INTERNATIONAL JOURNAL
OF HEALTH
& ANIMAL SCIENCE
FOOD SAFETY

UNIVERSITÀ DEGLI STUDI DI MILANO
DIPARTIMENTO DI SCIENZE VETERINARIE
PER LA SALUTE, LA PRODUZIONE ANIMALE
E LA SICUREZZA ALIMENTARE
Vespa

Spatial dependence of genetic data related to human health and livestock disease resistance: a role for geography to support the One Health approach

Stéphane Joost[1,2,3]*

1 Laboratory of Geographic Information Systems (LASIG), School of Architecture, Civil and Environmental Engineering (ENAC), Ecole Polytechnique Fédérale de Lausanne (EPFL), Lausanne, Switzerland

2 Unit of Population Epidemiology, Division of Primary Care Medicine, Department of Community Medicine, Primary Care and Emergency Medicine, Geneva University Hospitals, Geneva, Switzerland

3 Group of Geographic Information Research and Analysis in Public Health (GIRAPH; www.giraph.org)

Abstract

The spatial dependence of located health and/or genetic data can be used to detect clusters likely to reveal disease prevalence or signatures of adaptation possibly associated with characteristics of the local environment (high temperatures, air or water pollution), be it in humans or animals (Murtaugh *et al.* 2017).

Most often, geographic maps are produced to represent health data. Medical information is transmitted through thematic choropleth maps. For instance administrative units are colored according to the variable of interest.

But it is key to analyse health and/or genetic data by explicitly including geographic characteristics (distances, co-location) and also the potential and power of spatial statistics to detect specific patterns in the geographic distribution of disease occurrences ("make visible the invisible"). A classic example using clusters is the map produced by John Snow (Snow 1855) showing the number of deaths caused by a cholera outbreak in London. Looking at a detail of Snow's original map, it is possible to realize how he graphically represented the number of deaths, with short bold lines representing death occurrences (frequencies forming a kind of histogram) placed on the street at the addresses where it happened - what we currently name georeferencing. A cluster of death people is an effect observed on the territory, and the existence of such a cluster *depends* on an infected water pump located at the same place (the cause). How can this spatial dependence be detected and measured?

It is possible to identify spatial patterns in the geographic space by means of spatial statistics. We need to determine whether the variable of interest is randomly distributed or spatially dependent, and to check if the patterns observed are robust to random permutations. We also need to explore the data, to find out what is the range of influence of this spatial dependence. Here we focus on the functioning of one among several measures of spatial autocorrelation named Moran's (Moran 1950). Moran's I translates the global relationship between the behavior of points and of their neighborhood.

Measures of spatial dependence are key to detect and visualize spatial patterns in health and/or genetic data because spatial statistics can reveal signals that remain often hidden using thematic mapping. On the basis of the clusters highlighted by these exploratory methods, it is possible to formulate hypotheses about possible environmental or socio-economic causes and to test them with the help of confirmatory statistics. «Ideas come from previous explorations» John Tukey said in a paper published in 1980 in The American Statistician, in a paper entitled «We Need Both Exploratory and Confirmatory» (Tukey 1980). First explore and then confirm was already the reasoning applied by John Snow to detect death "hot spots" in London, which then allowed him to hypothesize that a particular water pump was infected, and finally to take public health steps to check the cholera epidemic.

Two examples illustrate the use of these spatial statistics: first, a cohort named COLAUS and established in the city of Lausanne was used to replicate the results obtained with 120'000 adults from the UK Biobank study to test the hypothesis that high-risk obesogenic environments and behaviors accentuate genetic susceptibility to obesity (Tyrell *et al.* 2017). Our findings suggest that the obesogenic environment accentuates the risk of obesity in genetically susceptible adults. Of the factors we tested, relative social deprivation (Townsend Deprivation Index) best captures the aspects of the obesogenic environment responsible. We produced a map of Lausanne showing the results of bivariate Local Indicators of Spatial Association (LISA) involving: 1) the value of the genetic risk score (GRS) based on 69 genetic variants and associated with obesity as identified by the GIANT consortium (more than 330'000 individuals); 2) the Townsend Deprivation Index (TDI), a composite measure of deprivation based on unemployment, non-car ownership, non-home ownership and household overcrowding. The analysis permits to identify clusters where a high GRS depends on a high mean of the TDI calculated within a spatial lag of 800m. Compared with a previous analysis applied to BMI in Lausanne, we were able to delimit areas where genetic susceptibility and deprivation result in observed obesity.

The second example is an application of landscape genomics (Joost *et al.* 2007) to goat breeds in Europe and to cattle in Uganda to show how measures of spatial autocorrelation can be used to identify similarities or differences in genotype occurrences between neighboring individuals that cannot be explained by chance (Stucki *et al.* 2016). In Uganda, LISA indicators applied to genomic data in the Ankole cattle breed reveal a pattern corresponding to the known geographic distribution of *Trypanosoma brucei gambiense*.

References

Joost S, Bonin A, Bruford MW et al., 2007. A spatial analysis method (SAM) to detect candidate loci for selection: towards a landscape genomics approach to adaptation. Molecular Ecology, 16, 3955–3969.

Moran PAP (1950) Notes on Continuous Stochastic Phenomena. Biometrika, 37, 17–23.

Murtaugh MP, Steer CJ, Sreevatsan S et al., 2017. The science behind One Health: at the interface of humans, animals, and the environment. Annals of the New York Academy of Sciences, 1395, 12–32.

Snow, J., 1855. On the Mode of Communication of Cholera. John Churchill. London.

Stucki S, Orozco-terWengel P, Forester BR et al., 2016. High performance computation of landscape genomic models including local indicators of spatial association. Molecular Ecology Resources, Doi: 10.1111/1755-0998.12629.

Tukey, J.W., 1980. We Need Both Exploratory and Confirmatory. The American Statistician, 34, 23–25.

Tyrrell J, Wood AR, Ames RM et al., 2017. Gene–obesogenic environment interactions in the UK Biobank study. International Journal of Epidemiology, dyw337.