

Christovam
Barcellos

Departamento
de Informação e Saúde,
Centro de Informação
Científica e Tecnológica,
Fundação Oswaldo
Cruz, Rio de Janeiro,
Brasil.

The specificities of spatial health data analysis

The article by Gilberto Câmara and Antônio Miguel Monteiro describes various recently developed spatial analysis techniques which have been applied mainly to environmental, geological, and land cover/land use problems, etc. Their use in the collective health field is still not very frequent and can present some analytical limitations. I wish to touch on some of these problems based on the question contained in the title, i.e., questioning the specificities of health data and problems as compared to other areas where these techniques have been applied.

In the first place, all health events – birth, infection, illness, death – manifest themselves in persons. These individuals are not randomly distributed in space. Thus, when one works with health records to evaluate risks, one should estimate the probability of an event occurring, weighted by the population distribution. The most common way to consider population distribution in risk evaluation is to group demographic and health data in watertight spatial units and to subsequently calculate epidemiological indicators. This strategy poses serious limitations, such as ignoring interactions between spatial units and the instability of indicators created in small areas (King, 1979). However, this is not the only way to consider population distribution. For example, one can calculate case density (the number of cases per area), producing a surface of probabilities where areas with more proximate cases present greater risk. Analogously, one can calculate the density of persons (inhabitants per area, or simply population density) as a continuous surface to be used as the denominator for calculating rates. A third strategy to evaluate the spatial distribution of these events is to test the randomness of the “cases” in relation to a set of “controls” obtained by survey or drawing from a population with a similar profile. Population density is always an implicit variable in all spatial analyses of health. However, this variable is not neutral. At least in Brazil, it is associated with concentration of wealth and a particular way of life. This variable is the result of human capacity, through the territorial division of labor, to produce surpluses and technology and to organize power structures. In addition, population clustering can have important repercussions on the spread of diseases, especially transmissible ones. For example, the initial years of the AIDS pandemic were characterized

by the rapid dissemination of the virus in large cities and by its spread through a downwardly hierarchical network of cities. These cities, considered “central”, concentrate people, income, and cases, as well as fostering an intense exchange among individuals, a condition for HIV transmission. Thus, the population of a given place is both the denominator for evaluating risks and one of the conditioning factors for the spread of diseases, which could be expressed mathematically as a differential equation.

In addition, the macro-determinants of diseases, whether environmental, social, or economic, occur “outside” of persons. It is interesting to note that the World Health Organization defines the environment as “the totality of *external* elements that influence the health conditions and quality of life of individuals or communities”. Therefore, if we intend to relate health problems to their determinants, we should combine health data, referenced in the population, to environmental data, referenced to something “external” to the population, with each coming from different information systems. Geographic Information Systems (GIS) can allow for this type of data relationship by superimposing layers of health event incidence rates on other layers relevant to this association (Vine et al., 1997).

Third, in Brazil, epidemiological data are collected according to the territorial logic of the Unified Health System (SUS), with increasing hierarchical levels and primarily administrative objectives. Thus, data location is conducted based on the spatial reference of these units, which display a wide variation in dimensions and resident populations. These dimensions, as well as the form of the reference spatial units, can have a major impact on the visual and statistical results. The Geographical Analysis Machine (GAM), for example, searches for excess points in relation to an expected number within circles generated by the program. However, in various situations one should consider non-circular risk locations, non-Euclidean distances between cases (and between the latter and sources of risk), like the bands around power transmission lines, where exposure to low-frequency radiation can cause damage to human health. By selecting indicators, one should search for a territorial division that maximizes the variances of both exposure and the measured effects on the population. One explains – or makes explicit – the environmental and social determinants on the scale in which the greatest variability in indicators is found (Cleek, 1979). Form can be an important factor for constructing a risk model due to its

influence on the “exposure geometry”, studied through landscape ecology (Frohn, 1998). A more elongated unit can have more neighbors, while compact units have a smaller perimeter and can have less neighborhood relations with other units.

In general, studies in medical geography have been characterized by the search for explanatory factors for a given spatial distribution of diseases, viewing space as an *a posteriori* factor. This approach can produce theoretical simplifications through the association of climatic, cultural, and social characteristics with epidemiological ones, which led a major portion of studies by pioneers in medical geography to conclusions that ideologically reinforced colonialism (Bennett, 1991). The use of neural networks, as suggested by the authors, can reverse the direction of analyses of socio-spatial disease determinants, seeking combinations among factors, constituted *a priori*, to explain this distribution. This approach requires that researchers formally present their hypotheses and construct a series of “layers” representing human spaces and which, when combined, best characterize the places where these diseases occur.

With the improvement of information systems, the inclusion of addresses on health records, and the growing use of satellite positioning equipment or Global Positioning Systems (GPS) in health surveillance activities, one can access these health events as points on a map with a local scale. The main advantage of the data georeferencing strategy is the possibility of producing different forms of data aggregation, constructing indicators in different spatial units according to the study's purpose. The same point (health event) can be contained in different types of spatial units: a neighborhood, a river basin, a health district, etc., defined by polygons on the maps. This characteristic incorporates the principles of simultaneity and interaction between scales for spatial analysis. This property also involves adopting a geometrical rigor that must be present in the planning phase and construction of the mapping base. In order for there to be a univocal relationship between the point and the polygon, the spatial units must cover the entire working area, and one area cannot be covered by more than one polygon, i.e., there cannot be empty places between units or overlapping of them. Each spatial unit represents a slice of space, containing populations at risk of diseases and displaying disease rates. Geographic Information Systems allow one to construct rates for different exposure conditions by superimpos-

ing layers of disease data (points) and population data (polygons). These technical requirements in the handling of both tabulated and mapped data hinder the adoption of less rigid criteria for spatial studies, restricting the concept of space to watertight units. By using network analysis techniques, interpolation, and smoothing of spatial data, one can dissolve previously established boundaries between spatial units. The adoption of fuzzy boundaries for spatial units, ideal for studying place, is jeopardized by the operational norms of information systems (Oppenshaw, 1996).

Spatial analysis is defined as the capacity to generate new information based on existing spatial data (Bailey, 1994). To this end, software applications have been developed that facilitate the search for patterns and exceptions in space. Such techniques do not replace the researcher. Spatial analyses applied to health allow one to study health problems where they manifest themselves. Although this statement may sound obvious, it is important to recall that these analyses are only made possible through the increasingly deep knowledge of both the health problem and the health place.

- BAILEY, T. C., 1994. A review of statistical spatial analysis in geographical information systems. In: *Spatial Analysis and GIS* (S. Fotheringham & P. Rogerson, ed.), pp. 13-44, London: Taylor and Francis.
- BENNETT, D., 1991. Explanation in medical geography. Evidence and epistemology. *Social Science and Medicine*, 33:339-346.
- CLEEK, R. K., 1979. Cancer and the environment: The effect of scale. *Social Science and Medicine*, 13D: 241-247.
- FROHN, R. C., 1998. *Remote Sensing for Landscape Ecology: New Metric Indicators for Monitoring, Modeling and Assessment of Ecosystems*. Boca Raton: CRC Press.
- KING, P. E., 1979. Problems of spatial analysis in geographical epidemiology. *Social Science and Medicine*, 13D:249-252.
- OPPENSHAW, S., 1996. Fuzzy logic as a new scientific paradigm for doing geography. *Environment and Planning A*, 28:761-768.
- VINE, M. F.; DEGNAN, D. & HANCHETTE, C., 1997. Geographic information systems: Their use in environmental epidemiologic research. *Environmental Health Perspectives*, 105:598-605.