Public Health Services Assessments Using ArcView and Matlab

By Dr. Craig H. Newborn II, Todd C. Burwell, Yenenesh A. Andrews, and Janelle L. Blazier

Abstract: *We present a process which integrates a relational database with the mathematical capabilities of Mathworks® Matlab and the GIS analysis capabilities of ESRI® ArcView. The process takes the data stored in the database and allows an analyst to perform statistical and GIS analysis using two state of the art software programs. We present examples of the data and analysis related to global demographic health analysis, public health assessment, and services funding for HIV/AIDS. We use public health data and sociological measures to categorize health care systems and services. We also discuss how the selection of indicators affects clustering and demonstrate the difference between geographical clustering and clustering based on statistical analysis. This process shows how thematic mapping coupled with statistics provides unique insight into public health services assessments.*

The Comparative Assessment Model (CAM) was developed to assist analysts with the subjective assessment of national health care systems around the world. It consists of three fully integrated parts: a database of over 40 health indicators from seven sources, Matlab® based analysis and plotting tools, and ArcView® mapping and geographical analysis. All database functionality is built into the Matlab tools, so the data can be stored in either an Oracle® database or an MS Access® database without any noticeable changes. The model provides the ability to view the data directly or to plot it in a wide array of formats. From correlations and trend analysis to ranking global health care based on chosen indicators, CAM provides specialized functions for health care analysis.

One of the most attractive features of CAM is the fact that all of these abilities are seamlessly linked to ESRI's ArcView mapping software. Utilizing ArcView's application programming interface (API), CAM is able to open the software and load the appropriate shape files and data layers. Once the user has performed their analysis and isolated the data of interest, they are not required to create any type of special file to use ArcView. At the push of a button, ArcView is opened, populated, and fully operational.

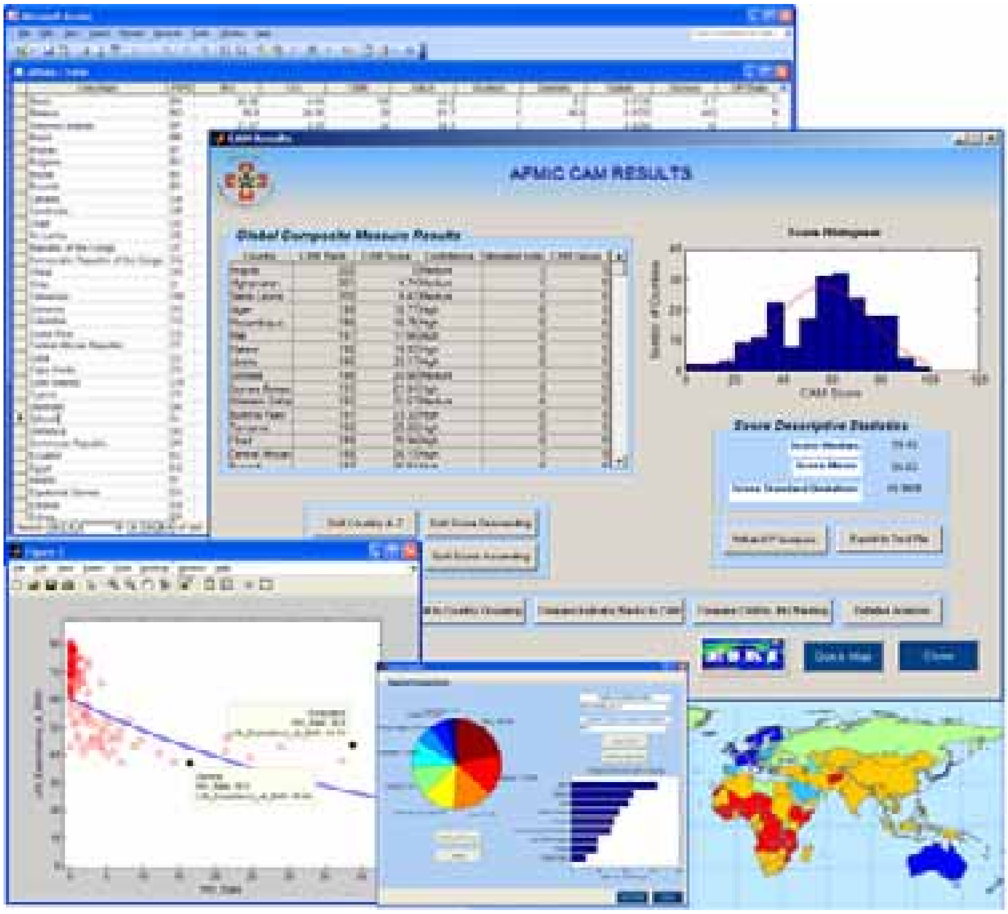
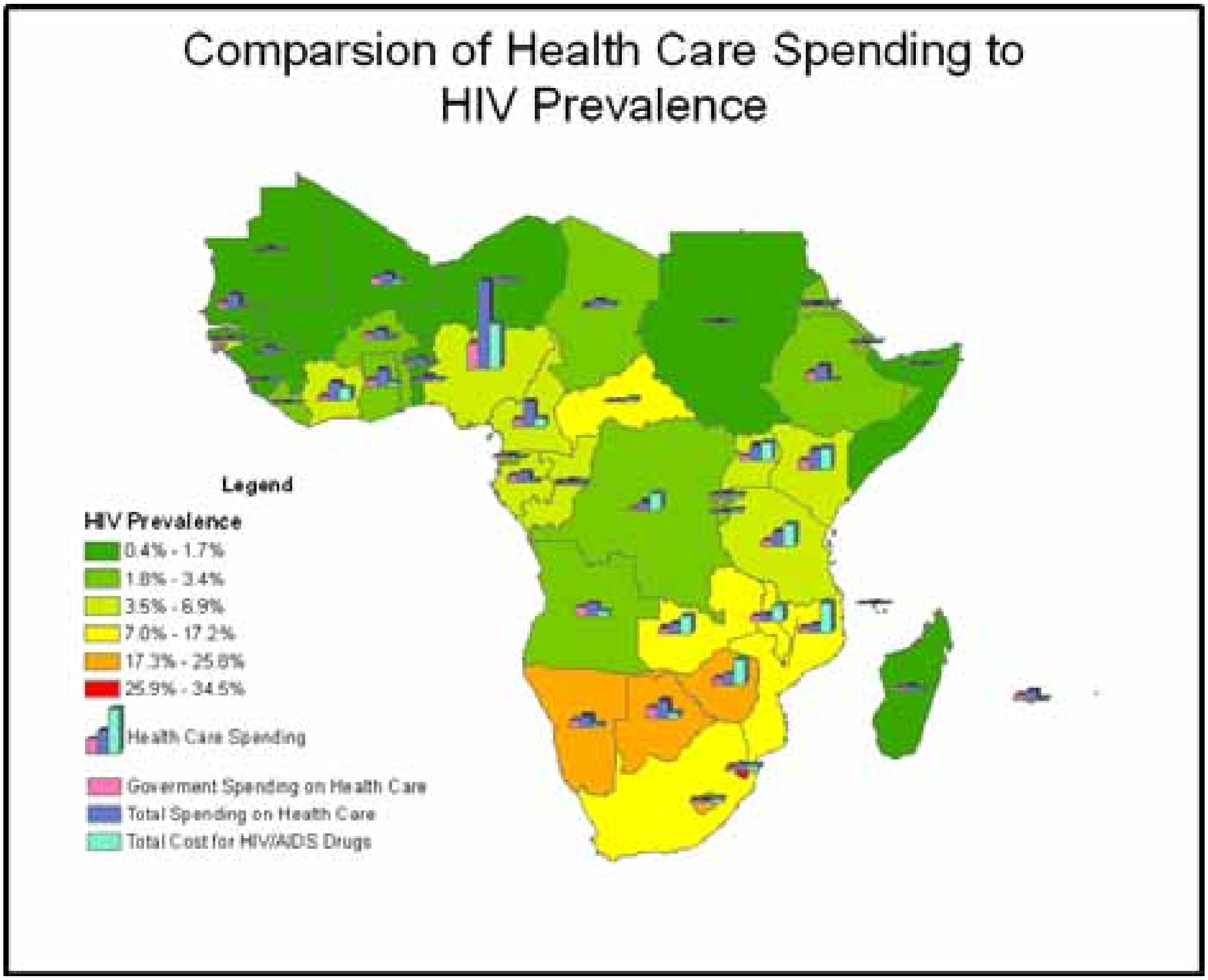


Figure 1. CAM seamlessly integrates a database, analysis tools, and GIS mapping.

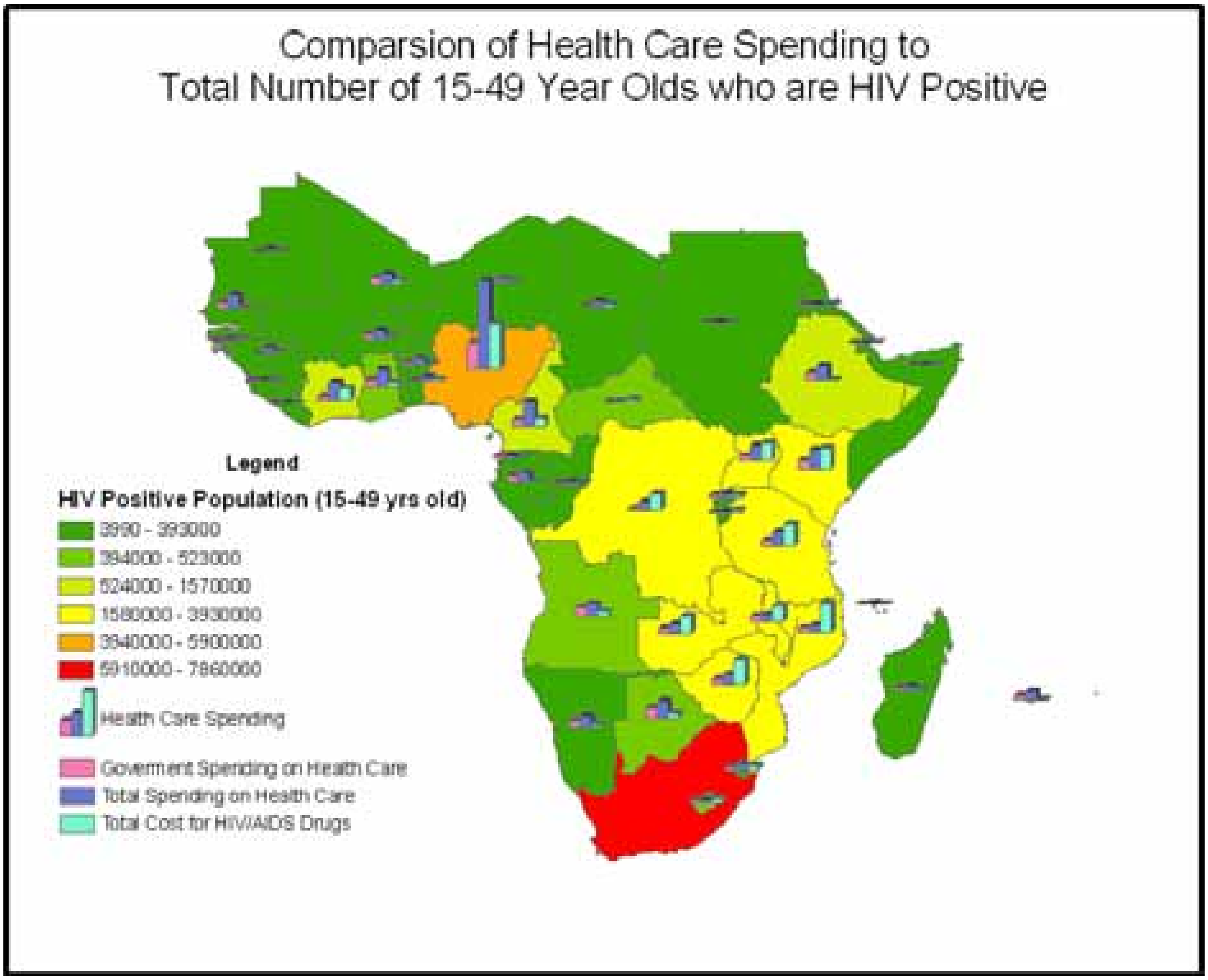
The ability to integrate a database, analysis software, and GIS makes CAM a powerful tool for studying global health issues. Beyond the statistical analysis CAM provides, mapping allows the user to consider regional effects and the influence of geography. Even a simple map can quickly reveal if there is a regional influence on an indicator. Services funding for HIV/AIDS provides a stark example of the benefits to visualizing data using GIS.

Health Care Funding and HIV/AIDS Measures

HIV prevalence is one of the most publicized measures in the on-going effort to combat the HIV/AIDS epidemic. As a ratio, it normalizes the data and provides a way to compare the effects of the epidemic on different countries and various segments of particular populations. However, for anyone who doesn't dig deeper, HIV prevalence can mask the enormity of the challenges faced by many countries especially in sub-Saharan Africa. This brief analysis doesn't dispute any other findings; it is intended to demonstrate how the effective use of maps can enhance analysis and communication by providing a more complete picture of a situation.



Map 1. HIV Prevalence shows no relationship to the yearly cost for HIV/AIDS drugs.



Map 2. The actual number of people needing HIV/AIDS drugs drives cost.

On Maps 1 and 2, the same financial data is represented by bar charts overlaid on shaded maps. (In each case, the financial data for South Africa is excluded. Health care spending in South Africa is so much greater than the rest of sub-Saharan Africa that it reduces the other bars to nothing.) The bars represent how much the national government spends on health care per year, the total spending on health care per year, and the projected total cost for providing HIV/AIDS drugs to the affected population.

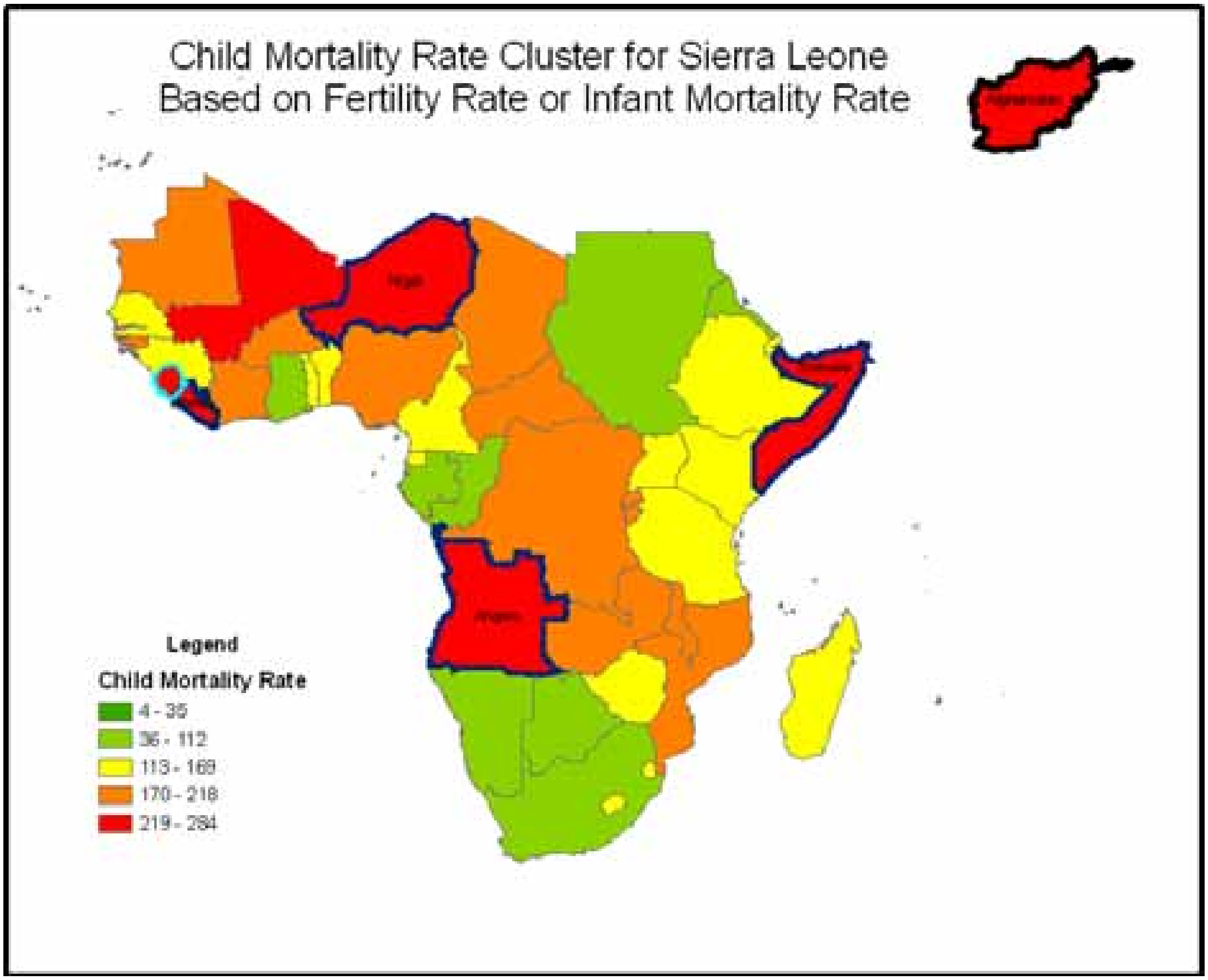
The shading in Map 1 corresponds to HIV prevalence. The five countries with the highest prevalence are all found in southern Africa, with the highest prevalence being found in the tiny nation of Swaziland. Except for Zimbabwe, which is plagued by a host of compounding problems, the countries with the highest HIV prevalence have total health care spending well in excess of the total cost for HIV/AIDS drugs. However, in eastern central Africa, there is collection of seven countries with relatively low HIV prevalence but the total cost of HIV/AIDS drugs soars well above the total spending on health care. Despite the tragic reality of HIV prevalence in excess of 20%, HIV prevalence is clearly not the best indicator for financial need.

Map 2 places the same financial data on a shaded map representing the number of HIV/AIDS cases in a given country. Here it can be said that "a picture is worth a thousand words." Even though Kenya, Uganda, the Democratic Republic of Congo, the United Republic of Tanzania, Zambia, Malawi, and Mozambique all have relatively low HIV prevalence, they have very large populations with HIV/AIDS. On the other hand, Namibia, Botswana, Lesotho, and Swaziland have relatively low numbers of people with HIV/AIDS. Thus, despite high HIV prevalence, it requires much less help for these countries to provide the necessary drugs, while countries with low HIV prevalence are unable to buy sufficient drugs even if they spend every single health care dollar on them.

It is easy to focus on prevalence in our fight with HIV/AIDS, but it is likely that the war will be won or lost elsewhere. We need to pay close attention to the low prevalence countries with large expanding populations. HIV does not spread like influenza, but it is a virus, and if we fail to pay sufficient attention millions of carriers in populace countries simply because HIV prevalence is low, the results could be devastating.

Categorizing Health Care Outcomes based on *k*-means Clustering

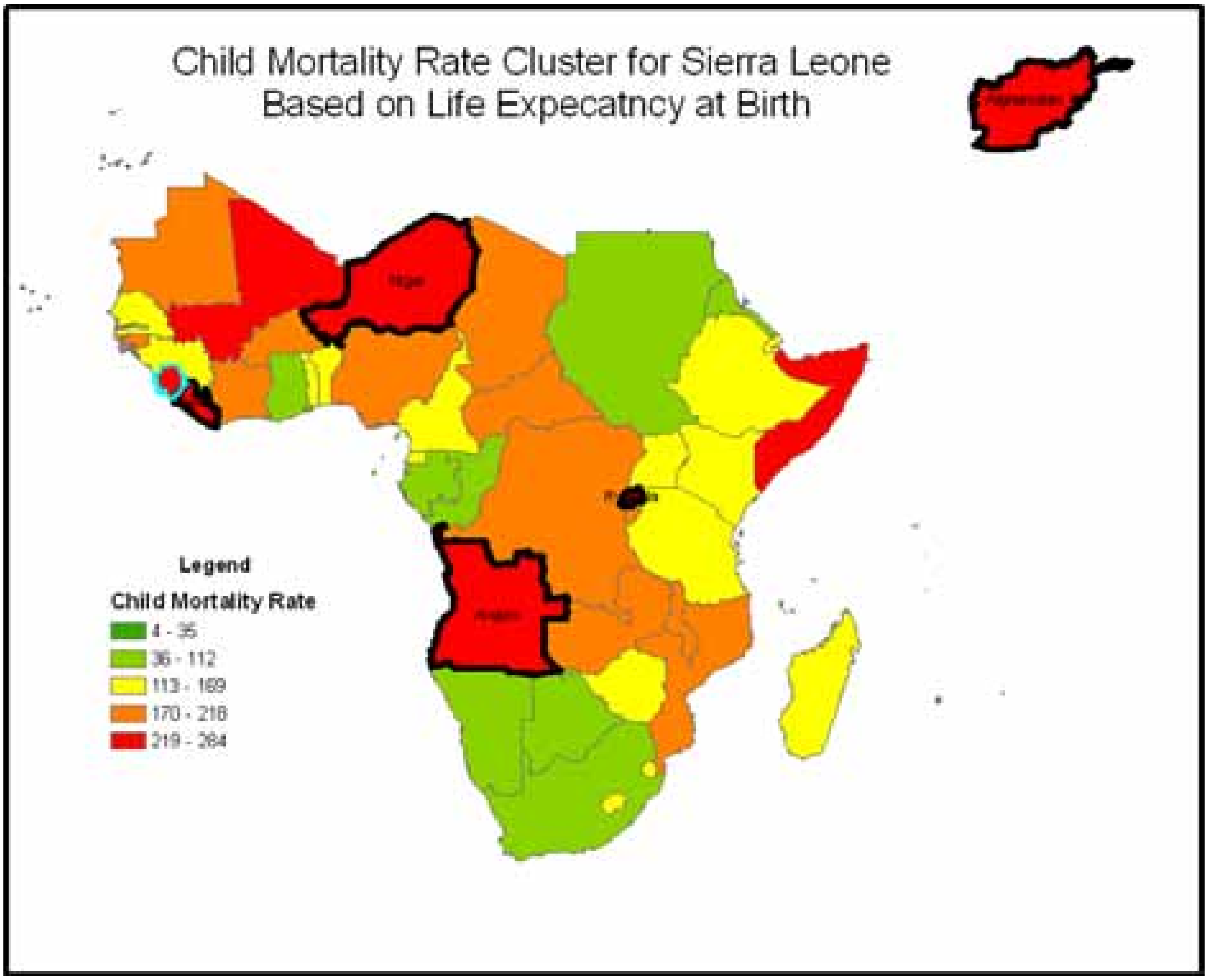
It is not uncommon to be interested in determining which countries are similar based on demographic, economic, or public health data. The similarities between countries can be used to group countries based on specific parameters of interest or help determine which countries can serve as surrogates for missing or suspect data. In many applications when data is missing, it is not uncommon to use the geographic nearest neighbors to determine surrogates for the missing data. Although this method can be of value, *k*-means clustering based on data of interest to the analyst can provide unique insights.



Map 3. Fertility Rate and Infant Mortality Rate select the same surrogates for Siera Leone.

Consider Maps 3 and 4. Here Sierra Leone is highlighted in light blue. The World Health Organization reports a child mortality rate of 284/1,000 for Sierra Leone in 2003. When using the geographic neighbors of Liberia, Guinea, and Guinea-Bissau as surrogates to estimate the child mortality rate of Sierra Leone one obtains a value of 199.7/1,000. When surrogates are determined based on Euclidean distance clusters for fertility rate or infant mortality rate, the countries of Afghanistan, Angola, Liberia, Niger, and Somalia are selected as the five most similar countries to Sierra Leone. These countries provide an estimate of 247.8/1,000 for Sierra Leone's child mortality rate. Map 4 shows a similar result for a cluster based on life expectancy at birth. For this cluster Angola, Afghanistan, Liberia, Niger, and Rwanda are selected as the five most similar countries. This cluster gives an estimate of 243.3/1,000 for the child mortality rate of Sierra Leone.

When a weighted average based on the level of similarity between the data for the selected countries is applied to Sierra Leone the estimate for child mortality rate increases to 252.5/1,000 for the fertility rate cluster, 247.2/1,000 for the infant mortality rate cluster, and 253.9/1,000 for the life expectancy at birth cluster.



Map 4. The Life Expectancy at Birth cluster selects war-torn countries with low HIV prevalence.

It is interesting to note that Rwanda shows up in the fertility rate and infant mortality rate cluster when the ten most similar countries are selected. Somalia shows up in the life expectancy cluster when the eight most similar countries are selected. Even more interesting is the additional insight one obtains from the clusters. The countries in all three clusters are nations suffering some type of internal conflict. Although Sierra Leone's life expectancy at birth is very low (42.8, WHO 2003 country data) in the range of the high HIV prevalence countries, all the countries in the clusters have HIV prevalence rates below 7.0%. The *k*-means clustering brings out the fact that low life expectancy at birth in Sierra Leone is not related to high HIV prevalence. Analysis of the countries in the clusters points to internal conflict as a factor for the high child mortality rates and low life expectancies.

Conclusion

The effective use of GIS is a powerful tool for analysis and communication. The geospatial representation of data allows an analyst to identify regional trends or the differences between neighboring countries. It can provide insight into cultural indicators that cross boundaries and national interests that are defined by borders. Making data geographically relevant in a visual way changes how individuals look at the world. By integrating ESRI's ArcView with analysis tools and an indicator database, CAM provides unique insight into public health services assessments and informative images for communicating those assessments.

References

The data displayed in this article is based on data from the World Health Organization, www.who.org.

Author Information

Dr. Craig H. Newborn II, craig.newborn@dynetics.com
Todd C. Burwell
Yenenesh A. Andrews
Janelle L. Blazier

Dynetics, Inc.
P.O. Box 5500
Huntsville, AL 35806-5500

256-922-9230 – phone
256-964-4038 – fax