Malawi, 1994-2010: Analyzing HIV/AIDS Trends Using Geographic Information Systems

Eliza Johannes

The Problem

Geospatial analytical methods, including geographic information systems (GIS), provide a critical method for understanding the nature and causes of spatial variation in HIV prevalence. GIS tools can enhance the ability to address public health problems and better inform the planning, implementation and assessment of interventions in affected areas. However, the utilization of GIS has been limited in sub-Saharan Africa in part due to scarcity of reliable spatially coded data. Malawi—one of the six low-income countries with the highest HIV prevalence rates globally—is the case study for this article.

The longer it took to travel to the nearest public transport, the higher the prevalence of HIV... Consequently, HIV prevalence decreased with mean distance from the main roads.

Malawi has made great strides in decreasing the number of HIV/AIDs victims. The 2012 global AIDS report ranked Malawi as one of 25 countries with declines of 50 percent or more in new cases of adult (aged 15 to 49) HIV infection globally. It is also among 32 countries to report a 25 to 49 percent decline in HIV deaths (2005-2011). Antiretroviral therapy also showed drastic improvement between 2001 and 2011. HIV prevalence, however, still presents a significant challenge (Nykiforuk and Flaman 2011).

An estimated 923,058 of Malawi's 14 million people were living with HIV/AIDS in 2010. Prevalence among sex workers was up to 70 percent, and nearly 600,000 children were orphaned as a result of HIV/AIDS. Disparities regarding who was affected and where were also glaringly apparent: HIV prevalence is more than 2 times higher among female than male youth aged 15 to 24 and almost 3 times higher among urban women than those in rural areas (22.7 percent versus 10.5 percent) (National AIDS Commission (NAC) 2011).

Spatial statistical tools were used to examine spatiotemporal trends in HIV prevalence between 1994 and 2010 and to identify and map the spatial variation/clustering of factors associated with HIV prevalence at the district level to determine "hotspots" and "coldspots" for potential targeting of interventions. HIV prevalence data for this analysis were gathered primarily from pregnant women attending antenatal clinics (ANCs).

Most GIS-based analyses of HIV/AIDS in Africa have largely been either at the coarse continental scale (which has limited national policy relevance) or at fine local scales requiring heavy data collection and having limited spatial scope. Despite increased use of such analyses in Africa, few have analyzed the spatiotemporal variation and clustering of HIV prevalence or its drivers at the lower (or meso) scales where such methods can improve understanding and enhance effectiveness of interventions. While analysis at this level can present challenges, ANC data are a good source of information and more readily available allowing for lower scale analysis. ANC data allow for longitudinal trend analysis and can be rescaled to district or other scales by using spatial interpolation techniques.

Drivers behind HIV epidemics are numerous and diverse and differ over time and space, making addressing such a public health problem challenging. In the case of sub-Saharan Africa, the proximate cause of the high incidences of HIV is well documented —unprotected sex with an infected person. However, recent research has shown that there are underlying drivers that merit more analysis to better combat the epidemic. These drivers include cultural, historical, socio-economic, demographic, and geographic factors—all of which contribute to the vulnerability of particular groups of women and men who engage in risky sexual behavior. For Malawi, identified underlying HIV drivers include low literacy, high poverty, gender inequity and low social and economic status of women, high rates of unprotected casual and transactional sex, and low male and female condom use, among others.

The 19 studied ANCs were particularly useful in that they provided a longitudinally rich network of surveillance material that was gathered from 1994 to 2010. The network nearly tripled in 2007 to 54 clinics although the original 19 were maintained for continuity and spatiotemporal analysis purposes. Data were collected annually from 1994 to 1999, then biennially, and included the latitude and longitude coordinates of the ANCs to allow for mapping. HIV prevalence rates were the only consistent data available for all years. Although this is considered a limited source of information, it remains the major source of HIV data in Malawi.

To determine spatial dependence in HIV prevalence, spatial interpolation, and spatiotemporal trends, HIV prevalence rates for pregnant women aged 15 to 49 were plotted. This allowed for a broad, multi-scalar, spatiotemporal perspective of HIV epidemics across the national, regional, urban, and rural levels. GIS tools such as ArcGIS desktop 10.0 empirically tested for spatial dependency in HIV prevalence at the national scale. The result was a continuous surface of HIV at 1x1 km spatial resolution generate prevalence estimates at the meso scale. This visualization enabled cluster/hotspot and regression analysis.

To produce a smooth surface for visualization and data generation at the district level, the Inverse Distance Weighted (IDW) spatial interpolation method was employed for the selected years. The variable setting of 6 to 10 points was used to predict values at each unknown location. GIS tools were then used to extract HIV estimates for the 31 "districts" (27 of Malawi's 28 districts and four main cities of Blantyre, Lilongwe, Zomba, and Mzuzu) by averaging prevalence values in constituent 1x1 spatial cells. Local measures of spatial association used within ArcGIS 10.0 allowed for "where the clusters or outliers are located" and "what type of spatial correlation is most important" (Tanser, et al. 2009). Anselin's Local Moran's I (Moran 1950) was employed to detect core clusters or outliers of districts with extreme HIV prevalence values not explained by random variation. These clusters were then classified into hotspots, coldspots, and spatial outliers.

Multiple regression analysis of HIV prevalence for 2010 determined and illustrated potential linkages of indicative drivers of observed spatial variation to particular hotspot/coldspot clusters of districts. Spatial distribution and clustering of selected factors were mapped. HIV prevalence per district among pregnant women attending ANCs served as the dependent variable. Multiple independent variables such as socio-demographic (e.g., education, poverty/wealth/consumption, and employment), HIV awareness and behavior, and geographic variables such as access to HIV related amenities were chosen based on literature and availability of the data at the district level. Thirty-seven independent variables were identified and subsequently narrowed down via a multi-stage statistical screening process in correlation analysis to 18 significant variables. A stepwise regression analysis further reduced the variable pool to 13.

The result of this analysis revealed several temporal and spatial trends in HIV prevalence at the national and regional scales. First, prevalence among pregnant women attending ANCs has significantly declined since the peak epidemic in 1999. Additionally, analysis revealed multiple geographically defined HIV "epidemics" with diverse spatiotemporal trends. The southern region consistently had the highest HIV prevalence—7.0 percent higher than the national prevalence (1996 and 2007). The northern region had the lowest but essentially merged with the central region's from 2003. Urban areas typically had higher and more severe epidemics whereas rural areas were less intense and slower/lower. Yet research revealed the intensity of these two epidemics has been converging from a 2.8-fold difference in 1995 to 1.5-fold by 2010.

Presence of a spatial structure was confirmed although there was significant temporal variation in spatial dependence including in the early years (1995, 1996, 1999). The HIV epidemic in prevalence and spatial extent intensified from 1994 to 1999 but declined from above 25 percent in 1999 to below 12.5 percent nationally by 2010. Spatial analysis revealed that core clusters of hotspots were present in much of the southern region. To determine potential drivers for HIV prevalence for 2010 and their local spatial patterns, Ordinary Least Squares (OLS) regression was used to produce four models. The fourth model was chosen as the "best" based on explanatory power and collinearity diagnostics.

The model revealed that the longer it took to travel to the nearest public transport, the higher the prevalence of HIV at the district level. Consequently, HIV prevalence decreased with mean distance from the main roads. These findings are consistent with other studies that have found an inverse relationship between HIV prevalence and distance to main roads. The only behavioral explanatory factor for HIV prevalence was the proportion of women who had ever taken an HIV test, which proved to be the second most influential factor and was positively associated with HIV prevalence. However, more nuanced analysis on HIV testing behavior determined that education was negatively correlated with HIV prevalence, suggesting that the more education pregnant women receive, the less likely they are to have HIV/AIDS and that some level of education is good for AIDS prevention.

This analysis utilizes GIS tools to challenge the logic and effectiveness of uniform policies and interventions given the unique circumstances of each area. It reveals that HIV epidemics are an aggregation of several spatially defined sub-epidemics (national, regional, urban, rural and local clusters) each of which demands its own specific and unique course of action given its context. One of the most prominent temporal trends to emerge was the general decline in HIV prevalence after the rise in 1999. This reflects Malawi's intense efforts to allocate human and financial resources to combat the epidemic. Another significant trend was the slow but developing spatial evening in HIV prevalence as the epidemic stabilized and declined after 1999. Cluster analysis revealing the southern region as an HIV epicenter was also a significant finding. The prevalence of HIV in the southern region could be due to its history of urbanization in addition to being the most densely populated, having the highest levels of rural poverty, and a prevalence of syphilis.

Our analysis revealed much with regard to analysis of spatial clustering and spatiotemporal trends of HIV/AIDS prevalence using GIS. However, there are limits to the study. More detailed analysis must be done to focus on local HIV variation at a finer spatial scales and the drivers of HIV incidence. Such research could better capture underlying factors of HIV prevalence such as HIV prevention knowledge, personal resources, and socio-economic access or exposure to health services to inform public health policy decisions. Additionally, further research could determine spatial limits of the spatial dependence of HIV on neighboring values at different spatial scales to better inform interventions among other aspects.

Our work represents one of the few spatially explicit longitudinal analyses of HIV prevalence to date though a growing number of researchers are attempting similar research. Acknowledging its limitations, our analysis demonstrates the need for more spatial analysis to understand the geographic nature of epidemics, examine trends, and use this knowledge for more targeted interventions to combat HIV/AIDS in affected regions.

Dr. Johannes *is a Research Staff Member in IDA's Intelligence Analyses Division. She holds a Doctor of Philosophy in education, policy studies from the University of Illinois at Urbana-Champaign.*

The original article was published in *BMC Infectious Diseases*, May 2014. Coauthors were Leo C. Zulu, affiliated with the Department of Geography, Michigan State University, and Ezekiel Kalipeni, affiliated with the Department of Geography, Colgate University.

"Analyzing Spatial Clustering and the Spatiotemporal Nature and Trends of HIV/AIDS Prevalence Using GIS: The Case of Malawi, 1994-2010"

http://www.biomedcentral.com/1471-2334/14/285

References

Moran, PAP. 1950. "Notes on Continuous Stochastic Phenomena., 37." *Biometrika* 37: 17–33.

National AIDS Commission (NAC). 2011. "HIV and Syphilis Sero–Survey and National HIV Prevalence and AIDS Estimates Report for 2010." Lilongwe, Malawi.

Nykiforuk, CU, and LM Flaman. 2011. "Geographic Information Systems (GIS) for Health Promotion And Public Health: a Review." *Health Promot Pract* 12: 63-73.

Tanser, F, T Bärnighausen, GS Cooke, and M L Newell. 2009. "Localized Spatial Clustering of HIV infections in a Widely Disseminated Rural South African Epidemic 38:1008–1016." *Int J Epidemiol* 38: 1008-1016.