Geographic Profiling for Epidemiology
Application to COVID-19 Outbreaks

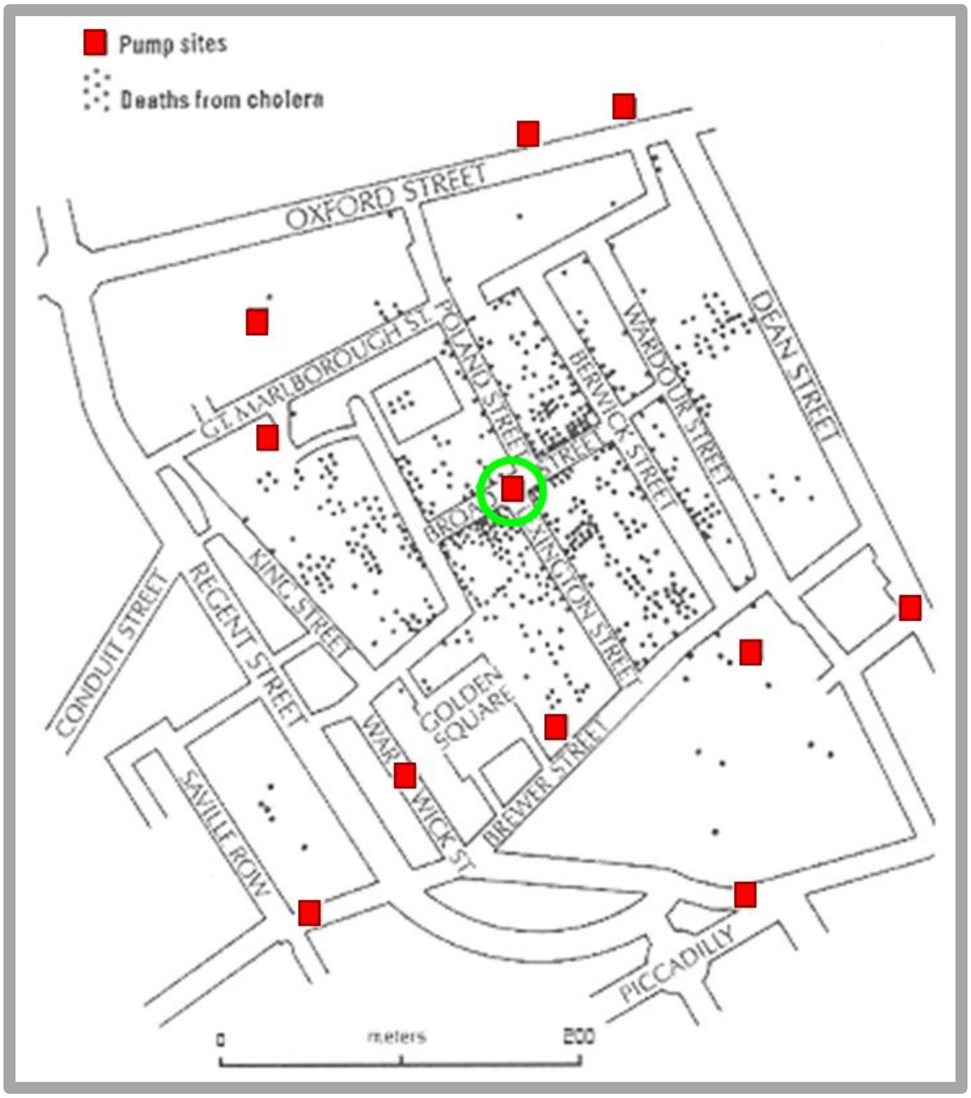
While geographic profiling analysis is best known for law enforcement, it is also useful in other applications involving finding an unknown location by analyzing the geographic pattern of known locations connected to the unknown location. One such application is in Epidemiology, specifically in identifying the common source of a local disease outbreak from the pattern of home locations of the victims.

The current rapid spread of the COVID-19 virus throughout the world calls for the use of every tool available to help slow the spread. The virus usually appears with a local outbreak involving a limited number of cases in one area. A priority is to identify the source of that outbreak so that it can be shut down and those potentially exposed can be notified. Geographic profiling is one of the tools that can be useful for this purpose. This briefing document gives some of the relevant background and references.

A Little History - The 1854 London Cholera Epidemic

In 1854, John Snow, a British medical doctor, performed the first known geographic profiling analysis for the purpose of medical epidemiology.

As early as 1849 he proposed that the "cholera Poison" was spread through the contamination of food or water. This theory was opposed to the more commonly accepted idea that cholera, like all diseases, was transmitted through inhalation of contaminated vapors. But Snow had no way to prove his theory until 1854, when cholera struck England once again.



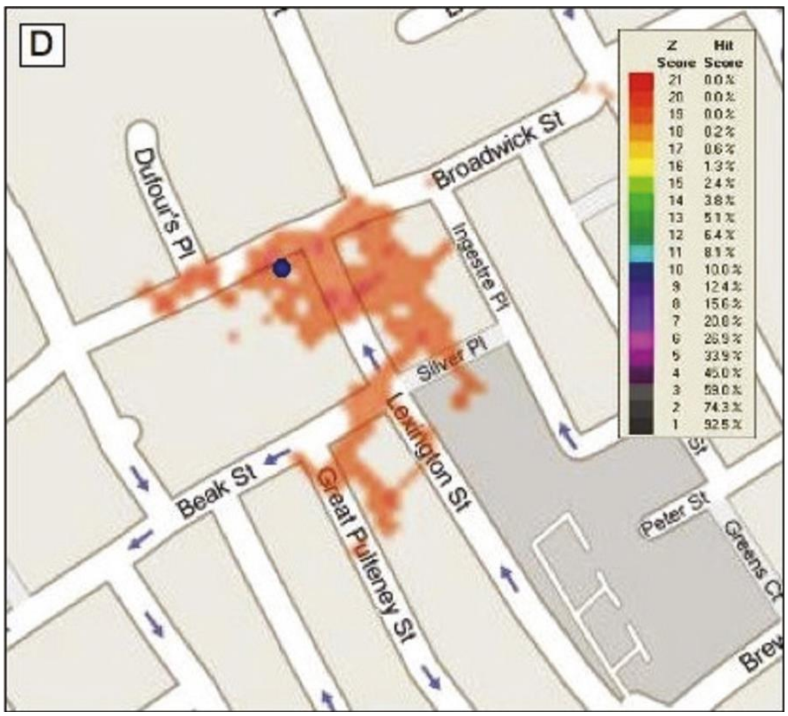
London cholera outbreak of 1854 with public water pump locations

In investigating the epidemic, Snow began plotting the location of deaths related to cholera. Snow found that in one particular location near the intersection of Cambridge and Broad Street (as they were then known), up to 500 deaths from cholera occurred within 10 days. Looking at the locations of public water pumps in the area, one was located in the heart of the most affected area.

After the panic-stricken officials followed Snow's advice to remove the handle of the Broad Street Pump that supplied the water to this neighborhood, the epidemic was contained. At the time, London was supplied its water by two water companies. One of these companies pulled its water out of the Thames River upstream of the main city while the second pulled its water from the river downstream from the city. A higher concentration of cholera was found in the region of town supplied by the water company that drew its water from the downstream location. Water from this source could have been contaminated by the city's sewage.

Through mapping the locations of deaths related to cholera, Snow was able to pinpoint one of the major sources of causation of the disease and support his argument relating to the spread of cholera. This is an early example of using a map of connected events to narrow the focus for a cause to the event, which we now know as geographic profiling.

A modern geographic profile of the home locations of the victims would produce a highlighted peak profile area calculated to have a 67% chance of containing the source of the outbreak.



Modern geographic profile of the 1854 cholera outbreak
Peak profile area with Broad St. pump marked

Theory of Geographic Profiling

Geographic profiling works by dividing the map of the entire area into a fine grid, and then adding up the "travel cost" from each known location to each grid cell, relying on the enormous computing power of modern computers to make feasible what could not be done manually. Grid cells are ranked in order of their score, with lowest travel cost (or "hit score") being most likely. Any suspect locations are ranked for likelihood of being the common origin point according to the score of their grid cell. The results are normally presented as a color-coded probability map, and a list of provided "suspect" sites ranked by profile hit score.

Geographic profiling relies on certain assumptions about the connected events being profiled:

1. A person has physically traveled between a home location and the event site, one of which must be a known location. Either the home site or the event site may be the unknown location.
2. All of the events included have the same unknown home site in common, or vice-versa, all the home sites must have the same unknown event site in common.
3. The journey should in general obey certain common behavioral patterns, such as rational choice and least-effort principle (for example, tendency to choose the shorter path or the nearer goal).
4. A substantial majority of the connected sites are known, or if only a smaller fraction are known, there must be no geographic bias to their pattern (for example such a bias would occur if the known sites are close to a border, and there may be many more unknown connected sites on the other side of the border).
5. There are sufficient known sites to support an analysis with useful probability. In general this is at least 5 sites that are reliably connected. With larger numbers, some degree of uncertainly and error in identifying connected sites can be tolerated.

These are not necessarily hard rules – for example it is possible to handle multiple origin sites in one analysis, and some degree of uncertainty and error in linking events is expected. Expert geographic profiling analysts are trained to work around these issues. More about geographic profiling:
http://www.ecricanada.com/Rigel%20Geographic%20Profiling/

http://rigelanalyst.net/iaca2013/Geographic%20Profiling%20Theory%20and%20Practice%20Tutorial.pdf

Tracing the source of local outbreaks of viral infections such as COVID-19

Geographic profiling can be used to locate likely origins of infection for local disease outbreaks, similar to Dr. John Snow's early application to the 1854 London cholera epidemic.

The basic requirements are:

1. The home address of each person who became infected is known.
2. The majority became infected from the same source location (for example a school, or a restaurant). Usually a cluster of cases is identified by time period and area, but it may be necessary to do some linkage analysis if there are multiple outbreaks from multiple sources close together in space and time.
3. The source of the infection is a location which the patients have traveled to physically (not someone who visited them)

If these requirements are met, then a geographic profiling analysis can help identify the most probable area where the source lies, and the most likely source sites ranked in order of probability if there is a set of suspect sites.

The process involved is very similar to criminal geographic profiling. The home locations of infected patients in the local area who are considered to be linked to a common infection source by space and time constraints are entered into the geographic profiling program, which then computes a geographic profile. In the case of a human disease outbreak, investigators would likely interview the patients and take their activity history for comparison, assuming the patients remember and are well enough to communicate it. It may be possible to list some common visit sites amongst the patients, in which case that will likely be where to start to identify the source of the local outbreak. It may be necessary to prioritize those sites for investigation given limited investigative resources, in which case the geographic profile will provide a likelihood ranking for those sites based on location. The geographic profile can also suggest the area to look in when there are no obvious common sites in the patient histories.

For More Information:

Geographic profiling as a novel spatial tool for targeting infectious disease control

Steven C Le Comber1*, D Kim Rossmo2, Ali N Hassan3, Douglas O Fuller4 and John C Beier5

. International Journal of Health Geographics 2011

http://www.ij-healthgeographics.com/content/pdf/1476-072X-10-35.pdf

Abstract

Background: Geographic profiling is a statistical tool originally developed in criminology to prioritise large lists of suspects in cases of serial crime. Here, we use two data sets - one historical and one modern - to show how it can be used to locate the sources of infectious disease.

Results: First, we re-analyse data from a classic epidemiological study, the 1854 London cholera outbreak. Using 321 disease sites as input, we evaluate the locations of 13 neighbourhood water pumps. The Broad Street pump -the outbreak’s source- ranks first, situated in the top 0.2% of the geoprofile. We extend our study with an analysisof reported malaria cases in Cairo, Egypt, using 139 disease case locations to rank 59 mosquitogenic local watersources, seven of which tested positive for the vector Anopheles sergentii. Geographic profiling ranks six of theseseven sites in positions 1-6, all in the top 2% of the geoprofile. In both analyses the method outperformed other measures of spatial central tendency.

Conclusions: We suggest that geographic profiling could form a useful component of integrated control strategies relating to a wide variety of infectious diseases, since evidence-based targeting of interventions is more efficient, environmentally friendly and cost-effective than untargeted intervention.