Spatial epidemiology and the information revolution

Quantifying the global burden of infectious diseases…

The preceding profile describes what we have achieved in defining the spatial distribution of malaria. The issues of data paucity that necessitate the modelling approach described for malaria are even greater for many other infectious diseases, which are known to impact on the health of individuals but may be less widespread, less prevalent where they are found, newer (i.e. emerging) and less well understood.

Our group has reviewed the current state of knowledge for 355 infectious diseases identified as being of clinical significance to human health, and has proposed a subset of 174 infectious diseases that can and should be mapped, both as a diagnostic resource and to enable health policy planning. Of these 174 infectious diseases, only seven have been reliably mapped to date leaving a large gap in the global knowledge base.

This information is important for human health in all countries. Health systems can be hampered by a lack of knowledge about the diseases present in an area, and there is a global need to prepare for the emergence or spread of new diseases. Both goals require baseline data describing the current distributions of known infectious diseases.

The advances made in modelling the distribution of malaria parasites have required substantial effort from a dedicated international team established in 2005. It is not feasible to replicate this effort for 174 other diseases, most of which lack the survey data available for malaria. To map these diseases, we must simplify our modelling methods and develop innovative approaches to obtain the data required within a meaningful timeframe. Our modelling approach for this suite of diseases will be based on ecological niche modelling, which uses reports of disease occurrence at specified locations to model the environments where a disease is found, and extrapolate from this to locations where the disease is likely to occur. These locations need to be limited to areas within the known range (the 'definitive extent') of the disease. Our approach to finding, extracting and validating the data required by the model builds on the current information revolution.

Developments in machine learning, crowd-sourcing and citizen science have the potential to revolutionise the field of spatial epidemiology. We are building on the work done by HealthMap and others to obtain sub-national disease occurrence data from internet-based health reports and will extend this to social media sources such as Twitter. Twitter, in particular, has been shown to identify disease outbreaks one to two weeks ahead of traditional surveillance systems. Applications built to semi-automate the geopositioning of this information will greatly speed up the process of assigning each disease occurrence to a location. A crowd-sourcing approach will validate these locations following the example set by GeoWiki.

The occurrence points will then feed into our models to produce a first map, which will be validated by a volunteer community, building on the work of projects like Zooniverse. This citizen science validation will be assessed with reference to gold standard, expert-validated datasets. The process of validation will feed back into the process together with new occurrence points to produce new iterations of each map. In this way, we can move the field of spatial epidemiology from static maps to dynamic evolving maps of risk that are constantly improved and updated, covering a range of diseases that have never previously been mapped on this scale.

Developments in machine learning, crowd-sourcing and citizen science have the potential to revolutionise the field of spatial epidemiology.

The benefits of this work are clear. Spatial distributions for multiple diseases can be combined to define the disease landscape at each location and study the underlying factors in disease diversity. Patterns in the changing distribution of a disease over time will feed into models of disease emergence and spread. Most importantly, only when we have this baseline information can we hope to quantify the global burden of these diseases in space and time as we are able to for malaria.

seeg
spatial ecology and epidemiology group

Dr Catherine Moyes
catherine.moyes@zoo.ox.ac.uk

Dr Peter Gething
peter.gething@zoo.ox.ac.uk

Professor Simon Hay
simon.hay@zoo.ox.ac.uk

Spatial Ecology and Epidemiology Group (SEEG)
University of Oxford

www.map.ox.ac.uk