Estimating District HIV Prevalence in Zambia Using Small Area Estimation Methods (SAE)

Chris Mweemba (chris.muna@gmail.com)
University of Zambia https://orcid.org/0000-0002-6834-9896
Peter Hangoma
University of Zambia
Isaac Fwemba
University of Zambia
Wilbroad Mutale
University of Zambia
Felix Masiye
University of Zambia

Research

Keywords: SAE, Small Area Estimation, HIV, Prevalence, District, Fay-Herriot, Auxiliary information

Posted Date: May 20th, 2021

DOI: https://doi.org/10.21203/rs.3.rs-526136/v1

Version of Record: A version of this preprint was published at Population Health Metrics on February 19th, 2022. See the published version at https://doi.org/10.1186/s12963-022-00286-3.

Estimating District HIV Prevalence in Zambia using Small Area Estimation Methods (SAE)

Chris Mweemba[1], Peter Hangoma[1], Isaac Fwemba[1,2], Wilbroad Mutale[1], Felix Masiye[3]

1. Department of Health Policy, Systems and Management, School of Public Health, P.O. Box 50110, Ridgeway Campus, University of Zambia, Lusaka, Zambia
2. School of Public Health, University of Ghana, P.O. Box LG 571,ACCRA
3. Department of Economics, School of Humanities and Social Science, P.O Box 32379, Great East Road Campus, University of Zambia

Corresponding Author:

Name: Chris Mweemba

Email: chris.muna@gmail.com

ABSTRACT

Background

The HIV/AIDS pandemic has had a very devastating impact at a global level, with the Eastern and Southern African region being the hardest hit. The considerable geographical variation in the pandemic means varying impact of the disease in different settings, requiring differentiated interventions. While information on the prevalence of HIV at regional and national levels is readily available, the burden of the disease at smaller area levels, where health services are organized and delivered, is not well documented. This affects the targeting of HIV resources. There is need for studies to estimate HIV prevalence at appropriate levels to improve HIV related planning and resource allocation.

Methods

We estimated the district level prevalence of HIV using Small-Area Estimation (SAE) technique by utilizing the 2016 Zambia Population-Based HIV Impact Assessment Survey (ZAMPHIA) data and auxiliary data from the 2010 Zambian Census of Population and Housing and the HIV sentinel surveillance data from selected antenatal care clinics (ANC). SAE Models were fitted in R Programming to ascertain the best HIV predicting model. We then used the Fay-Herriot (FH) model to obtain weighted, more precise and reliable HIV prevalence for all the districts.

Results

The results revealed variations in the district HIV prevalence in Zambia, with the prevalence ranging from as low as 4.2% to as high as 23.5%. Approximately 35% of the districts (n=26) had

HIV prevalence above the national average, with one district having almost twice as much prevalence as the national level. Some rural districts have very high HIV prevalence rates.

Conclusions

HIV prevalence in Zambian districts is driven by population mobility Districts located near international borders, along the main transit routes and adjacent to other districts with very high prevalence, tend to have high HIV prevalence. The variations in the burden of HIV across districts in Zambia points to the need for a differentiated approaches in HIV programming in Zambia. HIV resources need to be prioritized towards districts with high population mobility.

Keyword: SAE, Small Area Estimation, HIV, Prevalence, District, Fay-Herriot, Auxiliary information

BACKGROUND

The HIV/AIDS pandemic has continued to be a global public health problem, with an estimated 38 million people globally living with HIV in 2019 and the African region bearing the largest burden of the global HIV/AIDS cases (1). Interestingly, the burden of HIV varies considerably within Africa, with sub-Saharan Africa alone accounting for about 70% of all global HIV cases in (SSA) (2). However, a closer review of HIV in the SSA region reveals that the burden is mainly in Eastern and Southern African region (ESA) where, with only 6.2% of the world population, the ESA region accounted for approximately 54% of the total global HIV infections and 43% of all AIDS related deaths in 2019 (1). There is substantial variation in the distribution of HIV within the ESA region. For instance, of the 24 countries in this region, more than a quarter of the new HIV infection in 2018 were in South Africa while 50% of infections were in 7 other countries, namely, and in order of magnitude, Mozambique, Tanzania, Uganda, Zambia, Kenya, Malawi, and Zimbabwe (3).

Similarly, the distribution of HIV within countries has been shown to vary remarkably. In Zambia, for instance, some provinces such as Lusaka (16.1%), Western (16%) and Copperbelt (14.2%) have relatively high prevalence compared to provinces like North-western (6.9%) and Muchinga (5.9%) (ZAMPHIA, 2016). This trend is similar for South Africa where the burden of HIV among adult South Africans in 2016 ranged from as low as 12.6% in Western Cape to as high as 27% in Kwazulu-Natal (KZN) (4).

The information on the geographical variation in HIV prevalence at provincial level is certainly important for guiding government policy, prioritization of interventions, and resource allocation both across and within countries. It should, however, be noted that the burden of diseases within

the provinces can be heterogeneous. For example, within KZN province in South Africa, the district specific HIV prevalence in 2016 ranged from 16.1% in ILembe to 20.6% in uMgungundlovu (5). Similarly, a study that modelled district level estimates for HIV prevalence in South Africa found variations in the prevalence within the South African provinces (6). This means that, effective control and prevention strategies to combat HIV require knowledge of the burden of the disease at smaller and more similar areas such as districts (7). This is challenging, however, because most data currently in use is not sufficiently powered to provide reliable estimates at the small area levels such, as districts (7).

The Zambian Ministry of Health acknowledges the importance of district-level estimates for more focused approaches in HIV programming and in facilitating the achievement of the Fast Track targets (8). The MoH even advocates for targeting districts with high HIV prevalence to maximize HIV treatment enrolment. However, information on the prevalence of HIV at the district level is very limited and the MoH's Fast Track strategies are unlikely to be realized. Currently, existing districts estimates for HIV in Zambia are from routine health facility data which cannot be generalized to the general population due to the non-random nature of the people that present to test for HIV (9). This problem can only be remedied with the use of robust techniques, such as SAE methods, designed to provide valid estimates of the burden of HIV at district level.

District level HIV statistics are of particular importance for Zambia because, a district is the lowest level of decentralization where health services are organized and delivered (10). Unfortunately, as far as the literature search is concerned, no study has been conducted in Zambia to estimate HIV prevalence at the district level. This study, therefore, uses Small Area Estimation methods (SAE) methods to estimate district HIV estimates. Findings from this study

offers very useful insights for better targeting of HIV resources and facilitating the achievement of the Ministry of Health's strategic HIV goals.

METHODS

The district HIV prevalence was estimated using Small-Area Estimation (SAE) methods by utilizing multiple data sources. The SAE method is a statistical technique for obtaining reliable statistics for small areas that are mostly underrepresented in existing data sources due to small sample sizes. Using both direct and indirect methods, SAE models combine multiple data sources (censuses, surveys, etc) containing other related information—auxiliary data— for these small areas (11).

Put simply, small area estimates for HIV prevalence are a weighted average of the direct prevalence estimate from existing data, which due to sample size, may be too unreliable, and therefore requiring a statistical model that utilizes auxiliary data from outside the survey to improve the estimates (12). More weight is placed on predicted prevalence if the variance of direct prevalence is high, and vice-versa (6).

Data Sources

The outcome variable (i.e. HIV prevalence) was obtained from the ZAMPHIA of 2016 while ANC HIV prevalence was collected from pregnant women attending selected ANC clinics in the 74 districts, in 2017 and 2018. The rest of the variables were obtained from the 2010 Census of Population and Housing, and these included; the 2010 Zambian population; proportion of the population aged 15 to 36 years; dependence ratio; the proportion of the population in formal dwelling; proportion of the population with higher education attainment; proportion of the

population residing in the urban area; the population density and the proportion of females in the population.

The ZAMPHIA is a nationally representative cross-sectional, population-based survey of households across Zambia, aimed at measuring the status of Zambia's national HIV response (13). The 2016 ZAMPHIA used a two-stage stratified cluster sampling. The first stage selected 511 enumeration areas (EAs) using probability proportional to size method, and the second stage selected an average of 27 households per EA using equal probability method. A total of 13,441 households and 28,142 individuals were sampled for the survey; 19,168 were adults aged 15 to 59 years and 8,974 were children aged 0 to 14 years. Those aged 15 to 59 years received home-based counselling and testing for HIV. Additional information on the ZAMPHIA methodology is provided in the ZAMPHIA report (13).

Variable Description

HIV prevalence is the number of HIV positive cases per 100 people tested for HIV in the ZAMPHIA and in the selected ANC clinics dotted across all the district. According to the 20101 Census of Population and Housing (14), Population density is the total number of persons per square kilometer; Proportion of urban area is the area considered to be urban out of the total area of the district; Formal dwelling is defined as a room/ set of rooms in a permanent building that could be structurally separated from a permanent building; Dependence ratio is the ratio of the economically inactive persons to a 100 economically active persons; Higher education is the proportion of the population that have attained tertiary education.

Statistical Models

This study used SAE technique to model and estimate HIV prevalence in Zambia, adapting methods from a similar study in South Africa (6). Note that the response variable entered the modelling framework as a logit transformation of the direct district HIV prevalence from the ZAMPHIA survey. The ANC HIV prevalence rate was also modelled as a logit transformation. The HIV prevalence rate are the direct domain estimates of the Zambian district-level HIV prevalence proportions from the ZAMPHIA survey, while the ANC HIV prevalence rate are the prevalence proportions among pregnant women who obtained antenatal care services from clinics dotted across the various districts in Zambia. The logit transformation was necessary for converting prevalence proportions to the real line which helps in ascertaining the normality assumption test. Similarly, sampling error variance was estimated as Delta-method approximation using the variances of the domain estimates as reported and elaborated elsewhere [5]. The model estimated the true HIV prevalence by combining the direct estimate (i.e. direct methods estimation) from the ZAMPHIA survey and the indirect model based estimates, based on auxiliary predictors and district-specific random effects meant to improve the model prediction by borrowing strength from across the districts (6). The direct estimate of HIV prevalence, $\bar{y}_i$ for district i, was obtained as a weighted mean district specific HIV prevalence from the ZAMPHIA survey. This estimate can be viewed to be as follows;

$$\bar{y}_i = \Theta_i + \epsilon_i \qquad (1)$$

Where $\bar{y}_i$is the HIV prevalence estimate for district i estimated from the survey data; Θ_i is the district's true HIV prevalence being estimated and ϵ_i is the random error with mean 0 and variance σ_i^2 and is assumed to be normally distributed.

However, since the number of respondents sampled at district level, during the ZAMPHIA, is not sufficient to provide reliable district HIV prevalence estimates, the second part of the model, referred to as indirect method, was estimated to improve the reliability of the estimates. Therefore, in addition to the direct prevalence estimates obtained from ZAMPHIA, the indirect method used auxiliary information from within the district and neighboring districts, and other data sources to borrow strength and improve the precision of the HIV prevalence estimates (15). Since the outcome variable was a logit transformation of HIV prevalence, we assumed that, HIV prevalence is a linear function of covariates or HIV risk factors obtained from auxiliary data (6). The true HIV prevalence (Θ_i in equation 1) can therefore be thought of as;

$$\Theta_i = x_i\beta + v_i \qquad (2)$$

β is a set of regression coefficients obtained by regressing $\bar{y}_i$ on HIV risk factors (x_s) and v_i are normally distributed random errors with mean 0 and variance σ_v^2. Note that σ_v^2 and σ_i^2 are independent of each other. Combining equations 1 and 2, above, gives the following mixed-effects linear regression model;

$$\bar{y}_i = x_i\beta + v_i + \epsilon_i \qquad (3)$$

To improve precision of the HIV prevalence estimates from equation 3, there is need for a model that combines direct and indirect estimates into a single estimate, such as the Fay-Herriot (FH) small area estimator. The FH estimator is a linear combination of a direct and synthetic estimator which reduces estimation variance in the underrepresented small areas and in the whole model (16). The FH estimator is given by;

$$\hat{y}_i = \gamma_i\bar{y}_i + (1-\gamma_i)x_i\hat{\beta} \qquad (4)$$

Where γ_i and $1-\gamma_i$ are weights for the direct estimate $\bar{y}_i$ and the synthetic estimate,$x_i\hat{\beta}$, respectively, which constitute the FH estimator. Note that γ is simply the ratio of the model error variance to the total error i.e. $\frac{\sigma_v^2}{\sigma_v^2+\sigma_i^2}$. This means that, if the survey based estimates are precise, more weight is given to the direct estimate. Similarly, low precision of the survey based estimates results in more weight being given to the synthetic or indirect estimates.

Spatial Correlation

There is evidence that areas close to each other tend to have similar population dynamics, such as disease risk factors and disease burden (17). This highlights the importance of location and geographical clustering in determining the spread of, and burden of disease – especially infectious diseases, for areas that are in close proximity (18,19). A study in Ethiopia documented the importance of geographical clustering in determining the prevalence of HIV and Tuberculosis (TB) (20).

To account for this spatial correlation, we built a spatial Fay-Herriot (SFH) model and tested it against a non-spatial model to ascertain the best fitting model for this study. A Spatial Adjacency Matrix (W) was built in Excel, as follows;

Spatial Adjacency Matrix (W) is an n x n matrix where n is the number of district in Zambia.

The diagonal entries are $W_{ii}=0$*, indicating no correlation for district i to itself*

The off-diagonal row entries add up to 1, i.e. $W_{ij}=1$*. This can be thought of as follows, as presented by Yakoi and Ando* (21)*;*

$$w_{1ij} = \begin{cases} 1/d_{ij,}^{\alpha} \\ 0 \end{cases} \quad (5)$$

If $I \neq j$ otherwise

$$w_{oij} = w_{ij} / \sum_{k=1}^{N} w_{1ik} \quad (6)$$

Where d_{ij}, in Eq. (5) is the distance between districts i and j; α is a parameter of the distance decay (α =0 if ij not sharing border, otherwise 0 < α <1). According to Eq. (6), the total amount of influence that one area receives from other areas is fixed (21).

The data was conducted in R (22) utilizing the SAE package built in the software (23). Figures were produced with the ggplot2 package (24).

Model Selection

We fitted 19 variations of the basic area-level model, which differed in the inclusion of auxiliary predictors and assumptions about the random effects. The fitted models assumed independently normally distributed random effects and, where possible, Simultaneously Autoregressive (SAR) random effects, based upon spatial adjacency of the districts of Zambia, were also fitted. There are other models that can be used to account for autocorrelation effect, such as the Conditional Autoregressive (CAR) model, and its intrinsic version (Intrinsic Autoregressive [IAR] model, and the decision to use SAR is because these models are equivalent and in practice produce similar results (25,26). The first model had a single covariate, and was then augmented by adding an additional covariate to each subsequent model until all the covariates were added. The first nine models assumed an independent covariance structure while models ten to eighteen are similar to

models 1 -9 except for the inclusion of the simultaneously autoregressive covariance structure. A Spatial Adjacency Matrix, described earlier, accounted for the SAR covariance structure.

Model nineteen only contains the SAR covariance structure without any covariate. The best fitting model (Model 10) included only the population aged 15 – 34 years with the SAR covariance structure, and it was selected based on the Akaike Information Criterion (AIC). This model was thereafter used, in combination with the survey based HIV prevalence estimates, to model the prevalence of HIV in all the 74 districts of Zambia. The first ten model fitted are presented below;

***Model 1**; Logit (HIV Prevalence) = Proportion of population aged 15 and 35 years.*

***Model 2**; Logit (HIV Prevalence) = Proportion of population aged 15 and 35 years + Dependence Ratio.*

***Model 3**; Logit (HIV Prevalence) = Proportion of population aged 15 and 35 years + Dependence Ratio + Formal Dwelling.*

***Model 4**; Logit (HIV Prevalence) = Proportion of population aged 15 and 35 years + Dependence Ratio + Formal Dwelling + Proportion with higher education.*

***Model 5**; Logit (HIV Prevalence) = Proportion of population aged 15 and 35 years + Dependence Ratio + Formal Dwelling + Proportion with higher education + Proportion of urban area.*

***Model 6**; Logit (HIV Prevalence) = Proportion of population aged 15 and 35 years + Dependence Ratio + Formal Dwelling + Proportion with higher education + Proportion of urban area + Population in 2010.*

***Model 7**; Logit (HIV Prevalence) = Proportion of population aged 15 and 35 years + Dependence Ratio + Formal Dwelling + Proportion with higher education + Proportion of urban area + Population in 2010 + ANC HIV prevalence.*

***Model 8**; Logit (HIV Prevalence) = Proportion of population aged 15 and 35 years + Dependence Ratio + Formal Dwelling + Proportion with higher education + Proportion of urban area + Population in 2010 + ANC HIV prevalence + Population density in 2010.*

***Model 9**; Logit (HIV Prevalence) = Proportion of population aged 15 and 35 years + Dependence Ratio + Formal Dwelling + Proportion with higher education + Proportion of urban area + Population in 2010 + ANC HIV prevalence + Population density in 2010 + Proportion of Females.*

***Model 10;** Logit (HIV Prevalence) = Proportion of population aged 15 and 35 years + SAR Covariance Structure.*

RESULTS

Table 1 below shows the distribution of HIV against the auxiliary variables used to predict the district HIV prevalence. For instance, it can be seen that there are significantly more individuals aged 15 to 35 in the HIV positive category compared to the HIV negative category. Similarly, the HIV positive category comparatively more people that had achieved higher education and had more pregnant women testing positive for HIV, than in the HIV negative category. Table 1 provides more details.

Table 1: The distribution of HIV status across the selected auxiliary variables

Auxiliary Variables	Frequency*		P-value
	HIV Positive	HIV Negative	
Proportion of population aged 15 -35 years	0.379	0.368	<0.001
Proportion of dependence ratio	0.886	0.934	<0.001
Proportion living in formal dwelling	0.373	0.286	<0.001
Proportion of population with Higher Education	0.76	0.64	<0.001
Proportion of area in urban	51.1	41.2	<0.001
Proportion of population (***n=19,115***)	12.9	87.1	<0.001
Proportion of HIV among pregnant women at ANC clinics	72.7	60.2	<0.001
Population density	1017.8	679.9	<0.001
Average female population	50.8	50.8	0.72

*Note that the mean value is presented for Population density

Model diagnosis and validation

The results obtained using the SAE estimates model were consistently more precise than those obtained from the direct estimate methodology. For instance, the Relative Mean Standard Errors (RMSE) in Figure 1 and the Relative Standard Errors (See Additional file 1) for the SAE are continuously lower than those from the direct estimate model. In addition, the reduction in relative standard errors, due to SAE, was greatest in districts which produced the least precise direct estimates. For instance, districts like Chadiza, Milenge, Gwembe and Chavum, have relative standard errors reducing from 99.7% to 31%, 70.2% to 29.9%, 70.9% to 29.4% and 70.4% to 33.7%, respectively. Assuming, for example, that "useful" estimates are those for which RSE $\leq$ 20%, then our SAE model produced useful estimates in 53 of the 74 districts for which direct estimation failed to produce useful estimates.

Figure 1: Relative Mean Standard Errors (RMSE) for the FH HIV prevalence estimates and survey-based prevalence estimates: The RMSE show lower mean standard errors for the Fay-Herriot small area estimations over the survey-based estimation for all the 74 Zambian districts

It is worth noting that the estimates from the Fay-Herriot estimator had narrower 95% confidence intervals than the direct estimates (See Figure 2 below). Conversely, some point estimates for some districts such as Chadiza and Gwembe differed rather substantially between the design-based and model based estimates. The design-based survey domain estimate of HIV prevalence in Gwembe and Chadiza were of little value for lack of precision, and at most misleading. Smaller relative standard errors from the FH small area estimates are more likely to be true than those from the direct estimates, and are much more likely to be similar to surrounding districts.

The conclusion from this model diagnostics and validation is that, the FH estimator produces smaller standard errors compared to the survey based estimates, across all the 74 districts of Zambia. This means that SAE prevalence estimates are more reliable than those obtained from the direct estimates.

Figure 2: HIV prevalence estimates and confidence intervals for the FH and direct estimates in Zambia's districts

District HIV Prevalence Estimates

The district HIV prevalence in Zambia ranges from as low as 4.2% in Lundazi to as high as 23.5% Namwala. Other notable districts with high HIV prevalence, in order of magnitude, include Mongu (22.9), Mazabuka (18.5%), Kalulushi (17.4%), Choma (17.2), Kafue 17.1%), Itezhi-tezhi (16.8%) and Lusaka (16.6%). On the other hand, the five districts with the lowest

HIV prevalence, in descending order, were; Chama (5.1%), Zambezi (5%), Mafinga (4.7%), Kabompo (4.6) and Lundazi (4.2%). Figure 3 below shows the prevalence estimates obtained from the FH estimator. Note that the districts from the same province have the same color code, making it possible to have a visual sense of the distribution of HIV across the districts within the provinces.

Figure 3: Estimated district HIV prevalence in Zambia

The results in figure 3 shows that the districts in Central (Green), Copperbelt (Blue), Lusaka (Grey), Southern (Brown) and Western (Red) provinces had relatively high HIV prevalence compared to districts in Eastern (Orange), Luapula (Yellow), Muchinga (Pink), Northern (Black) and North Western (Purple) provinces.

Mapping district HIV prevalence

The distribution of district HIV prevalence is further illustrated with a map in figure 4. The results reveal that, 12 of the 74 districts had relatively low HIV prevalence (≤ 7.7%), 38 districts had relatively moderate HIV prevalence (between 7.7% and 12.1%), 21 districts had relatively high HIV prevalence (between 12.1% and 18.1%) while 3 districts had relatively very high HIV prevalence (between 18.1% and 23.5%). The spatial effect of HIV prevalence can also be seen from the map, with relatively high HIV being concentrated in areas around central, southern and western Zambia.

Figure 4: Zambia district HIV prevalence map

The mapping also shows that, generally, the districts in the north and eastern parts of the country have moderate HIV prevalence while districts in north-western and north eastern parts of the country, i.e. North-Western and Muchinga provinces, have the lowest HIV prevalence.

DISCUSSION

This paper is the first to use SAE methods to estimate the prevalence of HIV at district level in Zambia. Our study has demonstrated that, national HIV estimates currently being used for HIV programming fail to account for the full picture of the distribution, and extent of the variations in HIV prevalence at lower levels (6,7,27,28). Amoako Johnson (27), for instance, warns that relying on national estimates for planning could lead to an 'ecological fallacy', where planning and resource allocation fails to properly account for the variations that exist at small domains, but may not be apparent at national level. The one-size-fits-all approach, associated with national level estimates, will therefore, not achieve the desired results at local level (28).

In the midst of declining HIV funding (29), designing and targeting of HIV interventions require adequate knowledge on where the biggest resource need lies. In the context of Zambia, for example, national HIV estimates would demand that more resources be allocated to the Western province, based on the disease burden. However, these national level estimates do not provide any information on the district specific HIV burden, or sub-groups in greater need of HIV policy targeting within the province (30). The revelations of the wide variations in the burden of HIV within districts should be a policy concern, and effectively results in the redundancy of the "bigger picture" approaches from national estimates, especially if the intention is to make HIV programs at local levels more pragmatic and optimal (31,32). The importance of accounting for within district variation in HIV prevalence can be highlighted from our study. For instance, while

the average HIV prevalence for Southern province is around 13%, the within province prevalence varies from as low as 7.4% to as high as 23.5%. Ensuring effective service delivery, under such circumstances, requires recognizing and tailoring interventions to the needs of the different subpopulations at the level at which service delivery is organized and delivered (33). This remains a challenge for low resource countries, however, due to the higher cost of obtaining data to generate small area estimates (34).

Our study has also revealed important information on the predictors and drivers of HIV prevalence at district level. For instance, the age group 15-35 years, being the best out of survey predictor for district HIV prevalence, is an indication that young people account for the distribution of HIV at district level. This is a similar trends at national and global level (35–37). Similarly, a South African study found young females, aged 15-34 years, to be important determinants of district HIV prevalence (7).

Another important finding in this study is that district HIV prevalence is spatially correlated, i.e., HIV prevalence in one district is correlated with the prevalence in adjacent districts. This is reasonable and expected since district boundaries are arbitrary, and expected that individuals living in districts close to each other are likely to have similar characteristics and risk factors (27,38). Similar studies have acknowledged the importance of accounting for spatial correlation at small area levels (6,27). This may especially be true for communicable diseases such as HIV. It would be prudent, therefore, for neighboring districts to employ coordinated approaches to HIV programming and have a shared understanding of local HIV drivers and impact of the disease. The mapping of HIV prevalence in our study provides useful information to facilitate such a coordinated HIV response.

The national HIV prevalence for Zambia has generally been highest in urban areas (13,35,39) and this is similar to other countries within the region such as Malawi, Kenya, South Africa and Zimbabwe (40–43). However, district level estimates from our study have revealed that HIV prevalence in some rural districts is comparable and sometimes even higher than the prevalence in urban districts. For instance, we found that the two highest HIV prevalence in Zambia are in predominantly rural districts, with the highest district having almost seven percentage points higher prevalence than that of the most urbanized district of Lusaka. This is further proof that national level estimates mask very important HIV dynamics that can guide resource allocation at local levels (44). It is likely that the national level HIV dynamics observed in most countries is different to the situation at lower levels. As long as lower level prevalence estimates remain unknown, therefore, allocation of HIV resources will remain suboptimal (45).

The lessons that can be learnt from our study is that HIV at district level is driven by high mobility which comes with commerce and trade. The two rural districts with the highest HIV prevalence in Zambia are fishing districts which attract a large number of people to the districts every year (46–49). Therefore, districts that experience high population mobility due to commerce and trade should be marked for HIV interventions such as test and treat services, regardless of whether the districts are rural or urban. Other similar countries can draw important lessons from this finding. Population mobility has been shown to be a driver of HIV infections in other settings as well (50–52). To demonstrate the importance of population mobility in HIV transmission, our study found that districts that experience seasonality of employment, located along the main transit routes and those along the international border have higher HIV prevalence than the national average. The above factors have been shown to be associated with HIV in other settings as well (53–55). Districts experiencing high population mobility are

potential HIV hotspots and should be prioritized for HIV interventions. Similarly, areas that are in close proximity to districts with known high HIV prevalence need close attention due to the spatial nature of the HIV epidemic, as revealed by our study.

CONCLUSION

This is the first study in Zambia to present and map HIV prevalence estimates at district level using SAE methods. It is clear from the results that national estimates mask the wide variation in HIV prevalence within the districts. Ensuring that HIV resources are allocated where they are needed, requires knowledge on the distribution of HIV at smaller, more homogeneous areas such as districts. This study has been able to provide this information and mapped the distribution of district HIV in Zambia.

The revelation that HIV prevalence is very high in some rural districts is an important finding for HIV programming. It is useful for policy makers to realize that relying on national level prevalence to plan interventions at district level may not optimal because the HIV dynamics at district level may well be different. Utilizing results from SAE techniques for planning and resource allocation would ensure achievement of universal access to resources by underserved and underrepresented populations.

Our results have documented drivers and markers of high HIV prevalence at district level, information that can be used to plan prevention and treatment interventions. Population mobility is a key driver of HIV and should be an important consideration when designing HIV interventions. Profiling the burden of disease at appropriate levels is also a key aspect in designing responsive HIV interventions, and SAE models will increasingly become important tools in guiding policy making and decision making, especially for low resource settings.

LIST OF ABBREVIATIONS

HIV/AIDS: Human Immunodeficiency Virus/ Acquired Immunodeficiency Syndrome

SAE: Small Area Estimation

ZAMPHIA: Zambia Population-Based HIV Impact Assessment Survey

ANC: Antenatal Care

FH: Fay-Herriot

ESA: Eastern and Southern African region

KZN: Kwazulu-Natal

MoH: Ministry of Health

EA: Enumeration Area

TB: Tuberculosis

SAR: Simultaneously Autoregressive

CAR: Conditional Autoregressive

IAR: Intrinsic Autoregressive

AIC: Akaike Information Criterion

RSE: Relative Standard Errors

RMSE: Relative Mean Standard Errors

UNZABREC: University of Zambia Biomedical Research Ethics Committee

ZamStats: Zambia Statistics Agency

US: United States

NIH: National Institutes of Health

STUDY LIMITATIONS

The SAE model used in this study helped produce district HIV prevalence estimate, however, the use of relative mean standard errors and confidence intervals to validate the model has a potential bias. It should be noted that ZAMPHIA is not designed to collect representative data at district level, and by design therefore, SAE methods are always going to produce relatively better estimates, with smaller standard errors than ZAMPHIA estimates because they utilize additional data, in addition to the survey based estimates. An additional validation method would have been more useful. Additionally, the model was built with covariates as collected by the Census data and ZAMPHIA, and there is a chance that other HIV related covariates, not collected by the Census and the ZAMPHIA, e.g. the prevalence of transactional sex, could have strengthened the model. This study has, however, provided policy relevant information that can be utilized to improve targeting of HIV resources at local levels where interventions are planned and delivered.

DECLARATIONS

Ethics approval and consent to participate

Ethical approval was obtained from the University of Zambia Biomedical Research Ethics Committee (UNZABREC) (**REF. NO. 937-2020)** and permission from the Zambia National Health Research Authority. The study utilized secondary data from the Zambia Statistics Agency (ZamStats) and the Zambian Ministry of Health.

Consent for publication

Not applicable.

Availability of data and materials

The datasets used and/or analyzed during the current study are available from the corresponding author on reasonable request.

Competing interests

The authors declare that they have no competing interests.

Funding

This research was supported by the Fogarty International Center of the US National Institutes of Health (NIH) under Award Number D43 TW009744. The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institutes of Health.

Authors' contributions

CM contributed to the designing of the study, data analysis, interpretation and writing of the manuscript. IF contributed to data analysis, writing of the methodology section and model building. PH contributed to building the model and validating it, and also writing up of the manuscript. WM contributed to writing up of the results and discussion section. FM contributed to critically reviewing and finalizing the write up of the manuscript. All authors read and approved the final manuscript.

Acknowledgements

We are grateful to the Zambia Statistical Agency and the Zambian Ministry of Health for providing the data for this study.

REFERENCES

1. UNAIDS. Global HIV & AIDS statistics — 2020 fact sheet [Internet]. 2020 [cited 2021 Apr 15]. Available from: https://www.unaids.org/en/resources/fact-sheet
2. UNAIDS. AIDSinfo [Internet]. 2020 [cited 2021 May 10]. Available from: https://aidsinfo.unaids.org/
3. UNAIDS. AIDSinfo [Internet]. 2019 [cited 2020 Nov 3]. Available from: http://aidsinfo.unaids.org/
4. Simbayi L, Zuma K, Zungu NP, Moyo S. South African National HIV Prevalence, Incidence, Behaviour and Communication Survey, 2017 – The Human Sciences Research Council (HSRC) [Internet]. 2017 [cited 2020 Oct 21]. Available from: https://www.hsrcpress.ac.za/books/south-african-national-hiv-prevalence-incidence-behaviour-and-communication-survey-2017
5. KwaZulu-Natal Office of the Premier HIV/AIDS Directorate. The KwaZulu-Natal Provincial Multi-Sectoral HIV, TB and STIs Implementation Plan 2017-2022 [Internet]. 2017. Available from: https://sanac.org.za/wp-content/uploads/2019/02/PIP_KZN_Final-1.pdf
6. Gutreuter S, Igumbor E, Wabiri N, Desai M, Durand L. Improving estimates of district HIV prevalence and burden in South Africa using small area estimation techniques. PLoS ONE [Internet]. 2019 Feb 22 [cited 2019 Oct 3];14(2). Available from: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC6386240/

7. Kondlo LO, Manda SOM. Small Area Estimation of HIV Prevelance using National Survey data in South Africa. In 2011.

8. Ministry of Health. AIDS Response Fast Track Strategy 2015-2020 [Internet]. 2015 [cited 2021 Mar 30]. Available from: https://www.nac.org.zm/sites/default/files/publications/Zambia%20Fast%20Track%20Strategy_0.pdf

9. Ouma J, Jeffery C, Valadez JJ, Wanyenze RK, Todd J, Levin J. Combining national survey with facility-based HIV testing data to obtain more accurate estimate of HIV prevalence in districts in Uganda. BMC Public Health. 2020 Dec;20(1):1–14.

10. Ministry of Health. National Health Strategic Plan Monitoring and Evaluation Framework 2017-2021 [Internet]. 2019. Available from: https://www.moh.gov.zm/?wpfb_dl=121

11. Hidiroglou M. Small-Area Estimation : Theory and Practice. JSM proceedings, survey research methods section [Internet]. Alexandria, VA: American Statistical Association; 2007 [cited 2018 Sep 4]. Available from: /paper/Small-Area-Estimation-%3A-Theory-and-Practice-Hidiroglou/327266333da6fa5f51f71bf74f19fcfd0c2b24df

12. Chandra H, Chambers R. Multipurpose Weighting for Small Area Estimation. J Off Stat. 2009 Sep 1;25:379–95.

13. Ministry of Health, Zambia. Zambia Population-based HIV Impact Assessment (ZAMPHIA) 2016: First Report. Zambia. 2017.

14. Central Statistics Office. 2010 Census of Population and Housing: Zambia National Analytical Report [Internet]. 2012 [cited 2015 Apr 20]. Available from: http://www.zamstats.gov.zm/report/Census/2010/National/2010%20Census%20of%20Population%20National%20Analytical%20Report.pdf

15. Pfeffermann D, Tiller R. Small-Area Estimation With State–Space Models Subject to Benchmark Constraints. J Am Stat Assoc. 2006 Dec 1;101(476):1387–97.

16. Porter AT, Holan SH, Wikle CK, Cressie N. Spatial Fay–Herriot models for small area estimation with functional covariates. Spat Stat. 2014 Nov 1;10:27–42.

17. Legendre P, Legendre L. Numerical Ecology. Elsevier; 2012. 1007 p.

18. Anselin L, Bera AK. Spatial dependence in linear regression models with an introduction to spatial econometrics. Urbana-Champaign; 1996. (Office of Research working paper / University of Illinois at Urbana-Champaign, College of Commerce and Business Administration).

19. Riley S. Large-Scale Spatial-Transmission Models of Infectious Disease. Science. 2007 Jun 1;316(5829):1298–301.

20. Alene KA, Viney K, Moore HC, Wagaw M, Clements ACA. Spatial patterns of tuberculosis and HIV co-infection in Ethiopia. PLOS ONE. 2019 Dec 5;14(12):e0226127.

21. Yokoi T, Ando A. One-directional adjacency matrices in spatial autoregressive model: A land price example and Monte Carlo results. Econ Model. 2012 Jan 1;29(1):79–85.

22. R Core Team. R: A language and environment for statistical computing. 2013;

23. Molina I, Marhuenda Y. sae: An R Package for Small Area Estimation. R J. 2015;7(1):81–98.

24. Wickham H. ggplot2 [Internet]. New York, NY: Springer New York; 2009 [cited 2021 Apr 20]. Available from: http://link.springer.com/10.1007/978-0-387-98141-3

25. Ver Hoef JM, Hanks EM, Hooten MB. On the relationship between conditional (CAR) and simultaneous (SAR) autoregressive models. Spat Stat. 2018 Jun 1;25:68–85.

26. De Smith MJ. Statistical Analysis Handbook 2018 edition [Internet]. The Winchelsea Press; 2018 [cited 2021 Apr 25]. Available from: https://www.statsref.com/HTML/index.html?car_models.html

27. Amoako Johnson F, Padmadas SS, Chandra H, Matthews Z, Madise NJ. Estimating unmet need for contraception by district within Ghana: an application of small-area estimation techniques. Popul Stud. 2012 Jul;66(2):105–22.

28. Makurumidze R, Decroo T, Lynen L, Chinwadzimba ZK, Van Damme W, Hakim J, et al. District-level strategies to control the HIV epidemic in Zimbabwe: a practical example of precision public health. BMC Res Notes. 2020 Aug 26;13(1):393.

29. ten Brink D, Martin-Hughes R, Kelly SL, Wilson DP. What is the impact of a 20% funding cut in international HIV aid from the United States? AIDS. 2019 Jul 1;33(8):1406–1408.

30. Asian Development Bank. Introduction to Small Area Estimation Techniques: A Practical Guide for National Statistics Offices [Internet]. Asian Development Bank; 2020 [cited 2021 Apr 21]. Available from: https://www.adb.org/publications/small-area-estimation-guide-national-statistics-offices

31. Rao JNK. Some Methods for Small Area Estimation. Riv Internazionale Sci Sociali. 2008;116(4):387–406.

32. Niragire F, Achia TNO, Lyambabaje A, Ntaganira J. Bayesian Mapping of HIV Infection among Women of Reproductive Age in Rwanda. PLOS ONE. 2015 Mar 26;10(3):e0119944.

33. Jain S, Wilk AS, Thorpe KE, Hammond PS. A Model for Delivering Population Health Across the Care Continuum [Internet]. AJMC. 2018 [cited 2021 Apr 21]. Available from: https://www.ajmc.com/view/a-model-for-delivering-population-health-across-the-care-continuum

34. Bernal RTI, de Carvalho QH, Pell JP, Leyland AH, Dundas R, Barreto ML, et al. A methodology for small area prevalence estimation based on survey data. Int J Equity Health. 2020 Jul 31;19(1):124.

35. Nakazwe C, Michelo C, Sandøy IF, Fylkesnes K. Contrasting HIV prevalence trends among young women and men in Zambia in the past 12 years: data from demographic and health surveys 2002–2014. BMC Infect Dis. 2019 May 17;19(1):432.

36. Tyler KA, Handema R, Schmitz RM, Phiri F, Wood C, Olson K. Risk Factors for HIV among Zambian Street Youth. J HIVAIDS Soc Serv. 2016;15(3):254–68.

37. Avert. Young people, HIV and AIDS [Internet]. 2015 [cited 2021 May 6]. Available from: https://www.avert.org/professionals/hiv-social-issues/key-affected-populations/young-people

38. Pratesi M, Salvati N. Small area estimation: the EBLUP estimator based on spatially correlated random area effects. Stat Methods Appl. 2008 Feb 1;17(1):113–41.

39. Fylkesnes K, Musonda RM, Kasumba K, Ndhlovu Z, Mluanda F, Kaetano L, et al. The HIV epidemic in Zambia: socio-demographic prevalence patterns and indications of trends among childbearing women. AIDS Lond Engl. 1997 Mar;11(3):339–45.

40. Nutor JJ, Duah HO, Agbadi P, Duodu PA, Gondwe KW. Spatial analysis of factors associated with HIV infection in Malawi: indicators for effective prevention. BMC Public Health. 2020 Jul 25;20(1):1167.

41. Magadi MA. Understanding the urban–rural disparity in HIV and poverty nexus: the case of Kenya. J Public Health. 2017 Sep 1;39(3):e63–72.

42. Gibbs A, Reddy T, Dunkle K, Jewkes R. HIV-Prevalence in South Africa by settlement type: A repeat population-based cross-sectional analysis of men and women. PLOS ONE. 2020 Mar 17;15(3):e0230105.

43. Schaefer R, Gregson S, Takaruza A, Rhead R, Masoka T, Schur N, et al. Spatial patterns of HIV prevalence and service use in East Zimbabwe: implications for future targeting of interventions. J Int AIDS Soc. 2017;20(1):21409.

44. Kayeyi N, Fylkesnes K, Michelo C, Makasa M, Sandøy I. Decline in HIV Prevalence among Young Women in Zambia: National-Level Estimates of Trends Mask Geographical and Socio-Demographic Differences. PLOS ONE. 2012 Apr 4;7(4):e33652.

45. National Research Council (US) Committee on Health. Improving Health in the United States: The Role of Health Impact Assessment: Why We Need Health-Informed Policies and Decision-Making [Internet]. Washington (DC): National Academies Press (US); 2011 [cited 2021 Apr 24]. Available from: https://www.ncbi.nlm.nih.gov/books/NBK83538/

46. Ndubani P, Kamwanga J, Tembo R, Tete J, Buckner B. PLACE in Zambia: Identifying Gaps in HIV Prevention in Mongu, Western Province, 2005 — MEASURE Evaluation [Internet]. 2006 [cited 2021 Mar 18]. Available from: https://www.measureevaluation.org/resources/publications/tr-06-42

47. Singh K, Buckner B, Tate J, Ndubani P, Kamwanga J. Age, poverty and alcohol use as HIV risk factors for women in Mongu, Zambia. Afr Health Sci. 2011 Jun;11(2):204–10.

48. The OPC Fund for International Development. Improving Access to WASH and HIV Services in Limulunga District in Zambia - OPEC Fund for International Development [Internet]. https://opecfund.org. 2014 [cited 2021 Mar 18]. Available from: https://opecfund.org/operations/list/improving-access-to-wash-and-hiv-services-in-limulunga-district-in-zambia

49. Mweemba CE, Funder M, Nyambe I, Van Koppen B. Poverty and Access to Water in Namwala District, Zambia: Report on the Results from a Household Questionnaire Survey, Zambia [Internet]. 2011 [cited 2021 Mar 18]. Available from: https://www.diis.dk/en/research/poverty-and-access-to-water-in-namwala-district-zambia

50. Solomon S, Kumarasamy N, Ganesh AK, Amalraj RE. Prevalence and risk factors of HIV-1 and HIV-2 infection in urban and rural areas in Tamil Nadu, India. Int J STD AIDS. 1998 Feb;9(2):98–103.

51. Camlin CS, Charlebois ED. Mobility and its Effects on HIV Acquisition and Treatment Engagement: Recent Theoretical and Empirical Advances. Curr HIV/AIDS Rep. 2019 Aug 1;16(4):314–23.

52. Cassels S. Time, population mobility, and HIV transmission. Lancet HIV. 2020 Mar 1;7(3):e151–2.

53. Coulibaly I. The Impact of HIV/AIDS on the Labour Force in Sub-Saharan Africa: A Preliminary Assessment. Int Labour Organ Res Policy Anal No 3. 2005 Jan 1;

54. Avert. HIV and AIDS in Zambia [Internet]. Avert. 2018 [cited 2021 Mar 18]. Available from: https://www.avert.org/professionals/hiv-around-world/sub-saharan-africa/zambia

55. Jawando JO, Adeyemi EO. Sexual Exchange and Cross-Border Trade: Implications for HIV/AIDS in Nigeria. SAGE Open. 2020 Apr 1;10(2):2158244020917949.

FIGURE/TABLE LEGENDS

Figure 1: Relative Mean Standard Errors (RMSE) for the FH HIV prevalence estimates and survey-based prevalence estimates: The RMSE show lower mean standard errors for the Fay-Herriot small area estimations over the survey-based estimation for all the 74 Zambian districts

Figure 2: HIV prevalence estimates and confidence intervals for the FH and direct estimates in Zambia's districts: The confidence intervals of the FH estimates are narrower than those of the direct estimates for most of the districts

Figure 3: Estimated district HIV prevalence in Zambia: Relative mean standard errors (RMSE) show lower mean standard errors for the Fay-Herriot small area estimations over the survey-based estimation for all the 74 Zambian districts

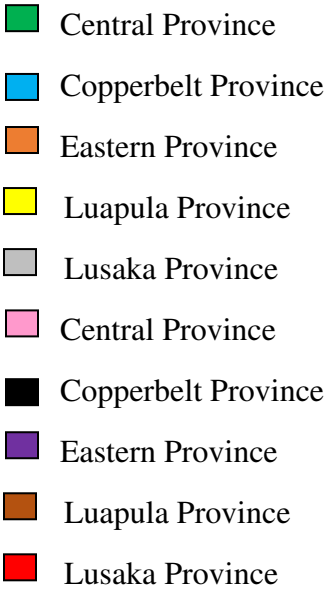

Figure 4: Zambia district HIV prevalence map: The color variations in the heat map shows the magnitude of the HIV prevalence in the 74 districts.

Additional File 1: The table shows a reduction in relative standard errors for the SAE HIV prevalence estimates (RSE_SAE) compared to the direct HIV prevalence estimates (RSE_RAW).

Figures

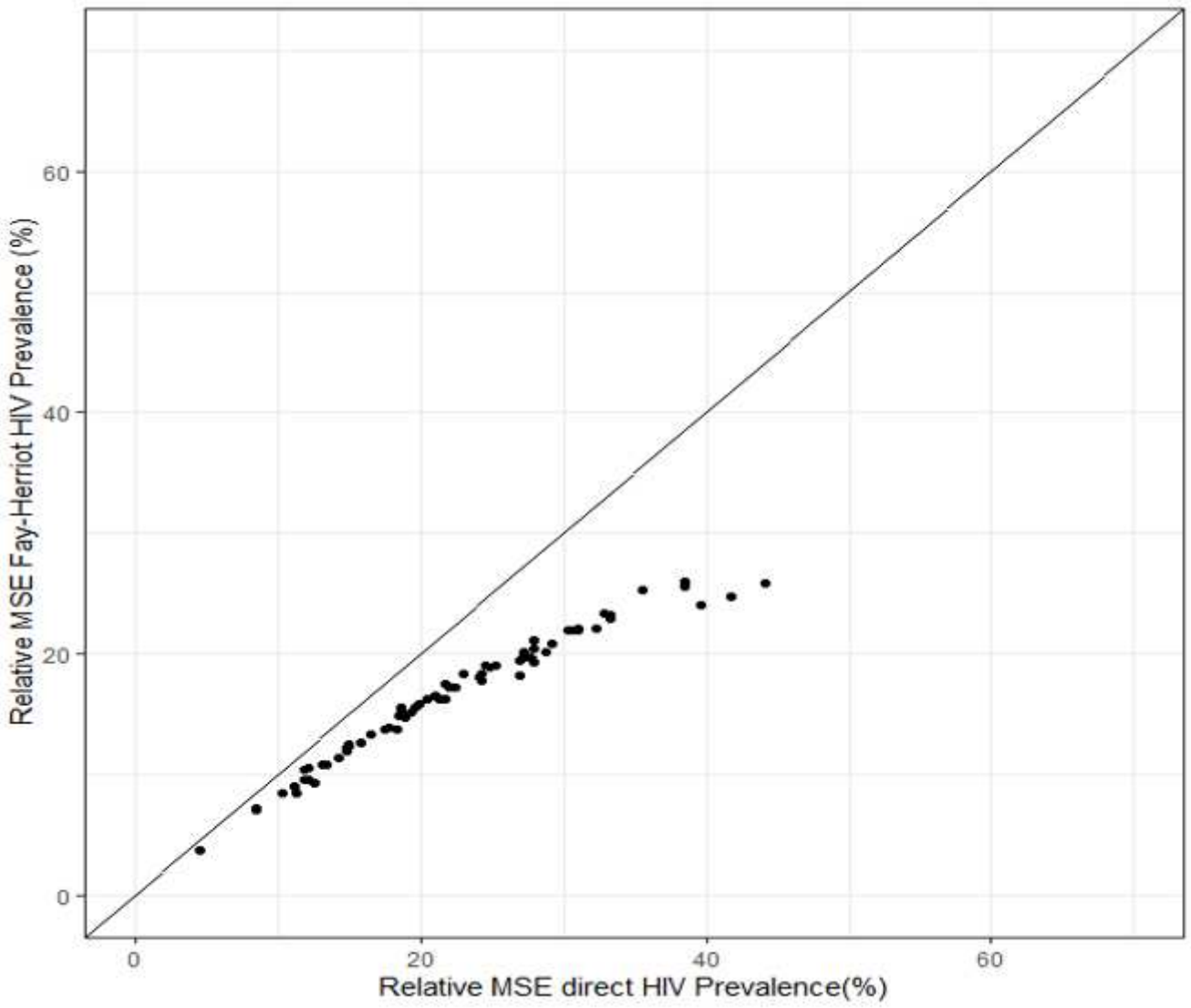

Figure 1

Relative Mean Standard Errors (RMSE) for the FH HIV prevalence estimates and survey-based prevalence estimates: The RMSE show lower mean standard errors for the Fay-Herriot small area estimations over the survey-based estimation for all the 74 Zambian districts

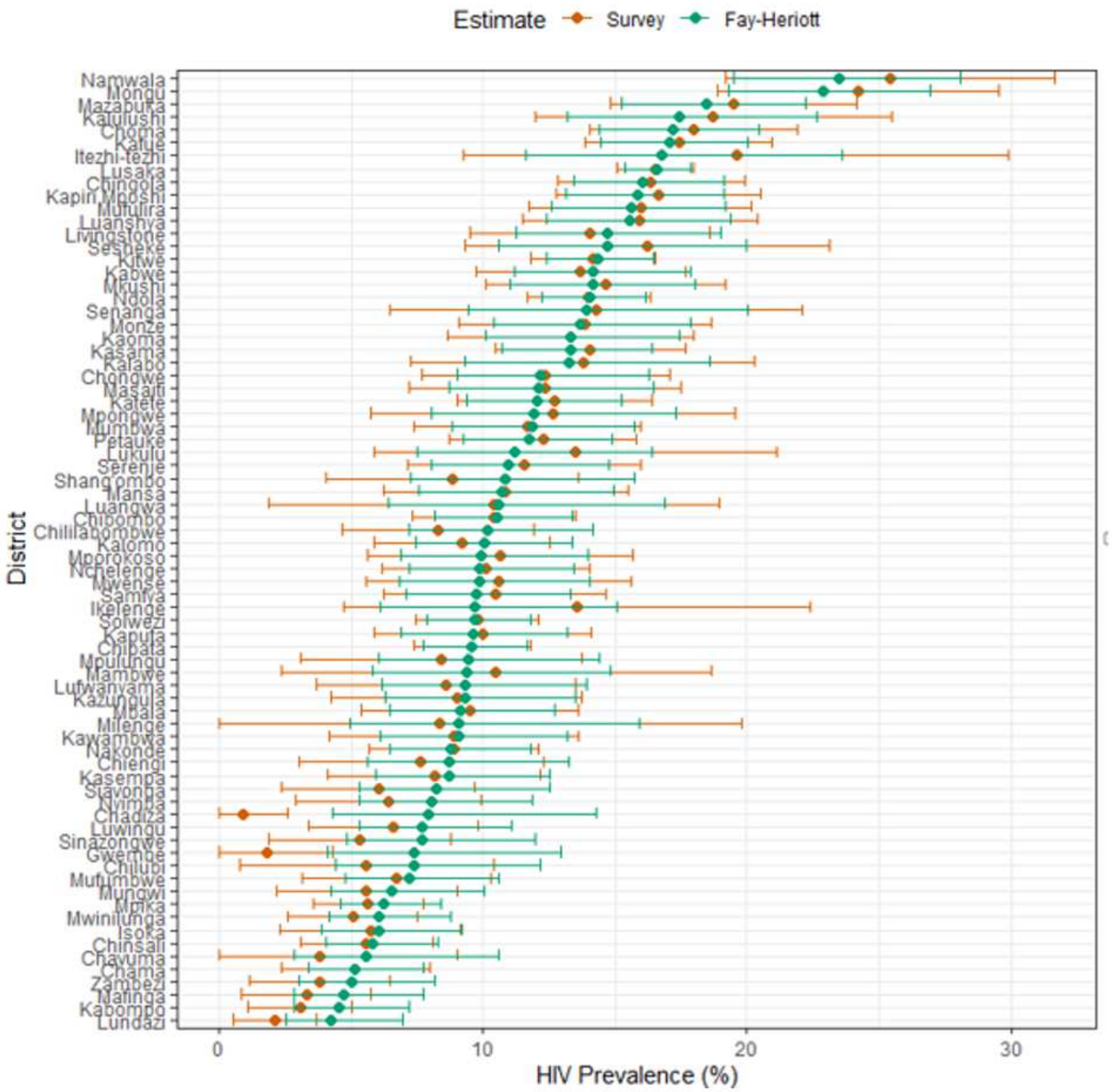

Figure 2

HIV prevalence estimates and confidence intervals for the FH and direct estimates in Zambia's districts: The confidence intervals of the FH estimates are narrower than those of the direct estimates for most of the districts

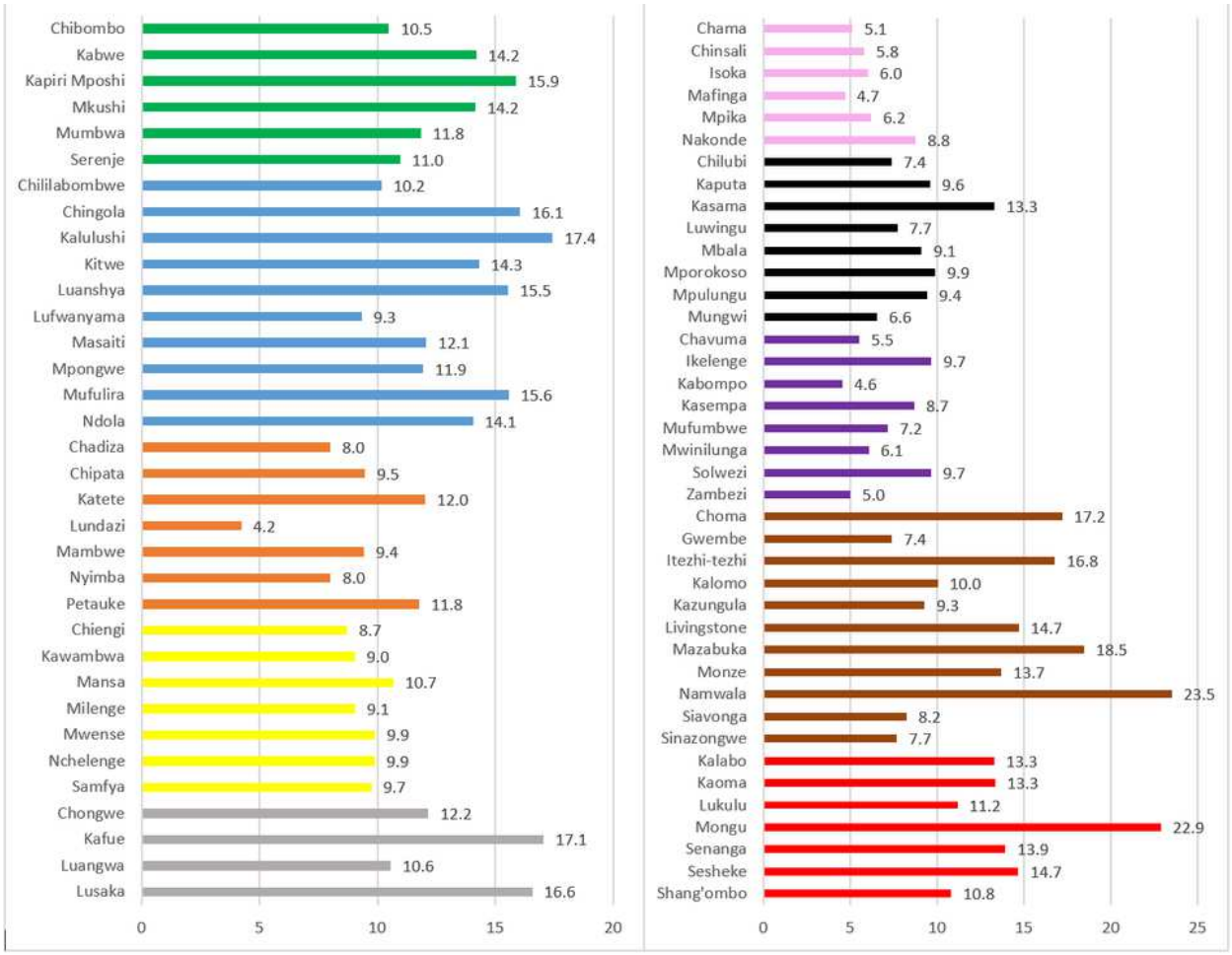

Figure 3

Estimated district HIV prevalence in Zambia: Relative mean standard errors (RMSE) show lower mean standard errors for the Fay-Herriot small area estimations over the survey-based estimation for all the 74 Zambian districts

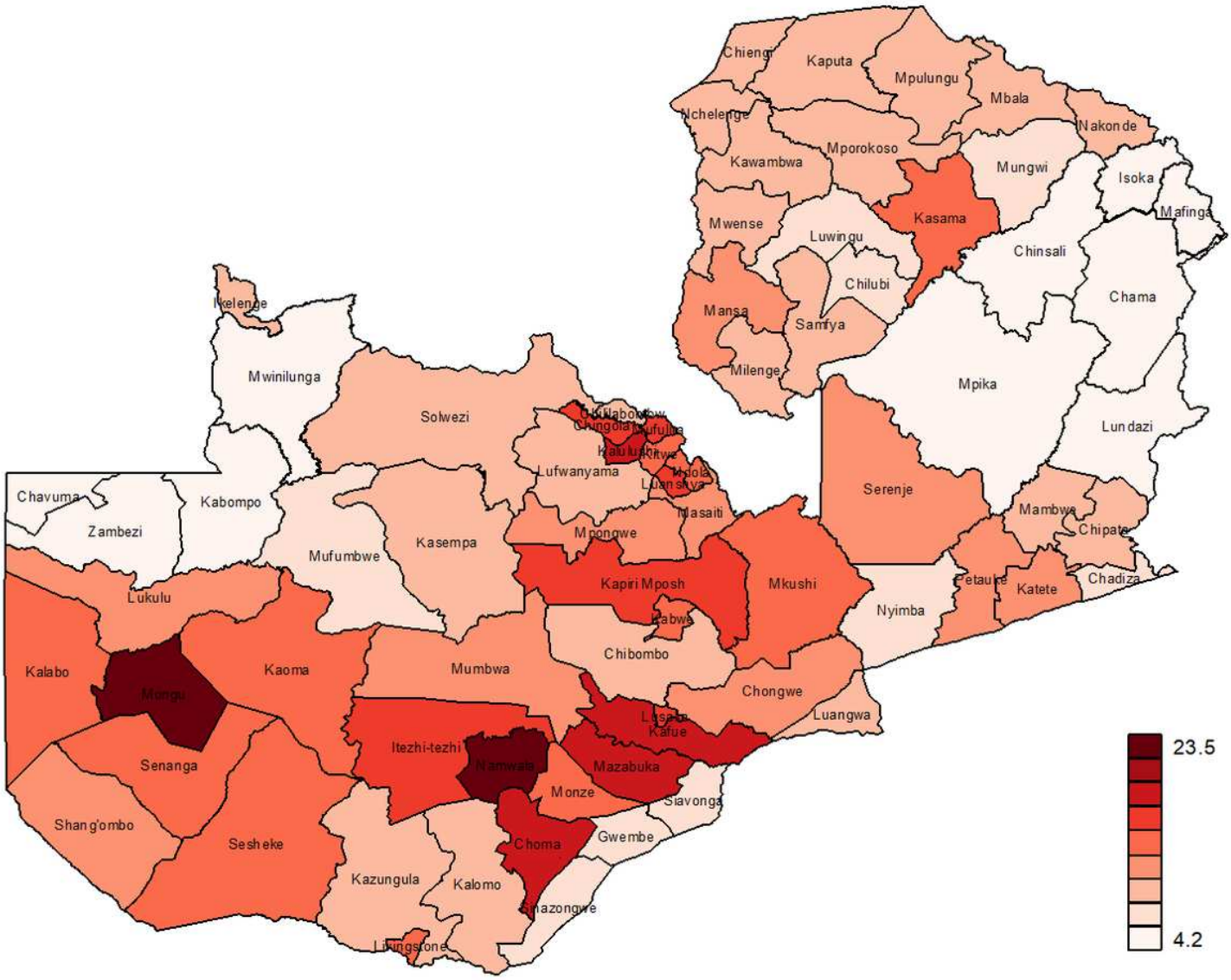

Figure 4

Zambia district HIV prevalence map: The color variations in the heat map shows the magnitude of the HIV prevalence in the 74 districts. Note: The designations employed and the presentation of the material on this map do not imply the expression of any opinion whatsoever on the part of Research Square concerning the legal status of any country, territory, city or area or of its authorities, or concerning the delimitation of its frontiers or boundaries. This map has been provided by the authors.

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- Additionalfile1.docx