Published in final edited form as:

J Biosoc Sci. 2013 November ; 45(6): 853–862.

An association between ethnic diversity and HIV prevalence in sub-Saharan Africa

Paul Henry Brodish, MSPH

MEASURE Evaluation, Carolina Population Center and Department of Public Policy, the University of North Carolina at Chapel Hill

Summary

This paper investigates whether ethnic diversity at the Demographic and Health Surveys (DHS) cluster level predicts HIV serostatus in three sub-Saharan African countries (Kenya, Malawi, and Zambia), using DHS household survey and HIV biomarker data for men and women ages 15–59 collected since 2006.. The analysis relates a binary dependent variable (HIV positive serostatus) and a weighted aggregate predictor variable representing the number of different ethnic groups within a DHS Statistical Enumeration Area (SEA) or cluster, which roughly corresponds to a neighborhood. Multilevel logistic regression is used to predict HIV prevalence within each SEA, controlling for known demographic, social, and behavioral and predictors of HIV serostatus. The key finding was that the cluster-level ethnic diversity measure was a significant predictor of HIV serostatus in Malawi and Zambia but not in Kenya. Additional results reflected the heterogeneity of the epidemics: male gender, marriage (Kenya), number of extramarital partners in the past year (Kenya and Malawi, but likely confounded with younger age), and Muslim religion (Zambia) were associated with lower odds of positive HIV serostatus. Condom use at last intercourse (a spurious result likely reflecting endogeneity), STD in the past year, number of lifetime sexual partners, age (Malawi and Zambia), education (Zambia), urban residence (Malawi and Zambia), and employment (Kenya and Malawi) were associated with higher odds of positive serostatus. Future studies might continue to employ multilevel models and incorporate additional, more robust controls for individual behavioral risk factors and for higher-level social and economic factors, in order to verify and further clarify the association between neighborhood ethnic diversity and HIV serostatus.

HIV prevalence rates in sub-Saharan Africa (SSA) are 50 times higher in some countries compared to the average outside the region. Prevention efforts have focused on individual-level behavioural interventions that try to influence knowledge, attitudes, and behaviours, and there is increasing evidence that infection rates in SSA are declining. For example, 22 of the most affected countries in SSA have reduced HIV incidence by more than 25% between 2001 and 2009 (Joint United Nations Programme on HIV/AIDS, 2010). However, it is unclear which of the range of behavioural interventions has been most effective nor to what extent, and even where, as in Uganda, national policy responses have been hailed a success, there has been recent backsliding attributed to a relative neglect of the broader sociocultural factors that constrain individual behavior. In Uganda prevalence increased to 7.3 percent in 2011 from 6.4 percent in 2005, despite the United States spending $1.7 billion there to fight AIDS over the same period through the President's Emergency Plan for AIDS Relief (PEPFAR) AIDS prevention strategy (Kron, 2012). Limited success in containing the SSA HIV epidemic has prompted renewed attention to the social and economic upstream

---

Correspondence: Paul Henry Brodish, MEASURE Evaluation, Carolina Population Center, the University of North Carolina at Chapel Hill, 206 W. Franklin Street, 2nd Floor, CB 8120, Chapel Hill, NC 27516, Phone: 919-966-3510, FAX: 919-966-2391, brodish@unc.edu.

contextual or structual factors, sometimes termed “the causes of the causes” of disease (Birn, 2009), which may facilitate viral transmission and undermine intervention effectiveness. These approaches are difficult because they involve deeply entrenched societal factors, such as gender, income distribution, and stigma/marginalization (Gupta et al., 2008).

A number of investigators have added insights into the ongoing controversy about the relative importance of socioeconomic or sociocultural factors and sexual concurrency (or other risk behaviours) in explaining the severity of the SSA epidemic. A combination of viral, host, transmission, and societal factors all contribute to the higher rates of infection in the region, but no single host factor can account for these high prevalence rates (Shandera, 2007). A recent country-level empirical study identified a number of social factors associated with HIV prevalence, finding little effect for poverty but large and significant effects for the predominant religious affiliation of the country (Nattrass, 2009). Within SSA countries, HIV prevalence rates are generally higher in urban compared to rural areas, but there is also much regional variation, with some poorer, rural areas like Nyanza Province in western Kenya having prevalence rates exceeding 20%. A study using the 2005–06 Zimbabwe DHS identified a strong and consistent contextual effect for ethnicity on sexual behavior among youth ages 15–24, controlling for several sociodemographic and social-cognitive factors (Sambisa, Curtis and Stokes, 2010), suggesting a need for prevention strategies that incorporate and address contextual factors that reproduce and perpetuate risky sexual behaviours.

This study utilized ecological systems theory applied to health (i.e., the social ecological model of health), which views individual health status as determined by a broad array of factors operating at multiple levels, often termed macro-, exo-, meso-, and micro-, which describe influences as intercultural, community, organizational, and interpersonal or individual. While this conceptual model, adopted by the World Health Organization’s Commission on the Social Determinants of Health (Commission on Social Determinants of Health, 2008) applies to general health status, it can be utilized to examine potential influences on specific disease susceptibility, and in the developing world the major threats to population health are infectious disease vulnerability and transmission.

One social health determinant is ethnic diversity or fractionalization, measured as “the likelihood that any two randomly drawn individuals from a population will be from different ethnic groups” (Lieberman, 2007) p.1410, and southern SSA has among the highest levels of ethnic diversity/fractionalization in the world, owing in part to the arbitrary drawing of regional boundaries across ethnic and tribal lines during colonial times (Alesina, Easterly and Matuszeski, 2011; Rosen, 2012). In the international development and political institutions literatures, ethnic diversity is associated with frequent inter-group conflicts, under-provision of public goods (Banerjee and Somanathan, 2007; Habyarimana *et al*., 2007), and generally poorer political and economic performance (Alesina, Easterly and Matuszeski, 2011). In a broad sample of developing countries Easterly found that greater social equality or a “middle class consensus” (i.e., a high share of income for the middle class and fewer ethnic divisions) is associated with a variety of positive outcomes, including higher income and growth; more education; better health, infrastructure, and economic policies; less political instability, civil war, and ethnic minority persecution; and more social “modernization” and democracy (Easterly, 2001). In a recent study of mostly low-income African countries, ethnic fractionalization was associated with lower public health spending (Platas, 2011).

Specifically with regard to HIV/AIDS, ethnic fractionalization in lower-income countries has been associated with lower per capita government AIDS expenditures, less anti-retroviral coverage, and a generally weaker policy response (Lieberman, 2007). The author

attributes these findings to the "important role of group esteem and the social construction of risk." That is, in the presence of ethnic fractionalization, groups are less likely to assess their risks as shared. Elites (or ordinary citizens) may downplay their own group's risk out of fear of reputational consequences and discount the indirect benefits of policies targeting "other" groups. Another study using the 2003 Zambian Sexual Behavior Survey found that the chances of men's involvement in extramarital sex increased with community-level (defined as the survey sampling cluster) ethnic heterogeneity (Benefo, 2008). And another study using data from the Nigeria 2008 DHS and a 2005 HIV seroprevalence sentinal survey found higher HIV prevalence in zones and states within Nigeria characterized by higher levels of ethnic-based violence, but this higher prevalence appeared to be mediated by HIV/AIDS-related knowledge and risk behaviours (Oluwadare and Dada, 2012).

To identify a distinct effect of ethnic diversity on HIV prevalence, a pooled analysis was conducted using DHS household survey data collected since 2006 from three southeastern SSA countries (Kenya, Malawi, and Zambia) with HIV prevalence rates exceeding five percent [rates above 1% are considered generalized epidemics (Joint United Nations Program on AIDS, 2011)] and HIV biomarker data and data on all covariates. Data are downloadable from the MEASURE DHS website at http://www.measuredhs.com/data/available-datasets.cfm. DHS surveys are nationally representative population-based surveys with large sample sizes (usually between 5,000 and 30,000 households). In all households, women age 15–49 are eligible to participate; in many surveys men age 15–54(59) from a sub-sample are also eligible to participate. There are three core questionnaires: A Household Questionnaire, a Women's Questionnaire, and a Male questionnaire. HIV biomarker data complements self-reported household survey information by providing an objective profile of a HIV status in the population. The sample is based on a stratified two-stage cluster design, the first stage being the SEA (or cluster), generally drawn from Census files, and the second stage, within each SEA, a sample of households is drawn from an updated list of households. The sample is generally representative at the national, residence (urban-rural), and regional (departments, states) levels. This paper evaluates a community-level (SEA) factor: ethnic diversity or pluralism. This key independent variable was defined as the number of different ethnic groups within a given SEA or cluster. It ranged from 1 to 18 with a mean (SE) of 4.6 (0.16) across all three countries. Because there were over 40 ethnic groups in Zambia, those with less than 10 observations were folded into the "Other" category. The binary dependent variable was HIV positive serostatus. The DHS provides anonymous, voluntary testing using blood spots collected on filter paper from a finger prick. An initial ELISA test is performed in the laboratory, with retesting of all positive tests and 5–10 percent of the negative tests with a second ELISA. For those with discordant results on the two ELISA tests, a new ELISA or a Western Blot is performed (Measure DHS, 2012).

The final empirical model regressed the dependent variable individual HIV serostatus on the key community-level independent variable ethnic diversity and included 12 individual-level control variables: number of sexual partners (other than husband/wife) in the past year (dummy-coded as 0, 1, 2, and 3 or more), lifetime number of sexual partners (coded as 1, 2, 3–5, 6–10, and >10), condom use at last intercourse, self-reported STD in the past year, poverty status, male gender, urban residence, age (in years), education (in years), employed (currently working, having worked in past year, or on leave in the past 7 days), married or living together, and religious affiliation (Catholic, Protestant/other Christian, Muslim, No/other religion). Multicollinearity was assessed by examining the correlation matrix for potential control variables included in the final model. Because the data were nested within clusters and within countries, the assumption of standard logistic regression that respondents are independent within each cluster and that there is equal variance among clusters did not hold, and a multilevel regression framework was needed to account for the hierarchical data structure. The model assumed random variation in intercepts across clusters (random

intercepts model) but constant slopes for beta coefficients. The two levels consisted of 23,345 respondents clustered within 1,561 SEAs across the three countries. The country-stratified analysis proceeded in three steps. First, a null or base model including only the dependent variable HIV prevalence established the degree of variance at the cluster level in order to validate use of a multilevel framework. Next, all the level-1 control variables were added to the model (Model 1) in order to assess the improvement in model fit and presence of significant effects for individual-level predictors of HIV serostatus. Finally, the key independent variable community-level ethnic diversity (Model 2) was added to test for significance of this predictor controlling for all other variables in the model.

Mean HIV prevalence was 15.3% in Zambia, 6.8% in Kenya, and 11.5% in Malawi and varied by ethnic group (Table 1). In Zambia nine ethnic groups had prevalence rates exceeding 20% and in Malawi four (Lomwe, Lambya, Nyanja, and Mang'anja) had rates exceeding 16%, while in Kenya a rate of 21.8% was confined to the Luo. In bivariate analyses, the number of ethnic groups in a cluster was significantly correlated with HIV prevalence in the cluster ($r = 0.37$, $p < 0.001$). The final multilevel weighted logistic regression models by country are shown in Table 2. The base model including only the dependent variable HIV prevalence rate showed significant variation by cluster. In Model 1 including the individual-level control variables showed a dose-dependent increase in the odds of being HIV positive with more lifetime sexual partners. Reporting 11 or more partners (reported by 6.1% of the sample or 1,428 individuals) increased the likelihood of being HIV positive five-fold in Kenya and Zambia and thirteen-fold in Malawi. Condom use at last intercourse and an STD in the past year increased this likelihood by three to five times in Kenya and Malawi and by 1.5 to 2.6 times in Zambia (Note: because this analysis is correlational, reverse causality likely explains the former association, i.e., condom use is likely to be more frequent among those who know they are HIV positive and/or who engage in higher-risk sex). Male gender was protective in all three countries and Muslim religious affiliation was protective in Zambia, while urban residence and age in Malawi and Zambia, and being employed in Kenya and Malawi, were associated with a slightly increased likelihood of being HIV positive.

In Model 2 adding the cluster-level ethnic diversity measure significantly improved the model fit in both Malawi and Zambia. The number of different ethnic groups per cluster in both Malawi (OR = 1.21, $p < 0.001$) and Zambia (OR = 1.05, $p < 0.01$) was associated with a significant increase in the likelihood of being HIV positive. The marginal effects of ethnic diversity were that a 1 point increase in the number of ethnic groups within a given SEA cluster was associated with a 1.21 times increased likelihood of being HIV positive in Malawi and with a 1.05 times increased likelihood of being HIV positive in Zambia, controlling for all other variables in the models. To my knowledge this report is the first to empirically demonstrate an association between neighborhood ethnic diversity and HIV prevalence using the DHS sampling cluster. Future studies might validate this finding and clarify its mechanism of action.

Acknowledgments

Support/Funding Disclosure

This work was supported in part by an R24 NIH center grant to the Carolina Population Center, the University of North Carolina at Chapel Hill

References

Alesina A, Easterly W, Matuszeski J. Artificial states. Journal of the European Economic Association. 2011; 9(2):246–277.10.1111/j.1542-4774.2010.01009.x

Banerjee A, Somanathan R. The political economy of public goods: Some evidence from India. Journal of Development Economics. 2007; 82(2):287–314.10.1016/j.jdeveco.2006.04.005

Benefo KD. Determinants of Zambian men's extra-marital sex: A multi-level analysis. Archives of Sexual Behavior. 2008; 37(4):517–529.10.1007/s10508-007-9243-7 [PubMed: 17999170]

Birn AE. Making it politic(al): Closing the gap in a generation: Health equity through action on the social determinants of health. Social Medicine. 2009; 4(3):166–182.

Commission on Social Determinants of Health. Final Report of the Commission on Social Determinants of Health. World Health Organization; Geneva, Switzerland: 2008. Closing the gap in a generation: health equity through action on the social determinants of health.

Easterly W. The Middle Class Consensus and Economic Development. Journal of Economic Growth. 2001; 6(4):317–335.10.1023/a:1012786330095

Gupta GR, Parkhurst JO, Ogden JA, Aggleton P, Mahal A. HIV prevention 4 - Structural approaches to HIV prevention. Lancet. 2008; 372(9640):764–775.10.1016/s0140-6736(08)60887-9 [PubMed: 18687460]

Habyarimana J, Humphreys M, Posner DN, Weinstein JM. Why does ethnic diversity undermine public goods provision? American Political Science Review. 2007; 101(4):709–725.

Joint United Nations Program on Aids. UNAIDS Terminology Guidelines. 2011.

Joint United Nations Programme on Hiv/Aids. MDG6: Six things you need to know about the AIDS response today. Geneva: 2010.

Kron, J. The New York Times. New York: 2012. In Uganda, an AIDS Success Story Comes Undone.

Lieberman ES. Ethnic politics, risk, and policy-making - A cross-national statistical analysis of government responses to HIV/AIDS. Comparative Political Studies. 2007; 40(12):1407–1432.10.1177/001041007306862

Measure Dhs. DHS Overview: HIV Prevalence. ICF International; Calverton, MD: 2012.

Nattrass N. Poverty, Sex and HIV. Aids and Behavior. 2009; 13(5):833–840.10.1007/s10461-009-9563-9 [PubMed: 19370407]

Oluwadare C, Dada A. The challenge of ethnic diversity and HIV prevalence in Nigeria. Journal of AIDS and HIV Research. 2012; 4(4):100–104.10.5897/JAHR11.029

Platas, M. Africa's Health Tragedy? Ethnic Diversity and Health Outcomes. Western Political Science Association Annual Meeting; Social Science Research Network (SSRN) San Antonio, TX. 2011.

Rosen, D. New York Times Sunday Book Review. 2012. Social Networks of Disease: 'Tinderbox,' by Craig Timberg and Daniel Halperin.

Sambisa W, Curtis SL, Stokes CS. Ethnic differences in sexual behaviour among unmarried adolescents and young adults in Zimbabwe. Journal of Biosocial Science. 2010; 42(1):1–26.10.1017/s0021932009990277 [PubMed: 19793404]

Shandera WX. Key determinants of AIDS impact in Southern sub-Saharan Africa. Ajar-African Journal of Aids Research. 2007; 6(3):271–286.

Table 1

Summary statistics for ethnic groups, by country and HIV sero-status[1]

Zambia					Kenya					Malawi				
	Total		HIV+			Total		HIV+			Total		HIV+	
	n	%	n	%		n	%	n	%		n	%	n	%
Gowa	13	0.2	4	32.1	Luo	661	13.5	144	21.8	Lomwe	1780	17.2	321	18.0
Soli	43	0.5	13	31.2	Luhya	752	15.4	54	7.2	Lambya	38	0.4	6	16.9
Kunda	78	1.0	20	25.1	Masai	46	0.9	3	7.0	Nyanja	161	1.5	27	16.6
Nsenga	451	5.6	109	24.2	Meru	251	5.2	16	6.3	Mang'anja	291	2.8	48	16.5
Lozi	468	5.8	103	22.1	Kamba	503	10.3	25	4.9	Yao	1378	13.3	190	13.8
Ila	32	0.4	7	21.9	Kikuyu	929	19.0	40	4.3	Ngoni	1296	12.5	160	12.4
Lungu	53	0.7	12	21.8	Mijikenda/Swahili	237	4.8	9	3.9	Sena	478	4.6	59	12.3
Chikunda	34	0.4	7	20.7	Kisii	365	7.5	14	3.8	Tonga	195	1.9	21	10.7
Lunda[2]	118	1.5	24	20.3	Other	176	3.6	6	3.5	Nkhonde	108	1.0	9	8.5
Lenje	123	1.5	24	19.4	Embu	89	1.8	3	3.4	Tumbuka	848	8.2	70	8.2
Swaka	32	0.4	6	18.9	Taita/ Taveta	62	1.3	2	2.6	Other	189	1.8	14	7.6
Lala	237	2.9	41	17.4	Kalenjin	690	14.1	14	2.1	Chewa	3551	34.3	265	7.5
Tabwa	37	0.5	6	17.3	Somali	119	2.4	1	0.9	Ndali	33	0.3	2	7.1
Luvale	158	2.0	27	17.3										
Chewa	603	7.5	102	16.9										
Nkoya	27	0.3	4	16.5										
Chishinga	33	0.4	5	16.2										
African	59	0.7	10	16.1										
Chokwe	40	0.5	6	16.0										
Lamba	158	2.0	24	15.5										
Toka-Leya	58	0.7	9	15.5										
Nyanja	16	0.2	2	15.3										
Ngoni	416	5.2	62	14.9										
Tonga	932	11.6	136	14.6										
Bemba	1679	20.8	242	14.4										
Kaonde	252	3.1	36	14.1										

	Zambia				Kenya				Malawi			
	Total		HIV+		Total		HIV+		Total		HIV+	
	n	%	n	%	n	%	n	%	n	%	n	%
Mambwe	236	2.9	33	13.9								
Other	181	2.3	24	13.4								
Mbunda	119	1.5	14	12.0								
Luchazi	48	0.6	6	11.5								
Senga	51	0.6	6	11.3								
Ushi	146	1.8	15	10.4								
Bisa	87	1.1	8	9.7								
Tumbuka	447	5.5	43	9.6								
Ngumbo	39	0.5	3	9.0								
Namwanga	238	3.0	21	8.9								
Mashi	41	0.5	3	6.3								
Lunda[3]	225	2.8	13	5.6								
Kabende	23	0.3	1	4.2								
Bwile	31	0.4	0	0.0								
Total	**8063**	**100**	**1233**	**15.3**	**4881**	**100**	**332**	**6.8**	**10346**	**100**	**1192**	**11.5**

[1] Numbers are estimated counts/percentages adjusted for survey sampling frame. Shaded groups have notably high prevalence rates compared to the national average.

[2] Luapula

[3] Northwestern

Table 2

Parameter estimates[a] (odds ratio) for multilevel weighted logistic regression models for cluster-level ethnic diversity

	Kenya (n = 4928)		Malawi (n = 10267)		Zambia (n = 8150)	
Explanatory variables	**Model 1[b]**	**Model 2[c]**	**Model 1[b]**	**Model 2[c]**	**Model 1[b]**	**Model 2[c]**
Fixed effects						
Community characteristic						
Number of ethnic groups		0.98		1.21***		1.05**
Individual characteristics						
Sexual behaviors						
# extramarital sex partners						
past year						
0	1.00	1.00	1.00	1.00	1.00	1.00
1	0.40*	0.40*	0.61*	0.61*	1.06	1.06
2	0.12**	0.12**	0.18*	0.70*	0.97	0.97
3+	0.24*	0.24*	0.30*	0.30*	0.87	0.86
# lifetime sex partners						
1 (RC)	1.00	1.00	1.00	1.00	1.00	1.00
2	1.67	1.67	3.05***	3.01***	2.52***	2.51***
3 – 5	4.25***	4.25***	6.80***	6.69***	4.26***	4.23***
6 – 10	4.88***	4.88***	9.65***	9.51***	4.21***	4.19***
11+	5.01***	5.01***	13.76***	13.36***	5.17***	5.09***
Condom use last intercourse						
No	1.00	1.00	1.00	1.00	1.00	1.00
Yes	3.75***	3.76	3.04***	3.02***	1.55***	1.54***
Any STD past year						
No	1.00	1.00	1.00	1.00	1.00	1.00
Yes	3.12**	3.12**	4.89***	4.96***	2.66***	2.63***
Demographics						
Poverty status						

	Kenya (n = 4928)		Malawi (n = 10267)		Zambia (n = 8150)	
Explanatory variables	**Model 1[b]**	**Model 2[c]**	**Model 1[b]**	**Model 2[c]**	**Model 1[b]**	**Model 2[c]**
Not in poverty	1.00	1.00	1.00	1.00	1.00	1.00
In poverty	0.66	0.66	0.84	0.88	0.70**	0.72*
Gender						
Female	1.00	1.00	1.00	1.00	1.00	
Male	0.24***	0.24***	0.24***	0.24***	0.39***	0.39***
Residence						
Rural	1.00	1.00	1.00	1.00	1.00	1.00
Urban	1.33	1.37	2.26***	1.80***	2.31***	1.83***
Age (years)	1.02	1.02	1.05***	1.05***	1.03***	1.03***
Education (years)	0.99	0.99	1.03	1.02	1.03*	1.03*
Working						
No	1.00	1.00	1.00	1.00	1.00	1.00
Yes	1.61*	1.61*	1.55**	1.55**	1.05	1.05
Married						
No	1.00	1.00	1.00	1.00	1.00	1.00
Yes	0.47*	0.47*	0.94	0.95	1.17	1.17
Religion						
Roman Catholic	1.00	1.00	1.00	1.00	1.00	1.00
Protestant/other Christian	1.02	1.03	1.12	1.11	1.17	1.16
Muslim	0.73	0.73	1.03	1.03	0.21*	0.20*
Other	1.44	1.43	0.90	0.94	0.84	0.83
Random effects						
Cluster-level variance (SE) [d]	1.12 (0.20)***	1.12 (0.20)***	0.92 (0.12)***	0.92 (0.12)***	0.39 (0.07)***	0.38 (0.07)***
Log likelihood	−1042.71	−1042.68	−2980.37	−2967.03	−3044.01	−3039.33
AIC	2127.42	2129.36	6002.74	5978.07	6130.01	6122.66
Log likelihood Ratio Test (χ^2)[e]		0.06		26.67***		9.35**

[a] Parameters for predictors (fixed effects) are reported as odds ratio; for random effects, the parameter is the variance

[b] Baseline model with control variables alone

J Biosoc Sci. Author manuscript; available in PMC 2014 November 01.

[c] Full model with control variables and predictor variables

[d] Significance of random effects evaluated by comparing model with a similar one in which random effects have been constrained to be zero.

[e] Compared to Model 1

* p<0.05

** p<0.01

*** p<0.001