between pairs is the average of the temporal distribution of pairs for that spatial distance, where the percentile position of any individual in the temporal distribution is determined by referring to the identified temporal distribution of pairs more than one mile apart [4]. Since average ridits are uniformly distributed, a ridit significantly less than 0.50 occurring in those spatial distance categories whose pairs are close together, is suggestive of aggregation.

The 271 pairs less than one mile apart were subgrouped into five spatial categories of successive 0.20 miles. Average ridits were calculated for each subgroup as well as for all pairs less than one mile apart.

3.2. *Sum of empirical clusters.* The temporal units adopted were two five years intervals (1946–50 and 1951–55) each one of which was further subdivided into five one year subintervals. In order to evaluate the effect of the size of the spatial unit selected on the results obtained with this method of analysis, the following categories were used in separate analyses of the observed temporo-

spatial clusters: the city's 118 single census tracts; the city's 15 major tracted areas; 13 aggregates of contiguous census tracts yielding equipopulated zones as regards individuals under age 15; the city as a single unit.

3.3. *The persisting rates approach.* This approach assumes that if temporospatial clustering exists, areas in the city characterized by high (or low) rates during a base line period will show during a subsequent observation period a rate gradient in the same direction. For this purpose the city's census tracts were considered as being "with" or "without" polio deaths according to whether fatal cases of polio had occurred during 1946, 1947, and 1948. Subpopulations were calculated for each of the two complementary portions of the city. Observed age specific polio mortality rates were calculated on the basis of the citywide mortality experience for 1949–55 and were used to calculate "expected" number of deaths in each of the two subpopulations. Differences between expected and observed were calculated with the chi square method.

4. Results and comments

Each one of the three methods discloses evidence for temporospatial aggregation of poliomyelitis deaths. Figure 1 summarizes the results of the ridit analysis

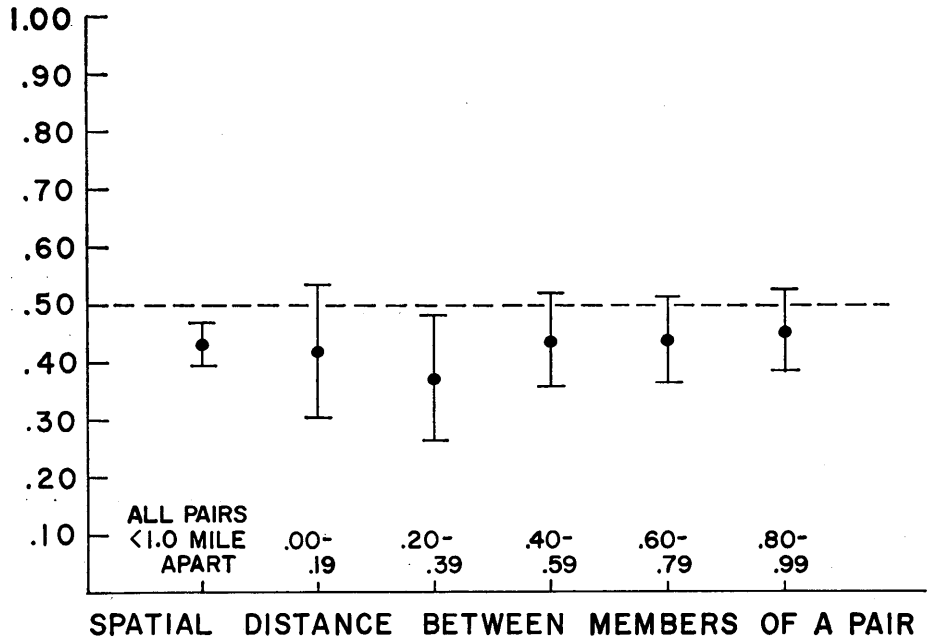

FIGURE 1

Poliomyelitis deaths, San Francisco 1946–55.
Average ridits and 95 per cent confidence intervals for pairs less than one mile apart.
Five spatial categories.

and the 95 per cent confidence limits for the pertinent ridits. It shows that there is a tendency among the entire group of 271 pairs less than one mile apart to cluster also in time. This evidence becomes more tenuous when spatial subcategories are analyzed individually. This seemingly paradoxical finding may be ascribed at least in part to the relatively small numbers of pairs included in each spatial subcategory.

A similar effect of a decreased number of observations leading to a dilution in the evidence for clustering was noted in Connecticut by Ederer [3]. The exceedingly high chi square value obtained on the basis of all 4020 cases of poliomyelitis shrank considerably when the calculations were based on a ten per cent sample. This effect is present also in the San Francisco data when results of analyses based on all cases are contrasted with those based on fatalities only (table I).

TABLE I

AN E-M-M ANALYSIS OF POLIOMYELITIS

		Number of Observations	χ_1^2
Connecticut	169 towns; all cases	4020	1767
1950–54	169 towns; 10% sample	397	57
San Francisco	118 census tracts; all cases	1177	267
1946–55	118 census tracts; deaths only	80	0.4
	18 tracted areas; deaths only	80	3.4
	13 equipopulated zones; deaths only	80	8.4
	the entire city; deaths only	80	65.0

The other findings summarized in table I show a trend compatible with this "dilution" effect. When there is a disproportionately large number of spatial units available to accommodate a relatively small number of cases, a substantial number of units can be expected to show only one case or none at all. Since no cluster can exist with less than two cases, this results in an appreciable decrease in the number of cases amenable to analysis. This is apparent from the suspiciously low chi square value obtained from an analysis of the distribution of the 80 deaths among the city's 118 census tracts. As the spatial units are enlarged, the loss of single cases becomes less crucial, and the evidence for clustering obtained with the E-M-M method gains momentum.

At the same time, with the adoption of large spatial units, another factor enters the picture, namely, the lesser likelihood that two biologically related cases be separated by an artificial boundary line and thus go unrecognized as a pair. This seems to be especially true when the entire city becomes the spatial unit for analysis. Moreover, in San Francisco, the shoreline which accounts for three quarters of the city's boundaries, makes it a geographical impossibility for any case to exist beyond it. Here the Ederer-Myers-Mantel analysis of the city as a single spatial unit is able to pinpoint quite effectively the two epidemic

years during the two five year periods of the study. They were: 1948 with 26 deaths; and 1952 with 15 deaths. A total of 45 poliomyelitis deaths had occurred during 1946–50 and 35 during 1951–55.

In a similar vein, the detection of significant clustering of poliomyelitis deaths when the city is divided into 13 equipopulated zones and the more hesitant conclusions reached when the 18 major tracts are analyzed may reflect different degrees of homogeneity with regards to a large group of susceptibles.

Accordingly, it appears that for this method to remain applicable in the absence of knowledge of the basic population, it may be necessary that the denominator population be contained in distinctly autonomous spatial units. On the other hand, since the effectiveness of the method depends on the deviation between m_1 and its expectation, it is easily seen that one or more spatial units could experience an excessive number of cases which could go undetected by the method simply because the increased incidence persisted throughout the entire span of the study period.

The considerations referable to the error incurred when a pair of related cases is separated by an intervening boundary line are applicable also to the results obtained with the third method used. However, in both instances the direction of the error is the same and would tend to reinforce any evidence of clustering. Table II summarizes the findings obtained with the last approach and, in agreement with the other two methods, concludes to the presence of temporospatial clustering of the fatal cases of poliomyelitis.

TABLE II

DEATHS FROM POLIOMYELITIS, SAN FRANCISCO
Expected on the basis of the age adjusted rates effective in the entire City during 1949–55.

Situation 1946–48	Population 1950	Deaths 1949–55	
		Observed	Expected
Census tracts without fatal poliomyelitis	514,843	19	28.1
Census tracts with fatal poliomyelitis	260,514	25	15.9
Total: the City	775,357	44	44

During the base line years 1946–48, clinical poliomyelitis occurred within more census tracts than experienced the fatal form of the disease. In other words, included in the subpopulation without poliomyelitis deaths during the three base line years, was a group which was allocated to the subpopulation without clinical polio by virtue of having experienced only nonfatal poliomyelitis. According to this analysis, during 1946–48 five sixths of the San Francisco population lived in tracts without clinical polio. During the subsequent seven years, 719 instead of the expected 680 cases occurred in the same tracts. Conversely, the complementary one sixth of the city's population contributed only 93 instead of the 132 expected cases. (A total of 812 cases were reported for the entire city during 1949–55.)

To the epidemiologist these results are of interest, because evidence of clustering in poliomyelitis can be obtained not only from an analysis of all clinical cases, but also when one adopts as a definition of the illness its most severe expression characterized by a very sporadic occurrence.

5. Summary and conclusions

In San Francisco, during 1946–55 the temporospatial distribution of fatal poliomyelitis was not random. A clustering pattern could be demonstrated with the ridit method, the empirical clusters device and the persisting high rates approach.

The smaller the spatial unit selected for the analysis of the empirical clusters, the less definite became the pattern. This could be related in part to the greater likelihood for spatial units to contain fewer than the minimum two cases necessary to form a cluster or to the increased probability that an intervening boundary line could separate two biologically related cases; in both instances paired cases would inevitably be lost to analysis. The persisting high rates approach may suffer from the second limitation but not from the first, and the ridit method would appear to be exempt from either.

We are indebted to Dr. Ellis D. Sox, Director, San Francisco Department of Health, and Dr. Erwin H. Braff, Chief, Bureau of Disease Control and Adult Health, for access to poliomyelitis records. Our thanks go to Miss Myrtle Cavanaugh, R.N. for her resourcefulness on countless occasions, and to Rick Greenspan and Mary Lou Schmidt who assisted in ridit analysis.

REFERENCES

[1] N. MANTEL, "The detection of disease clustering and a generalized regression approach," manuscript, Biometry Branch, NCI, NIH, Bethesda, 1965.
[2] I. D. J. BROSS, "How to use ridit analysis," *Biometrics*, Vol. 14 (1958), pp. 18–38.
[3] F. EDERER, N. H. MYERS, and N. MANTEL, "A statistical problem in space and time: do leukemia cases come in clusters?," *Biometrics*, Vol. 20 (1964), pp. 626–638.
[4] D. PINKEL, J. E. DOWD, and I. D. J. BROSS, "Some epidemiological features of malignant solid tumors in the Buffalo, New York area," *Cancer*, Vol. 16 (1963), pp. 28–33.
[5] P. MUSTACCHI, "Some intra-city variations in leukemia incidence in San Francisco," *Cancer*, Vol. 18 (1965), pp. 362–368.
[6] P. MUSTACCHI, unpublished results.