

Small Area Estimation of HIV Prevalence using National Survey data in South Africa

Kondlo, Lwando* and Manda, Samuel

Medical Research Council Biostatistics Unit

1 Soutpansberg Road

Pretoria 0001, South Africa

lwando.kondlo@mrc.ac.za* and samuel.manda@mrc.ac.za

Introduction

Sub-Saharan Africa is more heavily affected by HIV/AIDS than any other region of the world. An estimated 22.5 million people are living with HIV in the region - around two thirds of the global total. In 2009, an estimated 5.6 million people were living with HIV/AIDS in South Africa (SA), more than in any other country in the world (UNAIDS, 2010). In SA, data collected from national surveys for HIV provide the best available source of information for HIV epidemic at the national level. Such data have been used to understand the variation of HIV infection within a country; essential for determining where prevention and treatment programs need to focus resources; and to effectively target areas in need and to monitor intervention programs. However, these surveys do not provide sufficient data on which reliable estimate of HIV at small area (e.g., local and district municipalities) can be based. Reliable statistics on HIV are needed at small area levels for effective control and prevention of HIV policies and programs. In this effort, it is desirable to estimate HIV prevalence at small area level.

There are two national prevalence surveys that provide HIV statistics in SA, i.e., antenatal sentinel and household-based surveys. The antenatal sentinel prevalence estimates are drawn amongst pregnant women from antenatal clinics and have been shown to be a good proxy for prevalence in the overall population when the HIV is generalized, however, there are known limitation The survey does not include men, children, and in particular women who do not attend antenatal care. Furthermore, the sampled women are from public clinics; thus mostly are poor. Therefore, the sampled data from antenatal surveys cannot be used to estimate lower level HIV prevalence due to its representativeness and randomness in the country. The national household-based survey for HIV involves sampling a proportional cross-section of society, including a large number of people from each geographical, racial and other social group. The results are then adjusted to correct for possibly over- or under- sampling. The household-based survey is more representative and random as compared to the antenatal sentinel survey. These two national prevalence studies produce different estimates of HIV prevalence. In 2009, the estimated HIV prevalence of pregnant women aged 15-49 was 29.4% while 16.9% was found in the national household-based survey conducted in 2008 for the same age group. The prevalence found among pregnant women is significantly higher than those found among all adults. Some advantages of national household-based survey over antenatal survey were outline in Shisana, et al., (2002). Among other important advantages outlined were: *The household-based survey allowed for analysis of HIV against a range of demographic variables that are not gathered in the antenatal survey – e.g., race, residence genotype and marital status; HIV prevalence could be interpreted in relation to knowledge, attitudes, sexual behaviors and general dynamics of HIV infection, particularly differential infection rates between males and females*". The comparison and limitation of both antenatal and household-based surveys for HIV are available online <http://www.avert.org/safricstats.htm> for further details.

Kleinschmidt et al., (2007) advices that HIV prevalence in a small area is often only conjecture, based on hearsay or on non-representative samples of the population. We believe that prediction and

mapping of HIV prevalence has not been done in South Africa at a very local level where there is no data available. Therefore, this paper deals with the determination of HIV predictive model using HIV risk factors in SA. The model is then used to estimate and map HIV prevalence of young adult age 15-49 years, in non-representative local municipalities (small geographical area) using South African national household-based survey.

Montana et al., (2007) develop a simple model to predict HIV prevalence in Kenya in relation to its spatial and behavioural determinants, using data from the 2003 Kenya Demographic and Health survey. An extensive work on mapping of HIV prevalence in Sub-Saharan Africa has been done previously by, amongst others, Kleinschmidt et al., (2007) and Manda et al., (2010) in South Africa; Kandala et al., (2011) in Zambia. Kleinschmidt et al., (2007) derive spatial statistical models of HIV prevalence in 15-24 year-old men and women separately and, use these models to perform spatial prediction of HIV prevalence for the entire South Africa. They urge that detailed mapping of HIV prevalence in SA provides a good picture of the geographical variation of the epidemic that overall national (or provincial) HIV prevalence estimates are unable to do. Such maps can be used to further guide local prevention and care efforts to ensure their relevance to the local HIV situation. Kandala et al., (2011) used a Bayesian geo-additive mixed model based on Markov Chain Monte Carlo techniques to map the change in the spatial distribution of HIV prevalence at the provincial level in Zambia, accounting for important risk factors.

The predicted HIV prevalence estimates and its standard errors for people age 15-49 years at the local municipalities in SA are computed. The standard errors for the local municipality estimates were used to measure the accuracy of predictions. In cross validation, the predicted HIV prevalence estimates at the local municipalities were very close to the observed HIV prevalence. This indicates that the refined predictive model estimates is able to provide a good fit at the overall local levels. Smooth map of HIV prevalence at municipality levels in South Africa are provided. Multivariate analysis of the effect of various factors of HIV prevalence is given.

In general, this paper focuses on utilising national household-based HIV survey data in developing reliable small area statistics (i.e., HIV prevalence in SA), possibly in conjunction with the census and community survey data. But, the problem with using these national surveys data is the sparse data at small areas. Data source and approach of this study are provided in section 2.

Data Source

The Human Science Research Council (HSRC) has been conducting national household study of HIV/AIDS since 2002. The survey included gathering of data on HIV prevalence, behavior and communication. Based on the surveys, the HIV prevalence of all South Africans over the age of 2 years was 10.9% in 2008. In 2002 and 2005, this figure was 11.4 and 10.8%, respectively. The estimated prevalence among those 15-49 years old was 16.90% and 16.92% in 2008 and 2005 respectively, showing the degree of stabilization. The National HIV household survey of 2005 was used in this study instead of 2008 due to administrative constraints. It is a second series that allow for tracking of HIV and associated determinants over time. It includes all people over 2 years living in SA and residing in homes. The survey design applied is the multi-stage disproportionate, stratified sampling approach. Census 2001 was used as sampling frame, based on a master sample consisting of 1 000 enumeration areas (EAs). The sample was explicitly stratified by province and locality type of the EAs. The primary sampling unit (PSU) was the EA, the secondary sampling unit (SSU) was the visiting point or household, and the ultimate sampling unit (USU) was the individual eligible to be selected for the survey. Three people in each household could potentially be selected, with only one from each of the following groups: 2-14 years, 15-24 years, and 25 years and older. Selected respondents were asked to answer questions relating to HIV risk behavior, sexual behaviors, and socioeconomics and to provide a dry blood spot specimen of oral fluid to test for antibodies to HIV. All HIV testing was anonymous, but linked to questionnaires by unique bar-coded number. More men refuse to do

HIV test than women The overall response rate was above 70%. The reporting of results was on the level of province, age-race group, type of locality. Further details of the survey results and methodology can be found in the main report by Shisana, et al., (2005).

A sample of young adult age 15-49 years was extracted from the national survey and found 12600 respondents in this age group from 930 EAs. We then allocated the 930 sampled EAs to the local municipalities and aggregated HIV status to municipalities. We discovered that only 209 out of 257 local municipalities had data. The remaining 48 municipalities have no HIV data. That is, no respondent were sampled from those municipalities (zero samples). Table 1 below provides the distribution of samples across the local municipalities. The maps in Figure 1 & 2 show the observed distribution of un-adjusted HIV prevalence in district and local municipalities respectively.

Table 1: Distribution of observed samples and HIV prevalence in age 15-49 individuals across municipalities

	Samples (percentage)	HIV prevalence
Min	1 (0.0008%)	0%
Mean	44 (0.04%)	14.16%
Median	22(0.02%)	12%
Max	858 (0.5%)	60%

Clearly, with a sample of 12600 young adults at national level many local municipalities of SA will have very small or zero samples. Therefore, local municipality samples cannot support direct estimates of adequate precision. Often, indirect estimates are used to “borrow strength” by using response variable from related areas and/or time periods to increase effect sample size. The small area estimation method and the adopted method in this work are discussed below.

Small area estimation method

Traditionally, small area estimation relies on a mixed model relating the response of interest (i.e., HIV prevalence) in the small areas to each other and to covariates (HIV risk factors). The model allows the estimates to “borrow strength” from other small areas through random effects terms. A comprehensive review of small area estimation can be found in Gosh and Rao (1994), Marker (1999, 2001), Lohr and Prasad (2003) amongst others. Suppose the area-specific auxiliary information

$$x_i = (x_{i1}, \dots, x_{ip})^T \quad i = 1, \dots, m$$

is available and the parameter of interest θ_i , is assumed to be related to x_i . Then, the model for random area specific effect is defined as

$$\theta_i = x_i^T \beta + v_i z_i, \quad i = 1, \dots, m \tag{1}$$

where β is a known vector of regression parameters/effect of covariates, z_i is a known positive constant and the independent and identical distributed random effects v_i are assumed to be normal with zero mean and variance σ_v^2 . m is the number of small areas. To make inferences about the variable of interest θ_i under model [1], first assume that direct estimators, $\hat{\theta}_i$ are available and that

$$\hat{\theta}_i = \theta_i + e_i \quad i = 1, \dots, m \tag{2}$$

where e_i are sampling errors with zero mean and variance σ_e^2 . Note that $\hat{\theta}_i$ are design unbiased estimate. Combining equation [1] and [2], a special case of the general mixed linear model is produced:

$$\hat{\theta}_i = x_i^T \beta + v_i z_i + e_i \quad i = 1, \dots, m \quad [3]$$

Equation [3] above involves design-induced and model-based random variables, e_i and v_i , respectively. Fay and Herriot (1979) use model [3] with $z_i = 1$; where $\hat{\theta}_i$ is a direct estimator of $\theta_i = \log P_i$ and P_i is the average percapita income in the small area i . Further, $x_i^T \beta = \alpha_0 + \beta x_i$ with x_i denoting the associated county value of $\log P_i$ from the census data. A number of methods are available in the relevant literature and it is clear that there is not a unique best solution to the small area estimation problem. Normally, the choice of a particular method for small area estimation depends on the data needs and on the richness and availability of data sources, which differ from country to country and within countries from one subject matter to another (Drew et. al., 0000). We opted for a simple binomial model also suggested by Montana et al., (2007) to model and map HIV prevalence.

Statistical model and prediction

Suppose P_i is the observed local municipality HIV prevalence in area ($i = 1, \dots, 209$). Furthermore, X_i be a predictor vector of municipality level covariates. The simple binomial probability model is given by

$$Y_i = \text{bin}(N_i, P_i)$$

where Y_i is the observed number of HIV positive in municipality i , N_i is the size of sample in municipality i . We opted for a logistic regression model to predict P_i . The logit model is a generalized linear model used for binomial regression. Similar to other forms of regression analysis, this model makes use of several predictor covariate variables that may be either numerical or categorical. The predictive model used here is given by

$$\text{logit}(P_i) = \log\left(\frac{P_i}{1 - P_i}\right) = \alpha_0 + \beta X_i$$

where α_0 is the intercept and β is the coefficient (effect) of the covariate variables. The coefficient describes the size of the contribution of that risk factor. A positive regression coefficient indicates that the covariate increases the probability of the outcome, while a negative coefficient indicates that the covariate decreases the probability of that outcome; a large coefficient indicates that the risk factor strongly influences the probability of that outcome; while a near-zero coefficient indicates that that risk factor has little influence on the probability of that outcome. From the literature search, a number of spatially relevant covariate variables associated with HIV/AIDS were identified. The identified spatial covariate/risk factors were first assessed in univariate logistic regression models to determine their potential for being significant with HIV/AIDS.

Variables that were associated with HIV/AIDS were used in a predictive model to estimate the prevalence of HIV in the 48 un-sampled local municipalities. The following covariates were deemed significant associated with HIV/AIDS:

Proportion of black African; Proportion of an urban area; Proportion of young females 15-34 years; Proportion of person living under R800 income/month and Population density

They were extracted from census 2001 data and community survey of 2006 from Statistics South Africa. Summary of predictors used in the model is given in Table 2.

Table 2: Summary of predictors that were deemed significant associated with HIV/AIDS

Factors/Predictors	5th Percentiles	Mean	Media	95th Percentiles
Proportion of black Africans	0.0329	0.7204	0.8699	0.9988
Proportion of Urban area	0.0	0.4501	0.5107	0.9437
Sex ratio of young adults age 15-34 years old	0.800	0.9636	0.9300	1.2800
Proportion of young females age 15-34 years	0.1452	0.1731	0.1730	0.1971
Proportion of person living under R800 income/month	0.1448	0.3604	0.3560	0.6247
Population density (person/km ²)	0.52	90.73	32.99	267.31

We then categorize (from lowest and largest proportion) municipalities into five groups and use the municipality with the lowest proportion as the reference. Table 3 contains multivariate analyses of the effect of various risk factors of HIV prevalence.

Table 3: Multivariate analysis of the effect of various factors of HIV prevalence

Risk factors	Quantiles	Odds ratio and 95% Confidence Interval	P-value
Black African	I (Lowest)	1.00	
	II	3.44 (2.23;5.30)	0.0001*
	III	5.84 (3.81;8.96)	0.0001*
	IV	6.59 (4.13;10.53)	0.0001*
	V(Highest)	5.54 (3.35;9.16)	0.0001*
Urban areas	I (Lowest)	1.00	
	II	1.02 (0.79;1.30)	0.896
	III	1.11 (0.84;1.48)	0.460
	IV	1.26 (0.91;1.76)	0.167
	V(Highest)	1.27 (0.92;1.77)	0.151
Young females 15-34	I (Lowest)	1.00	
	II	1.34 (0.97;1.84)	0.075
	III	1.18 (0.87;1.61)	0.280
	IV	0.97 (0.71;1.32)	0.859
	V(Highest)	1.39 (1.01;1.91)	0.044
Population density	I (Lowest)	1.00	
	II	1.00 (0.70;1.44)	0.984
	III	0.91 (0.63;1.30)	0.606
	IV	1.07 (0.74;1.54)	0.711
	V(Highest)	1.04 (0.74;1.46)	0.820

* p-value close to zero (0);

Statistical accuracy of the model

Cross-validation is one of the techniques used to estimate the performance of a predictive model. One round of cross-validation involves partitioning a sample of data into complementary subsets, performing the analysis on one subset (called the training set), and validating the analysis on the other subset (called the validation set or testing set). To reduce variability, multiple rounds of cross-validation are performed using different partitions, and the validation results are averaged over the rounds.

Since our dataset is large, the dataset is partitioned into two complementary subsets (i.e., two equal halves). Then perform analysis on one subset (training set), i.e., half (105) of the municipalities with data are used as the training set to derive the prediction model. The other subset (validation set) is then used to validate the analysis performed in training set. In this case, the predictive model developed from the training set is used to predict HIV prevalence to the remaining half of the municipality where HIV prevalence data was observed. The final refined predictive model given below was used to predict the other half. The predicted HIV prevalence estimates were very close to the observed HIV prevalence. Figure 3 show the predicted HIV prevalence obtained from the training set plotted against observed HIV prevalence.

$$\log\left(\frac{P_i}{1-P_i}\right) = -4.36 + 1.42 \cdot Black_2 + 1.88 \cdot Black_3 + 2.59 \cdot Black_4 + 2.38 \cdot Black_5 \\ - 0.15 \cdot Urban_2 + 0.16 \cdot Urban_3 + 0.83 \cdot Urban_4 + 0.64 \cdot Urban_5 \\ + 0.21 \cdot Female_2 - 0.19 \cdot Female_3 + 0.06 \cdot Female_4 - 0.09 \cdot Female_5 \\ + 0.33 \cdot Density_2 + 0.24 \cdot Density_3 + 0.19 \cdot Density_4 + 0.56 \cdot Density_5$$

where

Black_{}* is a dummy variable indicating the proportion of black Africans in municipality that falls in the * quantile

Urban_{}* is a dummy variable indicating the proportion of Urban areas in municipality that falls in the * quantile

Female_{}* is a dummy variable indicating the proportion of female age 15 to 35 years in municipality that falls in the * quantile

Density_{}* is a dummy variable indicating the population density in municipality that falls in the * quantile

More than 80% of the data fall between the upper-lower confidence interval calculated from the training set. This means that discriminatory power of predictive model is 80%. This indicates that the predictive model estimates is able to provide a good fit at the local level. This model performance extremely well when we aggregated the HIV status to district level municipality relatively to local municipality. One of the reasons is that many local municipalities contain zero prevalence due to the smallness dataset and that might affects the performance of the model. The refined model was used to predict the HIV prevalence in the 48 municipality.

Results

Aggregated data to local municipalities of young adult's age 15-49 years old was used. Some local

municipalities did have data i.e., no respondents were selected to 48 municipalities due to the design of the survey. Others had a very small data. Waterberg district municipality contains local municipality with sixty (60) percent observed HIV prevalence which is the highest. KwaZulu-Natal has many local municipalities with prevalence above 36%. Northern Cape Province have many (8) local municipalities that were not sampled. Seven local municipalities in KwaZulu-Natal have the HIV prevalence above 36%. Municipalities with a lower prevalence are more in both Northern and Western Cape. The average district level HIV prevalence in Eastern Cape is between 9-12%, while OR Tambo district has the highest 16.4%. In Gauteng, West Rand district municipality has the highest prevalence (28.2%) and 40% in one of the local municipality. Only one municipality had no data in Gauteng province.

The summary statistics (i.e., percentiles, mean and median) of predictors of HIV used in the model is given in Table 2. Table 3 provides the multivariate analysis of the effect of various factors of HIV prevalence. It shows that the prevalence of HIV is high in more black African dominant areas. Urban or rural community has no effect on HIV. Population density has also no effect either. Urban/rural and population density are being confounded with black African. All the maps show a large municipality variation in the prevalence of HIV in South Africa.

The predictive model developed here was used to compute predictions of HIV prevalence for people age 15-49 years at the local municipalities in SA as well as its standard errors. The standard errors for the local municipality estimates were used to measure the accuracy of predictions. Also to see if the predicted HIV prevalence for a given municipality is statistically different at a certain significant level for the observed HIV prevalence. The predicted HIV prevalence estimates at the local level were very close to the observed HIV prevalence. This indicates that the predictive model estimates is able to provide a good fit at the local level.

Discussion and Conclusion

Determination of the HIV predictive model using HIV risk factors at the local level has been developed. The HIV predictive model was used to estimate HIV prevalence of young adult age 15-49 years using South African National HIV Prevalence, HIV incidence, Behaviour & Communication survey data. A number of spatially relevant covariate variables associations with HIV/AIDS were identified and used in the predictive model for HIV prevalence. Cross validation was used to determine the accuracy of the HIV predictive model and the model estimates were able to provide a good fit at the local level. The refined HIV model (model fit diagnostics) was then applied in mapping of HIV prevalence to provide a good picture of the geographical variation of the epidemic that overall national HIV prevalence estimates are incapable to do.

The prevalence of HIV is high in more black African dominant areas. Urban or rural community has no effect on HIV. Population density has no effect either. Urban/rural and population density are being confounded with black African. There was a large local municipality variation in the prevalence of HIV in South Africa. Understanding the variation of HIV infection within a country is also essential for determining where prevention and treatment programs need to focus resources; and to effectively target projects and monitor progress. The maps provided here are good guide local prevention and care efforts to ensure the relevance to the local HIV situation. Using small area techniques to predict local level HIV prevalence is a useful tool for informing policies to achieve Millennium Development Goal 6.

The predictive model developed in this paper assumes data are not related in space and we knew that areas in close proximity are more likely to have similar outcomes.

Future Research

The predictive model developed in this paper assumes data are not related in space and we know that areas in close proximity are more likely to have similar outcomes. The prediction must address this: by way of “borrowing/pooling strength” across related areas to find more accurate estimates for a given area or several areas. Related areas can simply mean sharing borders or more advanced by way of similar characteristics. Therefore, the extension of this paper will focus on small area model-based approach to estimate HIV prevalence in local municipalities incorporate auxiliary administrative information as explanatory variables in a regression equation; the predicted value of the regression will be combined with a direct estimate from the national survey to estimate the small area statistics. That is, the structured component introduced will allow us to „borrow strength“ from neighbours in order to cope with the sample variation of the municipality effect and obtain estimates for areas that may have inadequate sample sizes or not sampled at all. This will give more reliable estimates of the fixed effect standard error. This method was used by Lohr and Prasad (2003), to estimate the poverty rate statistics.

Acknowledgements

We thank Statistics South Africa especially Mzi Shabangu for providing census data and mapping of our results.

REFERENCES (RÉFÉRENCES)

1. Drew, J.D., Singh M.P. and Choudhry G.H. (0000). Evaluation of small area estimation techniques for the Canadian Labour Force survey. Statistics Canada
2. Fuller, W. A. (1999). Environmental survey over time. *Journal of agriculture, Biological, and Environmental Statistics*, 4, 331-345.
3. Hickson D. A., Waller L. A., Chao A., Dushimimana A., Bulterys M. (). Small-Area Spatial Analysis and Hierarchical Modeling of HIV-1 Infection: A case study among Pregnant women in a rural but densely populated region of Southern Rwanda. Unpublished manuscript.
4. Gosh, M., and Rao, J.N.K. (1994). Small area estimation: An appraisal. *Statistical Science*, 9, 18-20.
5. Groves, R. M. (1989). *Survey Errors and Survey Costs*. New York: John Wiley & Sons, Inc.
6. Avert. South Africa HIV & AIDS statistics (2010). Available at: <http://www.avert.org/history-aids-south-africa.htm>. Accessed 1 April 2011
7. Wikipedia: Available at: <http://en.wikipedia.org/wiki/Cross-validation>. Accessed 2 Januray 2011
8. UNAIDS (2010). Available at http://www.unaids.org/globalreport/Global_report.htm. Accessed 24 March 2011.
9. Ivaschenko, O and Montana1, L. (0000). Producing Geographic Profiles of HIV Prevalence Using Small Area Estimation Techniques: How Far are we from the “Golden standard” of the Population-based sero-surveys?
10. Kalton, G. (1994). Comments on Singh, Gambino and Mantel. *Survey Methodology*, 20, 18-20.
11. Kalton, G. (2009). Sampling and Oversampling Rare Subpopulations. *Proceeding of the 2009 International Statistical Institute*.
12. Kandala N., Brodish P., Buckner B., Foster S., Madise N., (2011). Millennium development goal 6 and HIV infection in Zambia: what can we learn from successive household surveys? , *Epidemiology and Social*, 25 (1), 95–106.
13. Kleinschmidt L., Pettifor A., Morris N., MacPhail C. and Rees H. (2007). Geographic Distribution of Human Immunodeficiency Virus in South Africa. *Am. J. Trop. Med. Hyg.*, 77(6). 1163-1169
14. Kish, L. (1965). *Survey Sampling*. Wiley, New York.
15. Lohr S. L., and Prasad N. G. N. (2003). Small area estimation with auxiliary survey data. *The Canadian Journal of Statistics*, 31, 383 – 396.
16. Manda S. O. M., Summer R. D. and Mayamise G. (2010). Modelling ARV Regimen Change in HIV/AIDS Patients using Multi-State Model. Unpublished manuscript
17. Marker, D. A. (1999). Organization of small area estimator using a generalized linear framework. *Journal of Official Statistics*. 15, 1-24.
18. Marker, D. A. (2001). Producing Small Area Estimation From national Survey: Methods for Minimising use of Indirect Estimators. *Survey Methodology*, 27, 183-188.
19. Montana L., Neuman M. and Mishra V. (2007). Spatial Modeling of HIV Prevalance in Kenya. *Demographic and Health Research-USAID*. No. 27.
20. Rao, J.N.K. (1999). Some recent advances in model-based small area estimation. *Survey Methodology*, 25, 175-186.
21. Shisana O., Rehle T., Simbayi L. C., Parker W., Zuma K., Bhana A., Connolly C., Jooste S., Pillay V., et al. (2005). *South African National HIV Prevalence, HIV Incidence, Behaviour and Communication Survey, 2005*. Cape Town Press
22. Shisana O., Simbayi L. and Dorkenoo, E. (2002). South Africa's first national population-based HIV/AIDS behavioural risks, sero-status and media impact survey (SABSSM) research project: third quarterly progress report. (Paper prepared for the NMCF/NMF/HSRC Reference Group Meeting held on 23rd July

- 2002).
23. Singh, M. P., Gambino, J., and Mantel, H.J. (1994). Issues and strategies for small area data. *Methodology*, 20, 3-14.
 24. Tamblay J., and Catlin G. (1995). Sample Design of the National Population Health Survey. *Health reports*, 7 (1). Statistics Canada, Cat. No. 82-003
 25. UNAIDS (2010) 'UNAIDS report on the global AIDS epidemic'
 26. Zita A. (2003). Estimating Sub-national Poverty Incidence Using Small Area Estimation Technique. Concluding Workshop of RETA 6007: Enhancing social and Gender Statistics, Bangkok, Thailand
 27. http://en.wikipedia.org/wiki/Logistic_regression

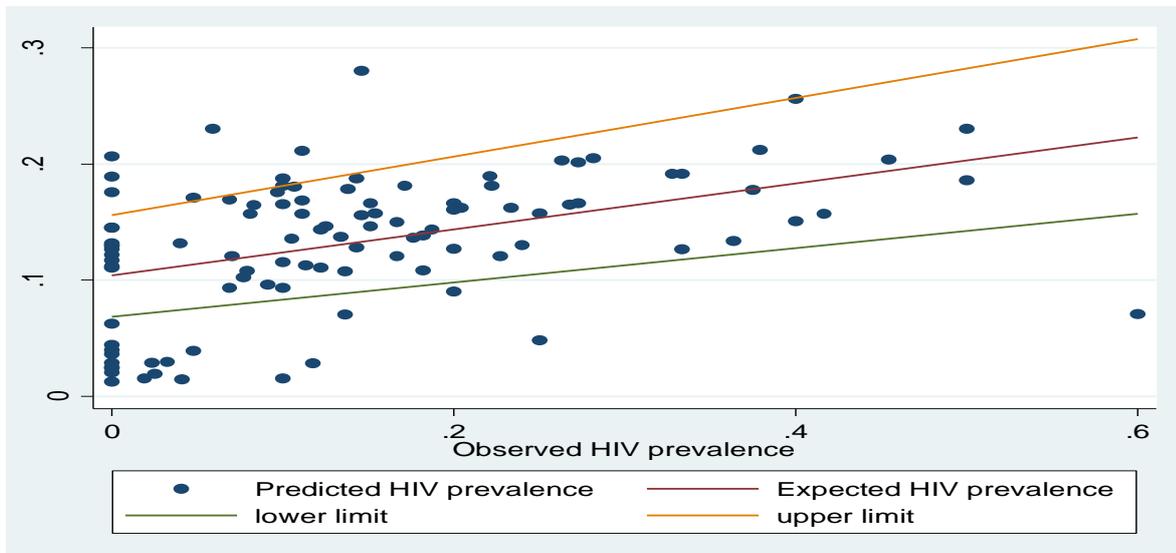


Figure 1: Predicted HIV prevalence plotted against observed HIV prevalence

RÉSUMÉ (ABSTRACT) — optional

National surveys for human immunodeficiency virus (HIV) provide the best available source of information at national level. These surveys do not, however, contain sufficient sample sizes to give reliable estimates by themselves of small areas. It is imperative to report the results of these surveys in small areas for more effective monitoring of local HIV prevalence. But, the reporting of HIV prevalence in South Africa has been reportedly been either at a national or at the level of province, but not at the local level/municipality.

This study deals with the determination of the HIV predictive model at the local level. The HIV predictive model is utilised to estimate HIV prevalence of young adult age 15-49 years, in the 257 municipalities using South African National HIV Prevalence, HIV incidence, Behaviour & Communication survey data. A number of spatially relevant covariate variables associations with HIV/AIDS were identified and used in the predictive model for HIV prevalence. Cross validation was used to determine the accuracy of the HIV predictive model. The refined HIV model (model fit diagnostics) was then applied in mapping of HIV prevalence to provide a good picture of the geographical variation of the epidemic that overall national HIV prevalence estimates are incapable to do.

The prevalence of HIV is high in more black African dominant areas. Urban or rural community has no effect on HIV. Population density has no effect either. Urban/rural and population density are being confounded with black African. There was a large local municipality variation in the prevalence of HIV in South Africa.

Understanding the variation of HIV infection within a country is also essential for determining where prevention and treatment programs need to focus resources; and to effectively target projects and monitor progress. Using small area techniques to predict local level HIV prevalence could be a useful tool for informing policies to achieve Millennium Development Goal 6.