Examining the association between HIV prevalence and socioeconomic factors among young people in Zambia: Do neighbourhood contextual effects play a role?

CURRENT STATUS: POSTED

Chola Nakazwe
Universitetet i Bergen

cholanakazwe@gmail.com Corresponding Author
ORCiD: https://orcid.org/0000-0002-7922-1222

Knut Martin Fylkesnes
University of Bergen. Centre for International Health, Department of Global Public Health and Primary Care

Charles Michelo
University of Zambia, School of Public Health

Ingvild F. Sandøy
University of Bergen, Centre for International Health, Department of Global Public Health and Primary Care

DOI: 10.21203/rs.3.rs-15670/v1

SUBJECT AREAS
Health Economics & Outcomes Research Health Policy

KEYWORDS
HIV infections, socioeconomic factors, young people, risk factors, multilevel modelling, Zambia
Abstract

**Background:** Evidently, socioeconomic factors are strongly associated with HIV infection, however, the pattern of association varies. Understanding how these factors influence HIV infection can help public health authorities design targeted comprehensive prevention strategies.

**Methods:** We used multi-level logistic regression models to examine the association of individual and neighbourhood level variables on HIV prevalence based on data from the 2013-14 Zambia Demographic and Health Survey, a population-based cross-sectional survey. The analysis was restricted to young people (15–24 years) with HIV serostatus results (n= 11,751). HIV serostatus was the outcome variable and socioeconomic status was measured by wealth, education and employment.

**Results:** Overall, at individual level, an additional year of education was associated with 5% and 2% reduced odds of infection among young women and men, respectively. Conversely, relative wealth and employment were associated with increased odds of infection among women, while among men the socioeconomic factors were associated with reduced odds of HIV infection. Living in neighbourhoods with medium proportion employed residents and neighbourhoods with higher average level of education were associated with a tendency of higher odds of HIV infection. Thirty percent and 9% of the variation in HIV infection among young men and women, was attributed to neighbourhoods. Inclusion of individual variables to the intercept-only model accounted for 34.9% and 58.1% of the variance for men and women, and reduced the unexplained variance by an average of 17%. However, adding neighbourhood variables to the intercept-only model explained a relatively small part of the variance (0.04 for men and 0.06 for women), and adding the cluster variables to the individual variables only lead to a slight increase in R-squared (explained variance) (36.2% and 58.4%).

**Conclusion:** Our study shows that individual-level socioeconomic factors are clearly associated with young people’s vulnerability to HIV infection in Zambia and that community level factors are less
important than previous studies in selected areas of Zambia have indicated. However, the high
neighbourhood variation among men, which was mostly unexplained, indicates that there are
dynamics within communities, possibly affecting the way people interrelate, that can put them at
increased risk.

Background
Globally the HIV epidemic differs by region in terms of prevalence levels and patterns, with a high HIV
morbidity and mortality burden in sub-Saharan Africa (SSA) in particular. In 2017, 66% of the
1.8 million new HIV infections estimated globally occurred in this region, and AIDS-related deaths in
SSA were estimated at 660,000 (1). The risk of HIV infection is associated with a number of
behavioural risk factors including but not limited to multiple and concurrent sexual partnerships,
which again, are linked to individual demographic characteristics (such as age, sex, and marital
status) and socioeconomic factors (such as education, wealth, and employment).
The associations between HIV prevalence and socioeconomic factors have changed over time in SSA
as sexual behaviours have changed in various directions for different socioeconomic groups (2, 3). In
Zambia for instance, premarital sex declined for both females and males between 2000 and 2009.
There was also a substantial decline in multiple sexual partnerships among males, while among
females, multiple sexual partnerships remained stable, over the same period (2). In addition,
associations sometimes vary between and within countries because factors such as cultural practices
(e.g., sexual cleansing, concurrent partnerships), urban/rural mobility, etc. may vary (4). However,
understanding how these factors influence HIV infection at different times and in different settings
can help public health authorities and planners design targeted and comprehensive prevention
strategies.
Available evidence indicates that individual-level socioeconomic factors are strongly associated with
HIV infection, though the pattern of association is mixed and depends on sex, urban/rural residence,
and the phase of the epidemic. The phase of the epidemic refers to how the HIV epidemic has evolved
over time. Modifying the focus of prevention and control activities to different subpopulations is
critical as the epidemic evolves (5). Relative wealth appears to have a mixed influence on HIV risk.
Mishra et al., in 2007, examined the association between HIV prevalence and wealth across eight sub-Saharan countries and found that wealthier individuals were more likely to be HIV positive (6).

Similarly, Brodish, in an ecological study that examined the association between wealth inequality and HIV prevalence, found that GINI coefficients and relative wealth ratios were positive predictors of HIV serostatus (7). Msisha et al., examining the extent to which geographic variation in HIV is associated with differential distribution of individual correlates and risk factors in Tanzania in a 2008 study, also observed that area poverty was associated with lower odds of being HIV positive (8).

Exploring the connections between HIV serostatus and individual, household, and community socioeconomic resources in Kenya, Ishida et al. found that the burden of HIV shifted from higher to lower socioeconomic status (SES) subgroups between 2003 and 2007 (9), highlighting the fact that the relationship between SES and HIV is context- and time-specific. Other studies have found a negative wealth and HIV gradient (10–12). In a 2013 study, Magadi’s analysis of data from 20 SSA countries revealed that urban poor in SSA had significantly higher odds of HIV infection than their non-poor counterparts (10). Lopman et al., in a 2007 population-based cohort study in Eastern Zimbabwe, also found that HIV incidence was lower in the higher wealth tercile (11). Likewise, a 2007 longitudinal population-based study in rural South Africa by Barnighausen et al. revealed that individuals living in households with middle-level relative wealth were at the highest risk of HIV acquisition (12). The mixed findings may be attributed to the phase of the epidemic and to contextual and community factors. According to Gillespie et al., accounting for a number of underlying contextual factors such as urban/rural residence, community wealth and mobility, reduced the positive association between wealth status and HIV in most cases (4).

The association between employment and HIV prevalence also shows inconsistent patterns. Msisha et al., in their 2008 study investigating socioeconomic status and HIV seroprevalence in Tanzania, found that employment status was positively associated with high HIV prevalence. Men and women employed in professional (i.e., non-agricultural or manual-related) occupations were at a higher risk of infection compared to those in the agricultural sector. However, for men the association changed after adjusting for other covariates, with unemployed men having a three-fold higher risk of being
infected compared to those in the agricultural sector (13). Conversely, Bunyasi and Coetzee found no significant association between employment and HIV prevalence in their study based on the 2007 and 2008 Prevention of Mother to Child Transmission of HIV, Effectiveness in Africa and Research Linkages to HIV Care survey (PEARL study) in South Africa (14).

Education in general appears to be protective with regard to HIV risk (12, 15–17). Barnighausen et al. in a 2007 study found that educational attainment significantly reduced the hazard of becoming infected in a rural community in South Africa (12). Investigating change over time in HIV prevalence by educational attainment in selected community populations in Zambia, Michelo et al. found that there was a shift towards reduced risk of HIV infection in subgroups with higher education from 1995 and 2003 (16). The association has transitioned from a higher risk of HIV infection among the educated, in the early phase of the epidemic, to a lower risk of infection among the more highly educated as the epidemic matures. According to Hargreaves et al., studies conducted from 1996 onwards were more likely to find a lower risk of HIV infection among the most educated (17). Most studies on factors associated with HIV infection have primarily focused on individual-level factors, but neighbourhood-level factors can also be independently associated with HIV infection (4).

A systematic review by Peterson et al. in 2017, evaluating the influence of contextual factors on HIV/AIDS, sexually transmitted infections, and risky sexual behaviour in sub-Saharan Africa showed contextual factors to be significantly associated with HIV infection (18). Two studies conducted among young women in selected communities in Zambia between 1997–1998 and 2003 found that community-level factors were as strongly associated with HIV infection as individual factors (15, 19). Our study extends previous research by using nationally presentative data from the 2013–14 Zambia Demographic and Health Survey (ZDHS) and applying multilevel analysis to investigate the association between HIV infection and individual and neighbourhood-level wealth, employment, and education among young women and men two decades after the peak of the epidemic.

Methods

Setting

A SSA country called Zambia situated in Southern part of Africa, with an estimated population of 16.9
million people in 2018. It has a relatively young population, with a median age of 16.9 years (20). Twenty two percent of the population are aged between 15 and 24 years, with an estimated HIV prevalence of 6.6% (21).

Study Design
We reanalysed data from the 2013-14 Zambia Demographic and Health Survey (ZDHS). This is a cross-sectional study based on a two-stage stratified cluster sample design. A total sample of 18,052 households were selected for the 2013-14 ZDHS, which provided reliable estimates of population and health indicators at the national and provincial levels, and in rural and urban areas within provinces. Details of the sample design and methodology are reported in the 2013–14 ZDHS report and in an article we published, which is based on the same data (21, 22).

Study Population
In the 2013–14 ZDHS, the eligible study population included women aged 15–49 years and men aged 15–59 years who were either permanent residents of the sampled households or visitors present in the households on the night before the survey (21). A total of 16,411 women and 14,773 men were successfully interviewed, yielding response rates of 96.2% and 91.1%, respectively. The principal reason for non-response among eligible adults was failure to find individuals at home despite repeated visits, followed by refusal to be interviewed. Of the respondents who completed the structured interview, 29,007 respondents aged 15-59 years were tested for HIV, corresponding to a response rate of 87.2%. Young people aged 15–24 years accounted for 39.9% of the people tested or 11,571 in total. Further details about participation in the survey and HIV testing response rates can be found the the 2013-14 ZDHS (21).

HIV Data Collection Procedure
The 2013-14 ZDHS HIV testing protocol allowed for linking of the HIV results to the socio-demographic data. Eligible women and men who consented to HIV testing were asked to voluntarily provide about five drops of blood from a finger prick for anonymous testing. Each Dried Blood Spots (DBS) sample was given a bar code label, and a duplicate label was attached to the individual’s questionnaire, while the third bar code was attached to the blood sample transmittal form to track the blood sample from
the field to the central laboratory (21). The HIV Laboratory testing algorithm used for the DHS surveys follow the UNAIDS/WHO guidelines for HIV testing in population based surveys. This testing algorithm included Vironostika HIV antigen/antibody combination assay (Biomerieux) as the first assay and Enzygnost HIV Integral II assay (Dade Behring) as the second confirmatory testing assay. Both tests are Fourth-generation enzyme-linked immunoassay assays (ELISAs) (23). Western Blot was used as a third confirmatory test. Further details about the testing methodology can be found in the 2013–14 ZDHS report (21).

Variables
The selection of variables to include in the analysis of this article was based on the proximate determinants framework for factors affecting the risk of transmission of HIV developed by Boerma and Weir (24). According to this conceptual framework underlying variables, such as socioeconomic status influence proximate determinants, which in turn have an effect on biological mechanisms to influence health outcomes (i.e. HIV infection). The proximate determinants include among others concurrent sexual partnerships, number of sexual partners and condom use. The conceptual framework helps to understand the causal pathways from distal socioeconomic factors to HIV infection The focus was on underlying determinants, i.e., the socioeconomic context of the neighbourhood (neighbourhood educational attainment, wealth, and employment status) and individual demographic and socioeconomic factors. The operational definitions are as follows:

**Dependent variable:**
The dependent variable for this study was HIV status, which is defined as serostatus determined by testing blood samples collected from each consenting individual (0 indicating “HIV negative” and 1 “HIV positive”).

**Independent variables:**
*Individual level:* The following variables were included in the analysis: sex, age, marital status, residence, educational attainment, wealth, and employment status. The **wealth index score** from the ZDHS was used, and separate wealth tertiles for urban and rural populations were created to reduce residence bias. The wealth score was calculated from the first component of principal
component analysis (PCA) of household assets, housing characteristics, and access-to-amenities data (e.g., roof and floor material, electricity, water supply, possession of goods such as a bicycle and television, and so forth). The methodology used to calculate the wealth index is based on the Filmer and Pritchett approach and a detailed explanation is available in the Measure DHS report (25).

Educational attainment was defined as the reported number of years spent in school, and this was used as a continuous variable in the regression model. Employment status was measured by asking respondents if they had been working in the 12 months preceding the survey and this was then categorised as unemployed and employed. Occupation was defined as the main type of work done in the 12 months preceding the survey, and this was classified using the ILO International Standard Classification of Occupation Codes (ISCO). However, this variable was not included in the regression model, as it partly represents the same information as employment status.

**Neighbourhood level:** In this study, neighbourhoods were defined as geographic areas located in close proximity to each other, and an enumeration area was used as a proxy for a neighbourhood. Variables describing the characteristics of the neighbourhoods were derived by aggregating individual responses within each cluster for all respondents and then categorising the means or proportions into three levels (low, medium, and high). Neighbourhood educational attainment was derived by calculating the mean number of years of school attained by the individuals who were interviewed in each enumeration area. Neighbourhood employment was the proportion of individuals interviewed in each cluster who were categorised as employed in the last 12 months. Neighbourhood relative wealth was derived by calculating the mean of the wealth scores of all respondents in the clusters.

**Analysis**

We restricted the analyses to young people (15–24 years). We conducted the analyses of both sexes and then stratified by women and men. Analyses were carried out in two phases. In the first phase we explored bivariate associations between HIV serostatus and individual demographic and socioeconomic factors, and neighbourhood-level socioeconomic factors. In the second phase, we used multilevel mixed-effects logistic regression analyses on HIV status with associated factors, which we conducted using Stata 15. Multilevel modelling takes into account both individual and neighbourhood-
level variability. The bivariate and multivariate analyses also controlled for the potential confounder age, which was adjusted for as a linear effect. We tested five separate models to examine the association between HIV prevalence and socioeconomic factors. Model 1 was the null or random intercept-only model, which did not include any socioeconomic or demographic factors. Subsequent models gradually included socioeconomic and demographic variables, starting with individual-level variables only, then neighbourhood-level variables only, and finally included both individual and neighbourhood-level variables (corresponding to model 2, model 3 and model 4, respectively). The final model included only significant variables at both individual and neighbourhood level. The log likelihood test was used for assessing the goodness-of-fit of the models and to determine whether adding independent variables to the intercept-only model significantly improved the fit.

Multilevel statistics estimated were, explained variance, the interclass correlation (ICC)), and log likelihood tests. To estimate explained variance ($R^2$), we use the approach explained in Hox, J (2010), and proposed by Snijders and Bosker (1999) for multilevel logistics regression models, which does not rely on the likelihood. The estimated variance is decomposed into the lowest-level residual variance ($\pi^2/3 = 3.29$), the second-level variance ($\rho$) and the variance of the linear predictor from the fixed part of the model ($\sigma^2$) (26). The explained variance is estimated using the formula:

**See formula 1 in the supplemental files.**

The ICC or rho indicates the proportion of the variance explained by the grouping structure in the population, which in our case are the neighbourhoods (26). This is same as the unexplained neighbourhood-level variance. We tested for interactions in the final model, for individual and neighbourhood variables, but none were found to be statistically significant.

**Ethics**

Ethical approval was obtained from the Tropical Disease and Research Centre (TDRC) ethical committee, the institutional board of ICF International, and the Centers for Disease Control and Prevention (CDC) Atlanta research ethics review board. Participation in the surveys was obtained by
soliciting verbal informed and voluntary consent. Participants were informed that the survey’s HIV testing results were anonymous. Home-based counselling and testing were offered in parallel to participants following the national HIV testing algorithm to ascertain HIV infection status for respondents who consented. Concurrent HIV testing with Determine™ HIV ½ (Alere Healthcare) and Uni-Gold™ (Trinity Biotechnology) was the home-based testing procedure, and nurses and lay counsellors provided pre- and post-test counselling. If either of the rapid tests was HIV reactive (positive), the respondent was referred to the nearest health facility for further assessment, treatment, and care.

Results

Participation Distribution

In total, 11,571 young people aged 15–24 years were interviewed and consented to HIV testing. Fifty one percent were living in rural areas, while 49% resided in urban areas. There were more young women than men in the study population (52% compared to 48% in both rural and urban).

HIV Prevalence and Associated Factors

An estimated 6.6% of the young people were HIV positive. The prevalence was notably higher among young women and men residing in urban areas than those in rural areas (10.7% and 7.3% compared with 4.8% and 3.6%, respectively). Age was associated with higher odds of being infected. Young people aged 20–24 were twice as likely to be infected as those in the younger age group of 15–19 years. Formerly married young women were at increased odds of being infected (aOR 1.69, 95% CI 1.17–2.44) compared to the never-married, when adjustment was made for age (formerly married include those who were separated, divorced and widowed). Conversely, married women had lower odds of being infected (aOR 0.73 95% CI 0.58–0.92) (Table 1).

HIV prevalence and age-adjusted odds ratios varied by the selected socioeconomic factors. HIV prevalence did not differ significantly by education. This was similar for both sexes. Among both young men and women, those with the highest wealth status had higher odds of being infected (aOR 1.57, 95% CI 1.08–2.29 for men, and aOR 2.29, 95% CI 1.76–3.00 for women) than those in the poorest tertile. Employment was associated with lower odds of being infected (aOR 0.81, 95% CI
0.69–0.96), but professional occupation was associated with higher odds of being infected among both young men and women (aOR 1.47, 95% CI 0.99–2.17 and aOR 1.51, 95% CI 1.18–1.92, respectively), compared to those who were unemployed. In contrast, agricultural occupation was associated with lower odds of HIV infection among both sexes. Living in neighbourhoods with a high proportion of employed residents was associated with lower odds of HIV infection among young women and men. Conversely, living in a neighbourhood with high wealth and education status was associated with higher odds of HIV infection (Table 1).

The analysis was also stratified by rural/urban residence, and some differences disappeared for selected variables, indicating interaction with residence. High wealth status was protective for young women residing in urban areas (aOR 0.68, 95% CI 0.51–0.91), but in rural areas it was associated with increased odds of being infected (aOR 1.64, 95% CI 1.08–2.47). For young men, the association between wealth status and HIV prevalence was associated with reduced odds of HIV infection in both urban and rural areas (aOR 0.79, 95% CI 0.51–1.24 and aOR 0.98, 95% CI 0.54–1.75, respectively).

Cluster-level variables were not significantly associated with HIV prevalence after stratification (results not shown) (see Additional file 1).

When data from young men and women were pooled, the intercept-only model showed that the overall HIV prevalence varied across neighbourhoods with a neighbourhood difference of 18%, with strong evidence of neighbourhood variation (Table II). In Model 2, the addition of individual-level variables to the intercept-only model explained more of the variance than the addition of neighbourhood variables (it explained a variance of 51% and a neighbourhood difference of 15% versus 5% and 15%, respectively). The inclusion of both individual and neighbourhood variables in Model 4 explained 51.4% of the variance in HIV prevalence, and the log likelihood test showed that the fit of the model significantly improved compared to the intercept model. The following variables were associated with higher odds of being HIV positive in Model 5: age, urban residence, formerly married, and residing in neighbourhoods with high educational attainment and medium-level employment. Educational attainment at the individual level was significantly associated with lower odds of being infected. In Model 5, where only statistically significant variables at the individual and
neighbourhood levels were included, the explained variance was still about 50%, and the neighbourhood difference was 15% (Table 2).

Among young males, the intercept-only model shows that HIV prevalence varied by 30% across neighbourhoods, but with weak evidence of neighbourhood variation. The inclusion of both individual and neighbourhood variables explained 36% of the variance in HIV prevalence, with a neighbourhood difference of 27%. The only background factors with strong evidence of association with HIV infection were age and urban residence (Table 3).

Among females, the intercept-only model shows that HIV prevalence varied across neighbourhoods, with a statistically significant neighbourhood difference of 9%. The addition of both individual and neighbourhood variables in Model 4 explained 58% of the variance, with a neighbourhood difference of 4%. The log likelihood test showed that the inclusion of both individual and neighbourhood variables to the intercept model improved the fit of the model. Residing in urban areas, older age, being formerly married, and residing in neighbourhoods with medium-level employment were associated with higher odds of being HIV positive. Educational attainment at the individual level was associated with lower odds of being infected. At the neighbourhood level, education was associated with increased odds of HIV infection (Table 4).

Discussion

Our analysis indicates that both individual and neighbourhood socioeconomic factors influence the likelihood of young people being infected with HIV in Zambia. At the individual level, older age and urban residence were found to be strongly associated with increased odds of infection for both young men and women. Formerly married women were also at increased odds of HIV infection. Education was associated with a reduced risk of HIV infection, the association was particularly strong among women, but among men there was no strong association. Living in communities with either a high proportion of educated respondents, medium proportion being employed, or higher average wealth status was associated with higher odds of HIV infection. In the full multivariate model only neighbourhood education and employment remained strong predictors of high odds of HIV infection,
and this was limited to young women.

The intraclass correlation (ICC), from the intercept-only model shows that 30% and 9% of the variation in HIV infection among young men and women, respectively, was attributed to neighbourhoods. When neighbourhood variables were added into the intercept-only model, there was a decline in the unexplained variance of HIV prevalence of 6% for young men and 33% for young women, and the explained variance was relatively small (4% vs. 6% for men and women, respectively). This is an indication that contextual factors are important in the likelihood of HIV infection among men, but in this study, most of the neighbourhood variation was unexplained. In contrast, the explained variance in the model including both individual and neighbourhood variables was 36% for men and 58% for women, indicating that individual-level factors have a more important role than neighbourhood-level factors in young people’s susceptibility to HIV infection. This is contrary to what previous studies in selected areas of Zambia have showed (15, 19).

Education was found to reduce the risk of infection in young women. An additional year of educational attainment reduced the odds of being infected by 4%. As mentioned earlier, some studies have documented similar findings (12, 16, 17). Bunyasi and Coetzee reported that every additional year of formal education was associated with a 10% and 1% decrease in risk of HIV infection in two specific regions of South Africa in 2007–2008 (14). A systematic review assessing the association between education and HIV by Hargreaves et al. found that studies in sub-Saharan Africa done from 1996 onwards showed a lower risk of infection among the most educated (17). We also found that in the multivariate analysis, after adjusting for other individual and neighbourhood factors, the protective effect of individual education was maintained. It may reflect that young women with high education may feel more empowered to make decisions that can protect them from HIV infection, contrary to the less-educated young women. Educated persons also tend to be early adopters of new practices; they internalize relevant information and translate this knowledge into behavioural change (27). We could have expected a similar strong association among male respondents, but this does not seem to be the case. Our results are in contrast to a study in Zambia, which found education to reduce the risk of HIV infection among men. Michelo et al. showed a steeper decline in HIV prevalence over time.
among higher-educated young men and women (16). This study was, however, conducted at a different stage of the epidemic, between 1993 and 2003.

It has been postulated that the spread of education changes the community environment and promotes socially acceptable behaviour, i.e., delayed sexual debut, limiting sexual partners, etc. (27). Therefore, people living in neighbourhoods with high educational attainment would be more likely to adopt health-promoting behaviours. Thus, it is logical to expect that neighbourhood-level educational attainment might also be protective. To the contrary, our study found that neighbourhood-level educational attainment was associated with increased risk of HIV infection. Van Stam et al., in a study assessing the relationship between sexual and reproductive health and educational attainment in eastern and southern Africa, taking into account the level of globalisation, obtained similar findings with regard to education increasing the risk of HIV and having more than four lifetime sexual partners. They postulate that educational attainment may be a booster of sexual networks, due to access to jobs and income, particularly in the less-globalised regions such as rural areas (28). These findings contradict the findings from the earlier studies done in selected areas of Zambia, which found that neighbourhood education was associated with a reduced risk of HIV infection. It could be that the two studies were not representative of the whole population or that the contextual factors affecting risk are not static and may change in their effect as the HIV epidemic evolves. Our findings indicate that the associations with socioeconomic status may change continuously over time, and that there is a need for more nuanced terms to describe the different stages of the HIV epidemic.

Living in a neighbourhood with medium-level employment was found to be a predictor of higher odds of HIV infection for young women. Neighbourhoods with a medium proportion of persons employed may be more heterogeneous in terms of the socioeconomic status of the inhabitants and have more disposable resources available than neighbourhoods with higher levels of unemployment, thereby negatively affecting young peoples’ sexual behaviour. We postulate that those with more resources may be more likely to engage in particular risky lifestyles, such as higher rates of partner change and multiple sexual partners, because of greater autonomy and spatial mobility. The statistical evidence of a positive association between neighbourhood employment and high HIV prevalence is surprising.
and needs further investigation.

Our study has some limitations. As in any cross-sectional study, the exposures and outcome were measured simultaneously in this given population. This limits our ability to draw causal inferences about the associations found. However, since the study was limited to young people, HIV infection was likely to be recent. This group is marginally affected by mortality and treatment, and only a small proportion would have been infected by their mother. According to the Zambia population-based HIV impact assessment (ZAMPHIA) survey, about 1% of children (0–14 years) were HIV positive, most likely infected by their mothers (29). Some important factors may have biased the associations with neighbourhood characteristics in our study, such as using a cluster as a proxy of neighbourhoods. Clusters which are defined as census enumeration areas are not necessarily representative of naturally occurring neighbourhoods where individuals reside. Studies have, however, found them to be small enough to be a useful proxy (30). A high proportion of the total variance was not explained by the neighbourhood variables used, especially among men. This could be partly because there was no information on more specific community-level characteristics that have been associated with HIV prevalence in other studies, such as the availability of markets, health facilities, bars, and proximity to trading areas and major roads. Mobility, which has been linked to the risk of HIV infection in numerous studies, was also not measured in our study, and thus we do not know how long respondents had lived in the neighbourhood where they were interviewed.

Conclusion

Our study indicates that individual-level socioeconomic factors play a more important role in young people’s vulnerability to HIV infection in Zambia. Though accounting for a relatively small proportion of the explained variance, the fact that medium neighbourhood employment was associated with increased odds of HIV infection in the full multivariate models indicates that community-level factors need to be considered by prevention programmes. Further, the high neighbourhood variation among men, which was mostly unexplained, indicates that there are dynamics within communities, possibly affecting the way people interact with each other, that can put them at increased risk. Therefore, it may be worthwhile to also invest in community-level interventions that impact individual practices
and contribute to social cohesion and supportive environments.

Declarations

Ethics approval and consent to participate

All procedures performed in the original study involving human participants were in accordance with the ethical standards of the institutional review board of ICF international, Tropical Disease and Research Centre (TDRC) ethics committee, the Centers for Disease Control and Prevention (CDC) Atlanta research ethics review board. Verbal Informed consent was obtained from all individual participants aged 18-59 years or emancipated minors (minors aged 15-17 years who are married, have children and/or are living without parental/guardian supervision). Emancipation is the achievement of adult legal status before the usual age of majority due to marriage, motherhood, military service, or other specific circumstances. Likewise, minors may be considered emancipated if they are functionally independent of parents, because of marriage or other circumstances (31). In Zambia a minor means a person below the age of eighteen years, and the National Health Ethics Act, 2013 stresses the importance of obtaining the consent of the parent or legal guardian of the minor; and assent from the minor, who is capable of understanding the nature and potential risks and benefits of the health research (32). Therefore, in this study, for minors aged 15-17 years, assent was obtained from the minor and permission/consent was obtained from the parents or guardians of the minors. All the three ethics committees approved the ethical procedure of obtaining verbal informed consent from study participants. The ethics committees also approved the procedure of obtaining verbal consent/permission from the parents of minors. The primary study protocol did not specify the reason why verbal consent was used instead of written consent. However, the consent statement was read in its entirety to the study participants before the interviewer signed the consent form, indicating whether the participant agreed to take part in the study or not.

Consent for Publication

Not applicable

Availability of data and materials

The datasets used and/or analysed during the current are available from the Measure DHS website
Competing Interests

The authors declare that they have no competing interests.

Funding

Not applicable

Authors’ contributions

CN conceived and designed the study. CN conducted the statistical analysis and wrote the first draft of the paper. IS contributed to writing the manuscript. Authors IS KF CM provided guidance on the analysis, interpretation of results and revised the manuscript critically for important intellectual content. CN in collaboration with the ZDHS study team was involved in the data collection and analysis of the primary study, the 2013-14 ZDHS. All authors gave approval on the final version of the manuscript to be published.

Acknowledgments

We would like to acknowledge the Norwegian government for providing funding for this study through the Quota programme. Further acknowledgement is made to the funders of the Zambia Demographic and Health Survey (ZDHS), the Government of Zambia, and ICF International with funding from the United States Agency for International Development (USAID). Special acknowledgements also go to the study participants and the Zambian study team that implemented the 2013–14 Zambia Demographic and Health Survey.

Authors’ Information

Not applicable

Abbreviations

AIDS: Acquired Immunodeficiency Syndrome

CDC: Centers for Disease Control and Prevention

DBS: Dried Blood Spots

DHS: Demographic and Health Surveys

EAs: Enumeration Areas
ELISA: enzyme-linked immunoassay assay
HIV: Human Immunodeficiency Virus
ICC: Interclass Correlation
ILO: International Labour Organisation
ISCO: International Standard Classification of Occupation Codes
PCA: Principal Component Analysis
PEARL: Prevention of Mother to Child Transmission of HIV, Effectiveness in Africa and Research Linkages to HIV Care survey
PSU: Primary Sampling Unit
SES: Socioeconomic Status
SSA: Sub-Saharan Africa
TDRC: Tropical Disease and Research Centre
ZAMPHIA: Zambia population-based HIV impact assessment survey
ZDHS: Zambia Demographic and Health Survey

References
1. Joint United Nations Programme on HIV/AIDS U. UNAIDS Data 2018. 2018.
2. Kayeyi N, Fylkesnes K, Wiium N, Sandoy IF. Decline in sexual risk behaviours among young people in Zambia (2000-2009): do neighbourhood contextual effects play a role? PLoS One. 2013;8(5):e64881.
3. Sandøy IF, Michelo C, Siziya S, Fylkesnes K. Associations between sexual behaviour change in young people and decline in HIV prevalence in Zambia. BMC Public Health. 2007;7:60.
4. Gillespie S, Kadiyala S, Greener R. Is poverty or wealth driving HIV transmission? AIDS. 2007;21:S5-S16.
5. Blanchard JF. Populations, pathogens, and epidemic phases: closing the gap between theory and practice in the prevention of sexually transmitted diseases. Sexually
6. Mishra V, Assche SBV, Greener R, Vaessen M, Hong R, Ghys PD, et al. HIV infection does not disproportionately affect the poorer in sub-Saharan Africa. AIDS. 2007;21:S17.

7. Brodish PH. An association between neighbourhood wealth inequality and HIV prevalence in sub-Saharan Africa. J Biosoc Sci. 2015;47(3):311-28.

8. Msisha WM, Kapiga SH, Earls FJ, Subramania SV. Place Matters: multilevel investigation of HIV distribution in Tanzania. 2008.

9. Ishida K, Arnold M, Stupp P, Kizito P, Ichwara J. Exploring the connections between HIV serostatus and individual, household, and community socioeconomic resources: evidence from two population-based surveys in Kenya. Soc Sci Med. 2012;74(2):185-95.

10. Magadi MA. The disproportionate high risk of HIV infection among the urban poor in sub-Saharan Africa. AIDS Behav. 2013;17(5):1645-54.

11. Lopman B, Lewis J, Nyamukapa C, Mushati P, Chandiwana S, Gregson S. HIV incidence and poverty in Manicaland, Zimbabwe: is HIV becoming a disease of the poor? AIDS. 2007;21 Suppl 7:S57-66.

12. Barnighausen T, Hosegood V, Timaeus IM, Newell ML. The socioeconomic determinants of HIV incidence: evidence from a longitudinal, population-based study in rural South Africa. AIDS. 2007;21 Suppl 7:S29-38.

13. Msisha WM, Kapiga SH, Earls F, Subramanian SV. Socioeconomic status and HIV seroprevalence in Tanzania: a counterintuitive relationship. Int J Epidemiol. 2008;37(6):1297-303.

14. Bunyasi EW, Coetzee DJ. Relationship between socioeconomic status and HIV infection: findings from a survey in the Free State and Western Cape Provinces of
South Africa. BMJ Open. 2017;7(11):e016232.

15. Kayeyi N, Sandøy IF, Fylkesnes K. Effects of neighbourhood-level educational attainment on HIV prevalence among young women in Zambia. BMC Public Health. 2009;9(1):310.

16. Michelo C, Sandøy IF, Fylkesnes K. Marked HIV prevalence declines in higher educated young people: evidence from population-based surveys (1995-2003) in Zambia. AIDS. 2006;20(7):1031.

17. Hargreaves JR, Bonell CP, Boler T, Boccia D, Birdthistle I, Fletcher A, et al. Systematic review exploring time trends in the association between educational attainment and risk of HIV infection in sub-Saharan Africa. 2008;22(3):403-14.

18. Ward-Peterson M, Fennie K, Mauck D, Shakir M, Cosner C, Bhoite P, et al. Using multilevel models to evaluate the influence of contextual factors on HIV/AIDS, sexually transmitted infections, and risky sexual behavior in sub-Saharan Africa: a systematic review. Ann Epidemiol. 2018;28(2):119-34.

19. Gabrysch S, Edwards T, Glynn JR. The role of context: neighbourhood characteristics strongly influence HIV risk in young women in Ndola, Zambia. Tropical Medicine & International Health. 2008;13(2):162-70.

20. Central Statistical Office. 2010 Census of Population National Analytical Report - 2010 Census. Lusaka, Zambia: Central Statistical Office; 2012.

21. Central Statistical Office (CSO) MoH, and ICF International. Zambia Demographic and Health Survey 2013-14. Rockville, Maryland, USA: Central Statistical Office, Ministry of Health, ICF International; 2014.

22. Nakazwe C, Michelo C, Sandoy 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;19(1):432.
23. UNAIDS/WHO. Guidelines for Measuring National HIV Prevalence in Population-based Surveys. Geneva: UNAIDS and WHO; 2005.

24. Boerma JT, Weir SS. Integrating Demographic and Epidemiological Approaches to Research on HIV/AIDS: The Proximate-Determinants Framework. The Journal of Infectious Diseases. 2005;191(Supplement_1):S61-S7.

25. Rutstein SO, Johnson K, Macro O. The DHS wealth index: ORC Macro, MEASURE DHS; 2004.

26. Hox JJ. Multi-level Analysis: Techniques and Applications. New York: Routledge; 2010.

27. Vandemoortele J, Delamonica E. The education vaccine against HIV. Current Issues in comparative education (Columbia University). 2000;3.

28. Van Stam MA, Michielsen K, Stroeken K, Zijlstra BJ. The impact of education and globalization on sexual and reproductive health: retrospective evidence from eastern and southern Africa. AIDS Care. 2014;26(3):379-86.

29. Ministry of Health Z. Zambia Population Based HIV Impact Assessment (ZAMPHIA) 2016: Final Report. Lusaka: Ministry of Health; 2019.

30. Diez Roux AV. Investigating Neighborhood and Area Effects on Health. American Journal of Public Health. 2001;91(11):1783-9.

31. Santelli J, Haerizadeh S, McGovern T. Inclusion with Protection: Obtaining informed consent when conducting research with adolescents. Florence, Italy: UNICEF Office of Research; 2017.

32. Health Research Act 2013, (2013).

Tables

Table I: HIV prevalence and age-adjusted odds ratios by background characteristics and sex among young people 15–24 years, 2013–14, Zambia.
| Individual Variable     | Number | HIV % | AOR (95% CI) | P-value | Number | HIV % | AOR (95% CI) | P-value |
|-------------------------|--------|-------|--------------|---------|--------|-------|--------------|---------|
| Residence               |        |       |              |         |        |       |              |         |
| Rural                   | 5897   | 4.2   | 1            |         | 2836   | 3.6   | 1            |         |
| Urban                   | 5674   | 9.1   | 2.57 (2.09 - 3.15) | < 0.001 | 2717   | 7.3   | 2.19 (1.62 - 3.22) | < 0.01 |
| Age                     |        |       |              |         |        |       |              |         |
| 15-19                   | 6519   | 4.4   | 1            |         | 3246   | 4.1   | 1            |         |
| 20-24                   | 5052   | 9.4   | 2.31 (1.98 - 2.70) | < 0.001 | 2307   | 7.3   | 1.82 (1.40 - 2.35) | < 0.01 |
| Marital Status          |        |       |              |         |        |       |              |         |
| Never married           | 8585   | 5.8   | 1            |         | 4947   | 5.2   | 1            |         |
| Married                 | 2657   | 8.0   | 0.98 (0.80 - 1.19) | 0.805   | 549    | 7.0   | 0.87 (0.55 - 1.36) | 0.542   |
| Formerly married        | 329    | 17.5  | 2.19 (1.55 - 3.09) | < 0.001 | 57     | 8.9   | 1.17 (0.41 - 3.34) | 0.737   |
| Education               |        |       |              |         |        |       |              |         |
| No education            | 279    | 8.5   | 1            |         | 93     | 7.2   | 1            |         |
| Primary                 | 4181   | 5.7   | 1.13 (0.68 - 1.88) | 0.648   | 1937   | 4.5   | 0.78 (0.28 - 2.16) | 0.632   |
| Secondary               | 6779   | 7.0   | 1.06 (0.64 - 1.77) | 0.812   | 3338   | 5.7   | 0.85 (0.31 - 2.34) | 0.760   |
| Higher than secondary   | 324    | 8.5   | 0.77 (0.39 - 1.15) | 0.444   | 181    | 9.1   | 0.88 (0.27 - 2.92) | 0.838   |
| Wealth                  |        |       |              |         |        |       |              |         |
| Low                     | 3369   | 4.3   | 1            |         | 1519   | 3.9   | 1            |         |
| Medium                  | 3566   | 6.7   | 1.66 (1.33 - 2.08) | < 0.001 | 1779   | 4.7   | 1.24 (0.86 - 1.79) | 0.557   |
| High                    | 4636   | 8.2   | 1.81 (1.43 - 2.28) | < 0.001 | 2255   | 7.1   | 1.57 (1.08 - 2.29) | 0.019   |
| Employment              |        |       |              |         |        |       |              |         |
| Not employed            | 6487   | 6.4   | 1            |         | 2373   | 5.4   | 1            |         |
| Employed                | 5084   | 6.9   | 0.81 (0.69 - 0.96) | 0.012   | 3180   | 5.4   | 0.79 (0.59 - 1.06) | 0.120   |
Table 2: Multilevel logistic regression models for HIV prevalence among young people aged 15–24 years in Zambia, both sexes, 2013–14

| BOTH SEXES | Model 1 | Model 2 | Model 3 |
|------------|---------|---------|---------|
| Unemployed | 6497    | 2379    | 607     |
| Professional | 1359 | 607     | 10.7    |
| Agricultural | 2950 | 1871    | 3.5     |
| Manual | 547    | 520     | 7.0     |
| Neighbourhood Variables | | | |
| Cluster Education | | | |
| Low | 3421 | 1646 | 2.9 |
| Medium | 3974 | 1894 | 5.6 |
| High | 4176 | 2013 | 7.3 |
| Cluster Wealth | | | |
| Low | 3472 | 1636 | 2.6 |
| Medium | 3326 | 1612 | 5.7 |
| High | 4774 | 2305 | 7.3 |
| Cluster Employment | | | |
| Low | 4816 | 2297 | 5.6 |
| Medium | 3624 | 1746 | 6.8 |
| High | 3132 | 1510 | 3.6 |
| Total | 11571 | 5897 | 5.4 |
### Fixed Effects

#### Independent Variables

| Variable       | AOR       | 95% CI       | AOR       | 95% CI       |
|----------------|-----------|--------------|-----------|--------------|
| Age            | 1.20      | (1.16 - 1.24)*** |           |              |
| Education      | 0.96      | (0.93 - 0.99)*** |           |              |
| Residence      |           |              |           |              |
| Rural          | 1         |              |           |              |
| Urban          | 2.59      | (2.02 - 3.33)*** |           |              |
| Marital Status |           |              |           |              |
| Never married  | 1         |              |           |              |
| Married/cohabiting | 1.03 | (0.83 - 1.27) |    1.03  | (0.83 - 1.27) |
| Formerly married | 2.15 | (1.51 - 3.06)*** | 2.15    | (1.51 - 3.06)*** |
| Wealth         |           |              |           |              |
| Low            | 1         |              |           |              |
| Medium         | 1.26      | 1.62)***     | 1.26      | 1.62)***     |
| High           | 1.15      | 1.55)        | 1.15      | 1.55)        |
| Employment     |           |              |           |              |
| Not employed   | 1         |              |           |              |
| Employed       | 0.89      | (0.75 - 1.05) | 0.89      | (0.75 - 1.05) |

#### Neighbourhood Variables

| Education       | AOR       | 95% CI       | AOR       | 95% CI       |
|-----------------|-----------|--------------|-----------|--------------|
| Low             | 1         |              |           |              |
| Medium          | 1.26      | 1.43)        | 1.26      | 1.43)        |
| High            | 1.58      | 2.40)***     | 1.58      | 2.40)***     |
| Wealth          |           |              |           |              |
| Low             | 1         |              |           |              |

*Note: AOR = Adjusted Odds Ratio, CI = Confidence Interval, *** p < 0.001, ** p < 0.01, * p < 0.05*
|               | Medium |            | High  |
|---------------|--------|------------|-------|
|               | 1.56   | 2.15***    | 1.85  |
|               |        | (1.20     | 2.85***|
| Employment    |        |            |       |
| Low           |        |            | 1     |
|               |        | (1.09     | 1.74***|
| Medium        |        | 1.37       |       |
|               |        | (0.80     | 1.40) |
| High          |        | 1.06       |       |
|               |        | (0.80     | 1.40) |

| Fixed-Effect Model Test |
|-------------------------|
| Wald                    | 263.47 |
| Degree of freedom       | 8      |

| Unexplained Neighbourhood-Level Variance (SE) |
|-----------------------------------------------|
| 0.721                                         |
| 0.584                                         |
| (0.10)***                                     |
| 0.579                                         |
| (0.10)*                                       |

| Model Statistics |
|------------------|
| Explained variance ( R²) | 0.510 |
| Intraclass correlation (ICC ) (rho) | 0.180 |
| Log likelihood       | 0     |
| Log likelihood*      | 278.05*** |

Figures with asterix are significant at the following * p<0.10, * *p<0.05, **** p<0.01; Log likelihood* - Tests whether adding individual and neighbourhood variables to the null model (model 1) significantly improves the fit of the model (P<0.05); Abbreviations R²- Explained variance; ICC – Inter-class correlation; AOR: age – adjusted Odds Ratio and CI: Confidence Interval.
Table 3: Multilevel logistic regression models for HIV prevalence among young people aged 15–24 years in Zambia, 2013–14

| MALE | Model 1 | Model 2 | Model 3 |
|------|---------|---------|---------|
|      | AOR     | 95% CI  | AOR     |
| Fixed Effects |         |         |         |
| Independent Variables |         |         |         |
| Age | 1.16 | (1.09 - 1.23)*** |         |
| Education | 0.99 | (0.93 - 1.04) |         |
| Residence |         |         |         |
| Rural | 1 |         |         |
| Urban | 2.46 | (1.60 - 3.76)*** |         |
| Marital Status |         |         |         |
| Never married | 1 |         |         |
| Married/co–habiting | 1.01 | 1.62 |         |
| Formerly married | 1.20 | 3.44 |         |
| Wealth |         |         |         |
| Low | 1 |         |         |
| Medium | 0.93 | 1.40 |         |
| High | 0.88 | 1.45 |         |
| Employment |         |         |         |
| Not employed | 1 |         |         |
| Employed | 0.86 | 1.17 |         |
| Neighbourhood Variables |         |         |         |
|          | Low | Medium | High |
|----------|-----|--------|------|
| Education | 1   | 1.27   | 1.46 |
| Wealth   | 1   | 1.80   | 1.87 |
| Employment| 1   | 1.16   | 0.93 |

Fixed-Effect Model Test

|          | Wald  | Degree of freedom | Unexplained Neighbourhood-Level Variance (SE) | Model Statistics |
|----------|-------|-------------------|---------------------------------------------|------------------|
|          | 53.96*** | 6                | 1.38 (0.28) | Explained variance ( R-squared) 0.349 |
|          | 22.39*** | 6                | 1.28 (0.27) | ICC (rho) 0.30  |

|          | Log likelihood | Log likelihood* |
|----------|----------------|-----------------|
|          | -1058.35       | 57.07***       |
|          | -1029.82       | <0.01          |
|          | -1046.46       | 24.44***       |

Figures with asterix are significant at the following * p<0.10, * *p<0.05, **** p<0.01; Log likelihood* - Tests whether adding individual and neighbourhood variables to the null model (model
1) significantly improves the fit of the model (P<0.05); Abbreviations R²- Explained variance; ICC - Inter-class correlation; AOR: age - adjusted Odds Ratio and CI: Confidence Interval.

Table 4: Multilevel logistic regression models for HIV prevalence among young people aged 15–24 years in Zambia, 2013–14

| FEMALE |                       | Model 1 | Model 2 | Model 3 |
|--------|------------------------|---------|---------|---------|
| Fixed Effects |                       |         |         |         |
| Independent Variables |                       |         |         |         |
| Age | 1.22 | 1.28*** |         |         |         |
| Education | 0.96 | 1.00)** |         |         |         |
| Residence |               |         |         |         |
| Rural | 1 |         |         |         |         |
| Urban | 2.51 | 3.29)** |         |         |         |
| Marital Status |               |         |         |         |
| Never married | 1 |         |         |         |         |
| Married/co—habiting | 0.82 | 1.05)* |         |         |         |
| Formerly married | 1.74 | 2.25)** |         |         |         |
| Wealth |               |         |         |         |
| Low | 1 |         |         |         |         |
| Medium | 1.57 | 2.14)** |         |         |         |
### Employment

| Status       | Value 1 | Value 2 |
|--------------|---------|---------|
| Not employed | 1       |         |
| Employed     | 1.12    | 1.38    |

### Neighbourhood Variables

#### Education

| Level | Value 1 | Value 2 |
|-------|---------|---------|
| Low   | 1       |         |
| Medium| 1.25    | 1.75    |
| High  | 1.55    |         |

#### Wealth

| Level | Value 1 | Value 2 |
|-------|---------|---------|
| Low   | 1       |         |
| Medium| 1.50    |         |
| High  | 1.85    | 2.00    |

### Employment

| Level | Value 1 | Value 2 |
|-------|---------|---------|
| Low   | 1       |         |
| Medium| 1.48    |         |
| High  | 1.10    | 1.38    |

### Fixed-Effect Model Test

| Test   | Value 1         | Value 2       |
|--------|-----------------|---------------|
| Wald   | 224.32***       | 66.96         |
| Degree of freedom | 8  | 6           |

### Unexplained Neighbourhood-Level Variance (SE)

| Value 1         | Value 2         | (SE) Value 3 |
|-----------------|-----------------|--------------|
| 0.339           | 0.167           | (0.09)***    |
|                 |                 | 0.196        |

### Model Statistics

| Test                  | Value 1         | Value 2       |
|-----------------------|-----------------|---------------|
| Explained variance (R-squared) | 0.581 | 0.056         |
| ICC (rho)             | 0.09            | 0.05          |
Figures with asterix are significant at the following * p<0.10, * *p<0.05, **** p<0.01; Log likelihood* - Tests whether adding individual and neighbourhood variables to the null model (model 1) significantly improves the fit of the model (P<0.05); Abbreviations R²: Explained variance; ICC - Inter-class correlation; AOR: age - adjusted Odds Ratio and CI: Confidence Interval.

Supplementary Files
This is a list of supplementary files associated with this preprint. Click to download. formulas.docx