SIS | 2022

51st Scientific Meeting
of the Italian Statistical Society

Caserta, 22-24 June

Università degli Studi della Campania *Luigi Vanvitelli*

Società Italiana di Statistica

www.unicampania.it

Book of the Short Papers

Editors: Antonio Balzanella, Matilde Bini, Carlo Cavicchia, Rosanna Verde

Città di Caserta

1222-2022 800 ANNI

UNIVERSITÀ DEGLI STUDI DI PADOVA

DIPARTIMENTO DI SCIENZE STATISTICHE

STATA

sas

UNIVERSITÀ DEGLI STUDI DEL SANNIO Benevento

Pearson

Matilde Bini (Chair of the Program Committee) - *Università Europea di Roma*
Rosanna Verde (Chair of the Local Organizing Committee) - *Università della Campania "Luigi Vanvitelli"*

PROGRAM COMMITTEE

Matilde Bini (Chair), Giovanna Boccuzzo, Antonio Canale, Maurizio Carpita, Carlo Cavicchia, Claudio Conversano, Fabio Crescenzi, Domenico De Stefano, Lara Fontanella, Ornella Giambalvo, Gabriella Grassia - Università degli Studi di Napoli Federico II, Tiziana Laureti, Caterina Liberati, Lucio Masserini, Cira Perna, Pier Francesco Perri, Elena Pirani, Gennaro Punzo, Emanuele Raffinetti, Matteo Ruggiero, Salvatore Strozza, Rosanna Verde, Donatella Vicari.

LOCAL ORGANIZING COMMITTEE

Rosanna Verde (Chair), Antonio Balzanella, Ida Camminatiello, Lelio Campanile, Stefania Capecchi, Andrea Diana, Michele Gallo, Giuseppe Giordano, Ferdinando Grillo, Mauro Iacono, Antonio Irpino, Rosaria Lombardo, Michele Mastroianni, Fabrizio Maturo, Fiammetta Marulli, Paolo Mazzocchi, Marco Menale, Giuseppe Pandolfi, Antonella Rocca, Elvira Romano, Biagio Simonetti.

ORGANIZERS OF SPECIALIZED, SOLICITED, AND GUEST SESSIONS

Arianna Agosto, Raffaele Argiento, Massimo Aria, Rossella Berni, Rosalia Castellano, Marta Catalano, Paola Cerchiello, Francesco Maria Chelli, Enrico Ciavolino, Pier Luigi Conti, Lisa Crosato, Marusca De Castris, Giovanni De Luca, Enrico Di Bella, Daniele Durante, Maria Rosaria Ferrante, Francesca Fortuna, Giuseppe Gabrielli, Stefania Galimberti, Francesca Giambona, Francesca Greselin, Elena Grimaccia, Raffaele Guetto, Rosalba Ignaccolo, Giovanna Jona Lasinio, Eugenio Lippiello, Rosaria Lombardo, Marica Manisera, Daniela Marella, Michelangelo Misuraca, Alessia Naccarato, Alessio Pollice, Giancarlo Ragozini, Giuseppe Luca Romagnoli, Alessandra Righi, Cecilia Tomassini, Arjuna Tuzzi, Simone Vantini, Agnese Vitali, Giorgia Zaccaria.

ADDITIONAL COLLABORATORS TO THE REVIEWING ACTIVITIES

Ilaria Lucrezia Amerise, Ilaria Benedetti, Andrea Bucci, Annalisa Busetta, Francesca Condino, Anthony Cossari, Paolo Carmelo Cozzucoli, Simone Di Zio, Paolo Giudici, Antonio Irpino, Fabrizio Maturo, Elvira Romano, Annalina Sarra, Alessandro Spelta, Manuela Stranges, Pasquale Valentini, Giorgia Zaccaria.

PUBLISHED BY PEARSON

WWW.PEARSON.COM

ISBN 9788891932310

Developing countries, migration and migrants

HIV Prevalence in some African Territories: Socio-Economic Drivers

La Prevalenza di HIV in alcuni Territori Africani: Fattori Socio-Economici

Micaela Arcaio,[1*] Daria Mendola,[1] Anna Maria Parroco[1]

Abstract In 2020, 35% of all HIV-positive people in the world lived in Eastern and Southern Africa. This work aims at assessing the relationship between socio-economic drivers and HIV prevalence at the sub-national level in these countries. The data used are drawn from the Demographic and Health Survey, in which a subset of respondents is tested for HIV. Using a fractional logistic regression model on clusters of individuals, middle-to-high wealth is positively associated with higher HIV prevalence, while a higher average number of children acts as a protective factor; moreover, higher proportions of people who have never been in sexual relationships lower their cluster's HIV prevalence but HIV-positive people are more likely to use condoms at last intercourse.

Abstract *Nel 2020 il 35% delle persone sieropositive nel mondo era concentrato in Africa del Sud e dell'Est. Questo studio mira a valutare la relazione tra fattori socio-economici e prevalenza di HIV a livello sub-nazionale in questi paesi. I dati derivano dalla Demographic and Health Survey, in cui un sotto-campione viene sottoposto al test dell'HIV. Utilizzando un modello di regressione logistica frazionaria su cluster di individui, il benessere medio-alto è associato positivamente a prevalenza di HIV più alta, mentre un numero medio di figli maggiore agisce da fattore protettivo; inoltre, maggiori proporzioni di persone che non hanno mai avuto relazioni sessuali abbassano la prevalenza del loro cluster ma persone sieropositive utilizzano con maggiore probabilità il profilattico.*

Key words: HIV, socio-economic status; fractional logistic regression; sexual behaviours.

[1] Micaela Arcaio, University of Palermo; email: micaela.arcaio@unipa.it;
Daria Mendola, University of Palermo; email: daria.mendola@unipa.it;
Anna Maria Parroco, University of Palermo; email: annamaria.parroco@unipa.it
*Corresponding author

1 Introduction

In 2020 an estimated 38 million people were living with HIV in the world. Among these people, about 50% are adult women (age 15+) and 5% are children aged 0-14 [1]. When it comes to HIV prevalence, it is estimated to be globally at around 0.7% among adults aged 15-49, with varying prevalence between both regions and countries. Indeed, the regional prevalence ranges between less than 0.1% in North Africa and the Middle East and 6.5% in Eastern and Southern Africa.

As for socio-economic drivers of the HIV epidemic, there is no evidence for an unequivocal relationship between some socio-economic characteristics of a certain population and HIV infection. Different authors have found both positive and negative association with education and wealth, with employment status and age, highlighting the heterogeneity in the relationship [3]. Indeed, some studies have found that higher standards of living, as well as higher household wealth, are positively associated with higher odds of HIV infection [4, 5, 6], while others have found that poorer people incur to income-generating strategies that put them at risk of seropositivity [7]. Education has also been shown to reduce vulnerability to HIV [6].

Focusing on the sexual transmission of HIV, this work aims to assess the relationship between socio-economic drivers and HIV prevalence at the sub-national level.

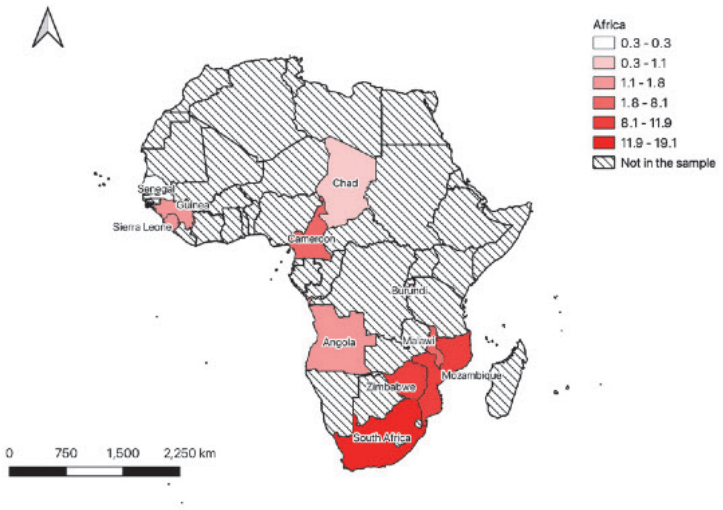
2 Data and Methods

The data used in this analysis are drawn from the Demographic and Health Survey (DHS). It is a nationally representative household survey that has covered over 90 countries in more than 350 surveys in the last forty years [8]. Standardized questionnaires are employed to improve comparisons in time and space on fertility, mortality, family planning, reproductive health, and HIV/AIDS – sub-samples are subjected to ELISA (enzyme-linked immunosorbent assay) tests to determine respondents' serostatus.

In this study, data are from 11 African countries: Angola, Burundi, Cameroon, Chad, Guinea, Malawi, Mozambique, Senegal, Sierra Leone, South Africa, and Zimbabwe. These were selected due to the availability of surveys in a period ranging from 2015 to 2019, choosing one survey (the most recent) for each country.

Indeed, despite accounting only for 3.21% of the global population in 2020 [2], these countries reached a total of 13.5 million cases in that year, accounting for almost 36% of all global HIV-positive patients [1]. The HIV prevalence in these countries ranges from 1% in Burundi to 19.1% in South Africa (see Figure 1).

Figure 1: *HIV Prevalence in the chosen countries*



Our sample is made of almost 150,000 people, aged 15-64, belonging to 11 countries. In the original dataset, households that participated in the survey are grouped in the so-called "clusters"; they are 5,960 with size from 2 to 113 individuals. The original DHS clusters were combined into purposive territorial areas with a greater demographic size to keep the analysis at a territorial level. A spatially-aware k-means clustering method was employed to group the original clusters. Here, the mean point is represented by the barycenter of the clustered geographical coordinates. The procedure [9] was iterated on QGIS for all countries in the dataset, so that the number of macro-clusters would amount to about 50 per country, with a mean population of 270 individuals. The final sample is made of 549 macro-clusters. All variables are aggregated at the macro-cluster level to check for macro socio-economic drivers of HIV.

HIV prevalence at the macro-cluster level is the response variable of the model in this study. It ranges from 0-3.73% in Burundi to 0.19-35.43% in Mozambique, with a total average 5.41% (SD = 7.10).

All variables in the model – including the response – are either presented as means (for continuous variables) or as proportions (for dichotomous variables). Given the nature of the phenomenon, and the way the response variable was built, a fractional logistic regression model was implemented to evaluate the effect of the selected covariates [10]. The log-likelihood function for this model is:

$$\ln L = \sum_{j=1}^{N} y_j \ln\left(\frac{\exp\left(\mathbf{x}_j'\boldsymbol{\beta}\right)}{1+\exp\left(\mathbf{x}_j'\boldsymbol{\beta}\right)}\right) + (1-y_j)\ln\left(1-\frac{\exp\left(\mathbf{x}_j'\boldsymbol{\beta}\right)}{1+\exp\left(\mathbf{x}_j'\boldsymbol{\beta}\right)}\right) \quad (1)$$

in which N stands for the number of territorial units, y_j is the local HIV prevalence, and $\mathbf{x}_j$ is the set of regressors for each territorial unit j.

The option "clustered standard errors" was used so to account for macro-clusters being nested in countries.

Regressors were classified into three different domains: 1) *Socio-demographic characteristics of the macro-cluster:* average age; average no. of years of education; proportion of middle-to-high wealth people; proportion of workers; 2) *Sexual behaviours:* proportion of people who have never had sexual relationships; proportion of married people; proportion of people who used a condom during their last intercourse; 3) *Context variables:* proportion of respondents in a rural setting; country dummies. Country dummies also account for regional specificity.

3 Results

Table 1 shows odds ratio estimates for the model in equation (1). Among socio-demographic characteristics of the macro-clusters, the average number of children is a particularly interesting characteristic, with more children on average lowering the odds for respondents' to be HIV positive – probably due to women having access to prevention education during antenatal care. While the model does not show any significant association between HIV and age at the territorial level, it shows that more educated areas are less likely to be affected by HIV. However, as expected, wealthy areas are twice as likely to have a higher prevalence of HIV.

As for sexual behaviours, a higher proportion in the area of people who have never been in sexual relationships lowers their cluster's HIV prevalence (83% less likely), and a higher prevalence of HIV-positive people is associated with higher proportions of people using condoms at last intercourse. On the other hand, being married is not statistically significant at the territorial level in explaining HIV prevalence.

Rural areas are not statistically different from urban ones in terms of HIV prevalence, while differences among countries emerge with, other things being equal, Burundi, Senegal, and Sierra Leone showing lower levels of HIV prevalence concerning Angola (assumed as a reference country).

Table 1: Fractional logit model for HIV prevalence (odds ratio estimates)

Domains	*Covariates*	*OR estimates*
Socio-demographic characteristics	Age (avg.)	1.001
	No. of years of education (avg.)	0.884***
	No. of children (avg.)	0.786***
	Middle-to-high wealth (prop.)	1.872**
	Workers (prop.)	0.507
Sexual behav-iours	Never in a sexual relationship (prop.)	0.172***
	Married people (prop.)	0.216
	Used condom at last intercourse (prop.)	3.665**
Context	Rural area	0.898
	Constant	0.289*
	Observations	549

*** p < 0.01, ** p < 0.05, * p < 0.10 *Note: controlled for country dummies*

4 Conclusions

This study is focused on eleven African countries, which, despite only accounting for 3% of the global population, represented more than 35% of all global HIV cases. This analysis aimed to assess the relationship between socio-economic drivers and HIV prevalence at the sub-national level, focusing exclusively on sexual transmission and the socio-economic background of the territorial areas.

Literature on determinants of HIV prevalence has found discordant results. Using a fractional logit model on aggregated data, our findings showed that while age and employment rates do not seem to have an effect on a macro level, a highly educated population is less likely to be HIV positive, but richer people are more likely to drive forward this epidemic.

This mechanism may be due to social networking: richer people have higher chances of having more encounters, which in turn increases their odds of being HIV positive. Social relationships and conventions may also explain why higher proportions of people using condoms during their last intercourse are highly associated with people being HIV positive: people aware of their serostatus are more conscious and, thus, try to avoid infecting their partners by using correct preventative methods.

There is no doubt that these effects need to be further investigated. The erratic way the socio-economic status affects HIV is a sign that the relationship between these two dimensions depends not only on the territories where these data are collected but also on how each population changes over time.

This study is not free from limitations, among these: the arbitrariness of aggregation of territorial clusters; a poor structure of context-level variables. Further steps include robustness tests to different spatial aggregation strategies and a deep reflection on such methodological and substantive aspects.

References

1. AIDSinfo | UNAIDS. Aidsinfo.unaids.org. https://aidsinfo.unaids.org. Published 2022. Cited 27 Feb 2022.
2. World Population Prospects - Population Division - United Nations. Population.un.org. https://population.un.org/wpp/. Published 2022. Cited 27 Feb 2022.
3. Fortson J. The gradient in sub-Saharan Africa: Socioeconomic status and HIV/AIDS. Demography. (2008) doi:10.1353/dem.0.0006
4. Msisha W, Kapiga S, Earls F, Subramanian S. Socioeconomic status and HIV seroprevalence in Tanzania: a counterintuitive relationship. Int J Epidemiol. (2008) doi:10.1093/ije/dyn186
5. Igulot P, Magadi M. Socioeconomic Status and Vulnerability to HIV Infection in Uganda: Evidence from Multilevel Modelling of AIDS Indicator Survey Data. AIDS Res Treat. (2018) doi:10.1155/2018/7812146
6. Seeley J, Malamba S, Nunn A, Mulder D, Kengeya-Kayondo J, Barton T. Socioeconomic Status, Gender, and Risk of HIV-1 Infection in a Rural Community in South West Uganda. Med Anthropol Q. (1994) doi:10.1525/maq.1994.8.1.02a00060
7. Bunyasi E, Coetzee D. 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) doi:10.1136/bmjopen-2017-016232
8. Croft T, M.J. Marshall A, Allen C. Guide to DHS Statistics. Calverton, Maryland, USA: Macro International Inc.; (2018)
9. K-means clustering. docs.qgis.org. https://docs.qgis.org/3.16/en/docs/user_manual/processing_algs/qgis/vectoranalysis.html#k-means-clustering. Published 2022. Cited 28 Feb 2022.
10. Fractional regression. Stata.com. https://www.stata.com/manuals/rfracreg.pdf. Published 2022. Cited 28 Feb 2022.