Putting Where Back into Epidemiology

Richard Hoskins

All epidemiology courses teach the obvious threesome about a disease investigation or study: who, when, and where. Epidemiology students proceed through the famous John Snow and cholera outbreak in London, 1854, which led to the locking of the Broad Street pump and the immediate decline in cholera incidence. Snow had no idea about the what with regard to cholera, as bacteria (in this case, *Vibrio cholerae*) had not been discovered. This story remains a marvelous piece of detective work that saved lives and resulted in a methodological advance that revolutionized public health. It became clear that the miasma (bad air), the famous London fog, was not responsible for the disease. Clearly, sewage-contaminated water was somehow in the causal pathway.

Snow definitely collected the who, when, and where. His famous map of cholera cases clustered around the Broad Street pump is an icon for public health practitioners. Until recently the where in epidemiology has languished and even, to a great degree, disappeared from general public health practice except for some outbreak investigations made with some pins and a paper map from the public works department.

GIS makes mapping easy

But that neglect of where has all changed. Geographic information systems (GIS) now allow ordinary public health folks to routinely make maps and carry out a descriptive spatial analysis at any scale appropriate for their problem: nation, state, county, a census region, or street. The limitation is, as with all studies, data availability and quality. Increasingly, health departments are geocoding their data (turning addresses into longitude and latitude). These data, with strict rules about their use, are becoming available to the public health community. All the census data and most geographic features can be found for no cost on the Internet or one's local spatial data center, which all states and most counties have.

The following example illustrates how a GIS analysis may offer some insight into what geographic information may contribute to understanding the incidence of infant death. Let’s start with a question: Is there any relationship between residential proximity to Interstate 5 and infant death rates? Earlier studies indicate that there might be. This is a “John Snow map” problem. The Broad Street pump in this study is I-5.
The map on the left shows a portion of the Interstate 5 corridor through Tacoma, which is in the study area—Snohomish, King, and Pierce (SKP) counties. The infant death rate for SKP for 1989-2001 was 5.92 per 1,000 live births. We'll take this figure as the comparison rate. Using a GIS, we'll build in two ½ kilometer-wide bands, or buffers, on each side of I-5. On the map, I-5 runs through the middle of the buffer zone. Using the geocoded birth data (not shown on the map), we count the number of births in each band; the total number of births is the denominator. Then using the geocoded, linked birth-death records (shown on the map), we count the number of infant death cases; the total number of infant deaths is the numerator (total infant deaths/total births=death rate). When we calculate and map the rate, we find that the inner buffer rate is 21% higher than the comparison rate (SKP rate) and the outer buffer, 26% higher. Both rates are statistically significant; in this case the likelihood that the measured rates are due to chance is less than 5%.

This preliminary result appears compelling and might support other studies indicating that proximity to high volume traffic is a health hazard for children. However, here is where the complications start. For one thing, using all infant deaths is too broad. Most infant deaths occur in the neonatal period (first 28 days) and most of those in the perinatal period (first week). It is not clear how traffic volume would affect those children. After 28 days, sudden infant death syndrome (SIDS) is the most prevalent cause of death. Since it is easy to make a map of selected causes of death, we could refine the original question and “ask the map” for more specific information.

Proceeding from the map, which represents just the beginning of formulating a question and is still far removed from framing a hypothesis, we wonder: is the location near I-5 just a proxy for some other risk factor such as low or high maternal age, or limited access to prenatal care, or is it a spatial factor (confounder) that distorts the association between I-5 and infant mortality? In other words, is location getting in the way of understanding something else that is going on? Or perhaps proximity to I-5, with the resulting air pollution, really is a risk to infants.

This example shows that, with just the linked infant birth-death, live birth records, and GIS technology, in a matter of just an hour (after the data is geocoded) an investigator can get a description of what is occurring near I-5. As always with a descriptive study, more questions appear than do answers. Notice the concentration of infant deaths just northwest of the center of the map. Are there just a lot of births there or is there really a cluster?

**Careful thinking still required**

It is possible to do a similar analysis around other, or all, highways, toxic waste sites, power lines, or specific industrial sites. However, making maps so easily should be considered with caution. Public health data that appears with tables and charts has always required clear and careful thinking to develop a valid conclusion or useful study. Throwing maps into the mix adds another dimension of complexity. Now we need clear and careful geographic thinking, too.

This preliminary analysis, which puts the where back into epidemiology here in Washington State, now comes full circle from London, 1854, updated with GIS, but the thinking required to understand what a map tells us is still the same.

**Author**

Richard Hoskins is the state public health geographer and a senior epidemiologist in the Office of the Assistant Secretary for Epidemiology, Health Statistics and Public Health Laboratories in the Washington State Department of Health. He is also on the clinical faculty in the Department of Epidemiology at the UW SPHCM.

**Resources**


