

HIV Prevalence and Associated Factors among Adults Aged 15-64 in KAIS 2012

Regenerated from KAIS 2012 manuscript inputs and saved Xalec AI analysis artifacts. Analyst review required before submission.

Abstract

Background: Kenya has a generalized HIV epidemic with important variation by geography, sex, age, and social position. National population-based surveys are central to monitoring the epidemic and guiding prevention, testing, and treatment programmes. **Objective:** To estimate HIV prevalence and describe demographic, behavioral, and socioeconomic factors associated with HIV positivity among adults aged 15-64 years using KAIS 2012 analysis outputs. **Methods:** KAIS 2012 was a cross-sectional, population-based household survey using a two-stage stratified cluster design. The selected analysis artifacts included survey-weighted prevalence tables by sex and selected factors, descriptive frequency tables, and a survey-aware simple logistic regression table. Estimates used the saved survey-analysis outputs, including weighted percentages, 95% confidence intervals, and odds ratios where provided. **Results:** In the selected prevalence output, 648 of 11,534 adults with analyzable HIV status were HIV-positive, corresponding to weighted HIV prevalence of 5.68% (95% CI: 4.98-6.37). Weighted prevalence was higher among women than men (6.92% vs 4.39%). Regional patterns were pronounced, with Nyanza showing the highest weighted prevalence among listed regions (15.21%; 95% CI: 11.49-18.93). In simple survey logistic regression, women had higher odds of HIV positivity than men (OR 1.62; 95% CI: 1.34-1.96; $p < .001$), and Nyanza had higher odds than Eastern region (OR 4.59; 95% CI: 2.90-7.28; $p < .001$). Several older age groups also had higher unadjusted odds relative to ages 15-19 years. **Conclusion:** The reviewed artifacts support a manuscript focused on weighted HIV prevalence and unadjusted associations in KAIS 2012. A full publication on associated factors should add and review the planned multivariable survey logistic regression before making adjusted-effect claims.

Introduction

Kenya has experienced a generalized HIV epidemic, with substantial differences in HIV burden by region, sex, age group, residence, and socioeconomic conditions. HIV prevalence peaked in the late 1990s and has declined alongside expanded prevention, HIV testing, prevention of mother-to-child transmission, voluntary medical male circumcision, antiretroviral therapy, and broader public-health investments. Despite this progress, HIV remains a major public-health challenge requiring nationally representative surveillance to monitor epidemic patterns, evaluate programme reach, and guide targeted response.

The Kenya AIDS Indicator Survey 2012 (KAIS 2012) was the second national AIDS Indicator Survey, following KAIS 2007. KAIS 2012 collected household, individual, and laboratory data and linked HIV serostatus with demographic, behavioral, clinical, and household characteristics. The survey provides an important basis for estimating HIV prevalence and examining factors associated with HIV infection among adults aged 15-64 years.

The objective of this manuscript draft is to estimate HIV prevalence and describe demographic, behavioral, and socioeconomic factors associated with HIV positivity among adults aged 15-64 years in Kenya using reviewed KAIS 2012 analysis artifacts. Because the selected regression artifact currently contains simple logistic regression only, this draft distinguishes unadjusted associations from adjusted associations and does not present causal interpretations.

Methods

Study design and setting: KAIS 2012 was a cross-sectional, population-based household survey. It was nationally representative for Kenya across 9 of 10 NASCOP programmatic regions; North Eastern was excluded because of security concerns. Sampling used a two-stage stratified cluster design based on the NASSEP V frame. In the first stage, clusters were selected, and in the second stage, 25 households were selected within each cluster. Data collection used electronic capture during household visits, with interviews and blood specimen collection conducted in participants homes.

Population: The target population was adults and adolescents aged 15-64 years who were usual residents of selected households or visitors present the night before the survey. Individuals with cognitive or hearing disabilities that prohibited participation were excluded. The provided study brief reported 16,383 eligible adults aged 15-64 years, 13,720 completed interviews, and 11,626 provided blood specimens. The current

analytic prevalence table contained 11,534 adults with analyzable HIV outcome data for the displayed prevalence estimates.

Data sources and variables: Data sources included the household questionnaire, individual male and female questionnaires, and laboratory testing for HIV and related markers. The primary outcome was laboratory-confirmed HIV serostatus. The reviewed outputs included residence, region, sex, age category, education, marital status, sexual debut, and age at last birthday. Additional variables described in the study brief included employment, wealth, sexual history, number of partners, condom use, circumcision status, HIV testing history, TB/STI history, PMTCT, ART use, CD4 count, viral load, and household characteristics; not all of these were present in the selected analysis artifacts.

Statistical analysis: The saved outputs used survey-aware estimation with weights and survey-design information as configured in the completed jobs. Weighted proportions, prevalence estimates, 95% confidence intervals, and p-values were reported in the frequency/prevalence tables. The regression artifact supplied simple survey logistic regression odds ratios with 95% confidence intervals and p-values. Because the selected logistic regression output does not include adjusted model columns, this regenerated manuscript reports unadjusted associations only and treats multivariable analysis as a required next step.

Ethics, funding, and data availability: KAIS consent procedures and funding should be verified against the final KAIS 2012 documentation before journal submission. The KAIS 2012 final report states that the survey was supported by PEPFAR through CDC and USAID, the Joint United Nations Team on HIV/AIDS, JICA, Global Fund, and the World Bank. Data availability and ethics approval identifiers require author confirmation.

Results

Analytic sample and HIV prevalence: In the selected prevalence output, 648 of 11,534 adults were HIV-positive, giving a weighted HIV prevalence of 5.68% (95% CI: 4.98-6.37). Among women, weighted prevalence was 6.92% (95% CI: 6.06-7.77), compared with 4.39% (95% CI: 3.58-5.19) among men. Weighted prevalence was 5.14% (95% CI: 4.29-5.99) in rural areas and 6.58% (95% CI: 5.40-7.75) in urban areas in the selected prevalence table.

Regional and age patterns: Regional prevalence varied across the displayed regions. Nyanza had the highest weighted prevalence among listed regions at 15.21% (95% CI: 11.49-18.93), while Eastern, Central, Coast, Rift Valley, Nairobi, and Western had lower prevalence estimates in the selected table. Age-specific prevalence rose from 1.05% (95% CI: 0.47-1.64) among those aged 15-19 years to 9.82% (95% CI: 7.14-12.49) among those aged 45-49 years before declining in older age groups.

Other participant characteristics: Weighted prevalence was 3.59% (95% CI: 2.40-4.79) among adults with no primary education, 5.99% (95% CI: 4.97-7.01) among those with complete primary education, and 5.86% (95% CI: 4.98-6.73) among those with secondary or higher education. By marital status, weighted prevalence was 2.28% (95% CI: 1.74-2.82) among adults recorded as ever widowed, 5.32% (95% CI: 4.46-6.18) among those married or cohabiting, 11.07% (95% CI: 8.31-13.83) among those never married or never cohabited, and 19.96% (95% CI: 16.26-23.65) among those separated or divorced. Sexual debut before age 15 years had weighted prevalence of 6.55% (95% CI: 4.91-8.20), compared with 6.06% (95% CI: 5.28-6.85) for first sex at age 15 years or older.

Simple logistic regression: In unadjusted survey logistic regression, urban residence had OR 1.30 (95% CI: 0.96-1.76; $p=0.089$) compared with rural residence. Nyanza had OR 4.59 (95% CI: 2.90-7.28; $p<.001$) compared with Eastern region. Women had OR 1.62 (95% CI: 1.34-1.96; $p<.001$) compared with men. Compared with adults aged 15-19 years, unadjusted odds were elevated in several older age groups, including ages 35-39 years (OR 8.99; 95% CI: 5.02-16.10; $p<.001$), 40-44 years (OR 9.65; 95% CI: 5.13-18.15; $p<.001$), and 45-49 years (OR 10.22; 95% CI: 5.73-18.24; $p<.001$). Complete primary education (OR 1.71; 95% CI: 1.17-2.49; $p=0.005$) and secondary or higher education (OR 1.67; 95% CI: 1.15-2.42; $p=0.007$) were associated with higher unadjusted odds compared with no primary education. First sex before age 15 years was not statistically associated with HIV positivity in the selected simple logistic output (OR 1.09; 95% CI: 0.85-1.40; $p=0.517$).

Discussion

This regenerated draft supports a clear descriptive and unadjusted association narrative for HIV among adults aged 15-64 years in KAIS 2012. The selected artifacts show overall weighted HIV prevalence of 5.68% and higher prevalence among women than men. Regional differences were also prominent, particularly the high prevalence

estimate for Nyanza, which is consistent with the need for geographically targeted HIV prevention, testing, and treatment strategies.

The simple logistic regression output reinforces several patterns observed in the prevalence table. Women, participants in Nyanza, and several middle-age groups had higher unadjusted odds of HIV positivity. These associations should be interpreted as descriptive epidemiologic patterns rather than causal effects. The selected job artifact does not include adjusted model output, so confounding by age, sex, region, residence, education, marital status, and behavioral factors cannot be resolved in this draft.

Public health implications: The results can support programme planning by identifying subgroups for closer prevention, testing, linkage-to-care, and treatment-continuity review. They also underscore the value of maintaining nationally representative surveys with laboratory-confirmed HIV outcomes. For policy use, however, adjusted analyses should be reviewed before ranking risk factors or designing interventions around specific predictors.

Limitations: This analysis is cross-sectional and cannot establish temporal or causal relationships. The regenerated manuscript is based on selected saved artifacts rather than the complete KAIS data set. Some variables listed in the study brief were not included in the selected tables. The selected logistic regression artifact reports simple logistic regression only; therefore, adjusted associations are not available in this report package. Survey variance calculations and p-values should be checked against the final analysis specification, especially if journal submission requires exact SAS-parity or design-based methods.

Future strengthening: A publication-quality version should add the planned multivariable survey logistic regression, document reference categories, include a complete Table 1, include a final model table with adjusted odds ratios, and reconcile all analytic denominators with the study flow diagram.

Conclusion

Among adults aged 15-64 years in the selected KAIS 2012 analysis artifacts, weighted HIV prevalence was 5.68%, with higher prevalence among women and pronounced regional variation. Simple survey logistic regression identified higher unadjusted odds among women, participants from Nyanza, and several older age groups. These findings support targeted public-health interpretation, but a full associated-factors manuscript

requires a reviewed multivariable survey logistic regression model before adjusted conclusions are drawn.

References

- National AIDS and STI Control Programme (NASCOP), Kenya. (2014). Kenya AIDS Indicator Survey 2012: Final Report. Nairobi, Kenya: NASCOP.
<https://stacks.cdc.gov/view/cdc/26189>
- Waruiru W, Kim AA, Kimanga DO, Ng'ang'a J, Schwarcz S, Kimondo L, Ng'ang'a A, Umuro M, Mwangi M, Ojwang JK, Maina WK, and KAIS Study Group. (2014). The Kenya AIDS Indicator Survey 2012: rationale, methods, description of participants, and response rates. *Journal of Acquired Immune Deficiency Syndromes*, 66(Suppl 1), S3-S12. <https://doi.org/10.1097/QAI.0000000000000114>
- von Elm E, Altman DG, Egger M, Pocock SJ, Gøtzsche PC, Vandenbroucke JP. (2007). The Strengthening the Reporting of Observational Studies in Epidemiology (STROBE) Statement: guidelines for reporting observational studies. *PLoS Medicine*, 4(10), e296. <https://doi.org/10.1371/journal.pmed.0040296>
- Lumley T. (2010). *Complex Surveys: A Guide to Analysis Using R*. Wiley.
- Hosmer DW, Lemeshow S, Sturdivant RX. (2013). *Applied Logistic Regression*. Wiley.
- Cohen MS, Chen YQ, McCauley M, Gamble T, Hosseinipour MC, Kumarasamy N, et al. (2011). Prevention of HIV-1 infection with early antiretroviral therapy. *New England Journal of Medicine*, 365, 493-505.
<https://doi.org/10.1056/NEJMoa1105243>

Section-by-section notes on assumptions and missing information

Title: Uses the user-provided title and KAIS 2012 context. Confirm whether the target journal prefers adults and adolescents instead of adults for ages 15-17 years.

Abstract: Reports only estimates available in saved artifacts. The abstract labels logistic regression as simple/unadjusted because adjusted columns were not present in the selected logistic artifact.

Introduction: Uses user-provided background plus verified KAIS references. Add 5-10 Kenya-specific HIV epidemiology references before journal submission.

Methods: Design, sampling, setting, and eligibility are drawn from the user brief and KAIS documentation. Ethics approval identifiers, consent wording, and data access language need author confirmation.

Results: Estimates are taken from the selected frequency/prevalence and simple logistic regression artifacts. Results do not include adjusted odds ratios because the saved artifact does not contain them.

Discussion: Interpretation is limited to descriptive and unadjusted association patterns. Avoid causal wording unless a prespecified analytic design supports causal inference.

Suggestions to strengthen the manuscript

Run and attach the final multivariable survey logistic regression model with selected covariates, reference categories, adjusted odds ratios, 95% confidence intervals, and global p-values.

Add a participant flow diagram reconciling eligible adults, interviewed adults, blood specimens, and analytic denominators in each table.

Complete a full literature review with Kenya-specific and regional HIV prevalence/risk-factor studies, including KAIS 2007 and KENPHIA where appropriate.

Add a complete methods subsection for weighting, strata, clusters, non-response adjustment, post-stratification, and software/version details.

Decide whether the manuscript is primarily a prevalence paper, an associated-factors paper, or a methods demonstration; the current artifact set supports prevalence plus unadjusted associations most strongly.

Journal-readiness checklist

- Final adjusted model reviewed and included.
- All denominators reconciled across text, tables, and figures.
- Survey design variables, weights, strata, clusters, and variance method stated exactly.
- Ethics approval, consent, funding, conflicts of interest, and data availability finalized.
- STROBE checklist completed for a cross-sectional study.
- References expanded, verified, and formatted to the target journal style.
- Tables and figures exported in journal-ready format with captions and footnotes.

- Coauthor/statistician review completed before submission.

Follow-up questions for the author

Should the paper present only prevalence and unadjusted associations, or should the final multivariable model be required before manuscript use?

Which adjusted covariates should be included in the final model, and which reference categories should be used?

What exact ethics approval identifiers and consent language should appear?

What is the target journal and required reference style?

Should KENPHIA or KAIS 2007 comparisons be added to the Introduction and Discussion?

Table A1. Selected weighted HIV prevalence estimates from saved artifacts

Variable	Category	Total Unweighted n/N	Total Weighted Prev % (or Median)	Total Weighted 95% CI (or IQR)	P-value
Rural/urban setting	Total	648/11534	5.68	4.98-6.37	
Rural/urban setting	Rural	373/7427	5.14	4.29-5.99	0.394
	Urban	275/4107	6.58	5.40-7.75	
Region	Nairobi	67/1310	4.89	3.73-6.06	0.062
	Central	60/1413	3.83	2.74-4.92	
	Coast	66/1452	4.36	3.06-5.67	
	Eastern	75/2297	3.76	2.43-5.08	
	Nyanza	242/1621	15.21	11.49-18.93	
	RiftValley	79/2049	3.77	2.78-4.77	
	Western	59/1392	4.79	3.01-6.57	
Age category (5 year categories)	15-19	17/1841	1.05	0.47-1.64	<.001
	20-24	58/1896	3.11	2.24-3.98	
	25-29	112/1786	6.27	4.83-7.71	
	30-34	97/1418	6.62	5.24-8.00	
	35-39	108/1220	8.73	6.74-10.73	

Variable	Category	Total Unweighted n/N	Total Weighted Prev % (or Median)	Total Weighted 95% CI (or IQR)	P-value
	40-44	90/984	9.32	7.09-11.54	
	45-49	66/715	9.82	7.14-12.49	
	50-54	60/734	8.41	6.34-10.48	
	55-59	24/523	4.39	2.39-6.40	
	60-64	16/417	4.01	1.96-6.05	
Educational level	No primary	44/1332	3.59	2.40-4.79	0.038
	Incomplete primary	46/974	5.01	3.26-6.76	
	Complete primary	223/3667	5.99	4.97-7.01	
	Secondary+	335/5561	5.86	4.98-6.73	
Marital status	Ever widowed	78/3395	2.28	1.74-2.82	<.001
	Married/cohabiting	342/6683	5.32	4.46-6.18	
	Never married/never cohabited	63/558	11.07	8.31-13.83	
	Separated/divorced	165/894	19.96	16.26-23.65	
	Unknown	0/4	0.00	0.00-0.00	

Table A2. Selected simple survey logistic regression estimates from saved artifacts

Characteristic	N	n (%)	Simple OR (95% CI)	Simple p-value	Simple global p-value
Rural/urban setting					0.089
Rural	7427	373.0 (5.1)	ref		
Urban	4107	275.0 (6.6)	1.30 (0.96-1.76)	0.089	
Region					<.001
Eastern	2297	75.0 (3.8)	ref		
Nyanza	1621	242.0 (15.2)	4.59 (2.90-7.28)	<.001	
Respondent sex					<.001
Men	4794	193.0 (4.4)	ref		
Women	6740	455.0 (6.9)	1.62 (1.34-1.96)	<.001	
Age category (5 year categories)					<.001
15-19	1841	17.0 (1.1)	ref		

Characteristic	N	n (%)	Simple OR (95% CI)	Simple p-value	Simple global p-value
35-39	1220	108.0 (8.7)	8.99 (5.02-16.10)	<.001	
40-44	984	90.0 (9.3)	9.65 (5.13-18.15)	<.001	
45-49	715	66.0 (9.8)	10.22 (5.73-18.24)	<.001	
Educational level					0.037
No primary	1332	44.0 (3.6)	ref		
Complete primary	3667	223.0 (6.0)	1.71 (1.17-2.49)	0.005	
Secondary+	5561	335.0 (5.9)	1.67 (1.15-2.42)	0.007	
Sexual debut					0.517
First sex <15 years	1342	94.0 (6.6)	1.09 (0.85-1.40)	0.517	
Age At Last Brithday	11534	648 (5.7)	1.03 (1.02-1.03)	<.001	<.001

Figures

Figure A1. Weighted HIV prevalence bars from the saved prevalence table.

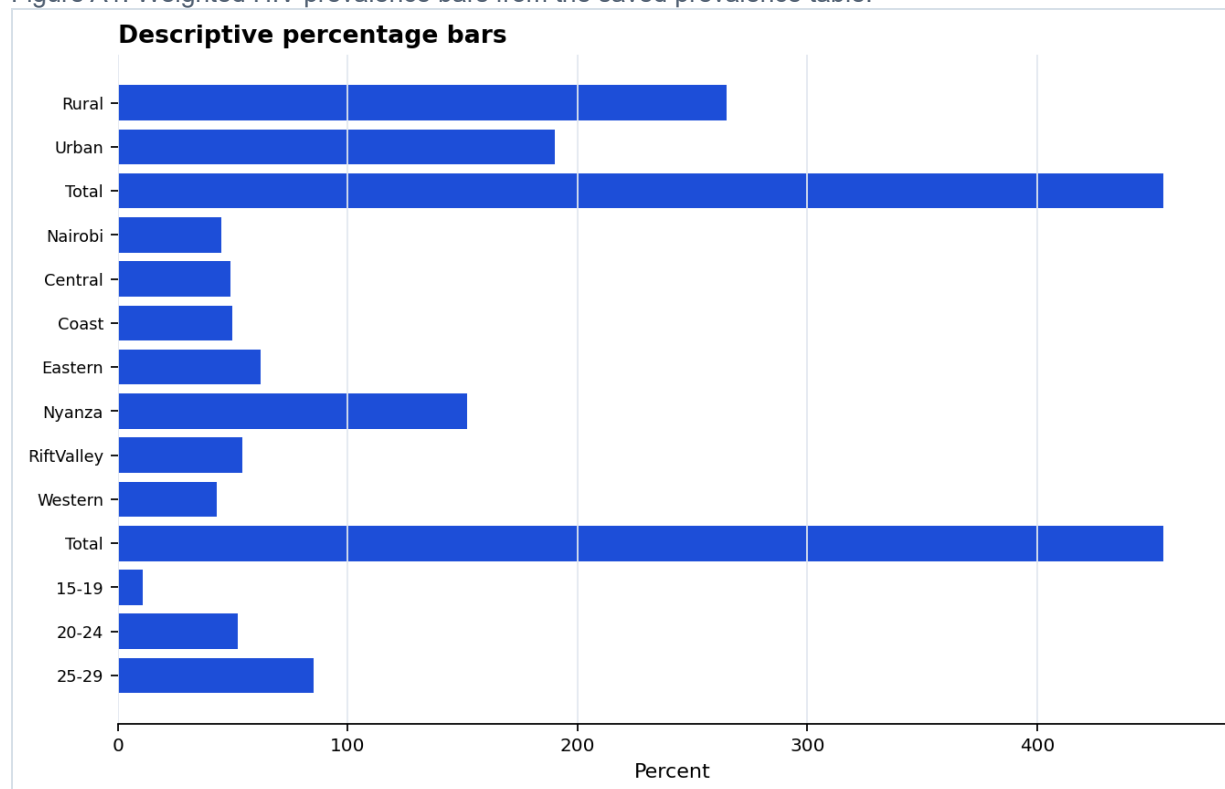


Figure A2. Descriptive distribution bars from the saved frequency table.

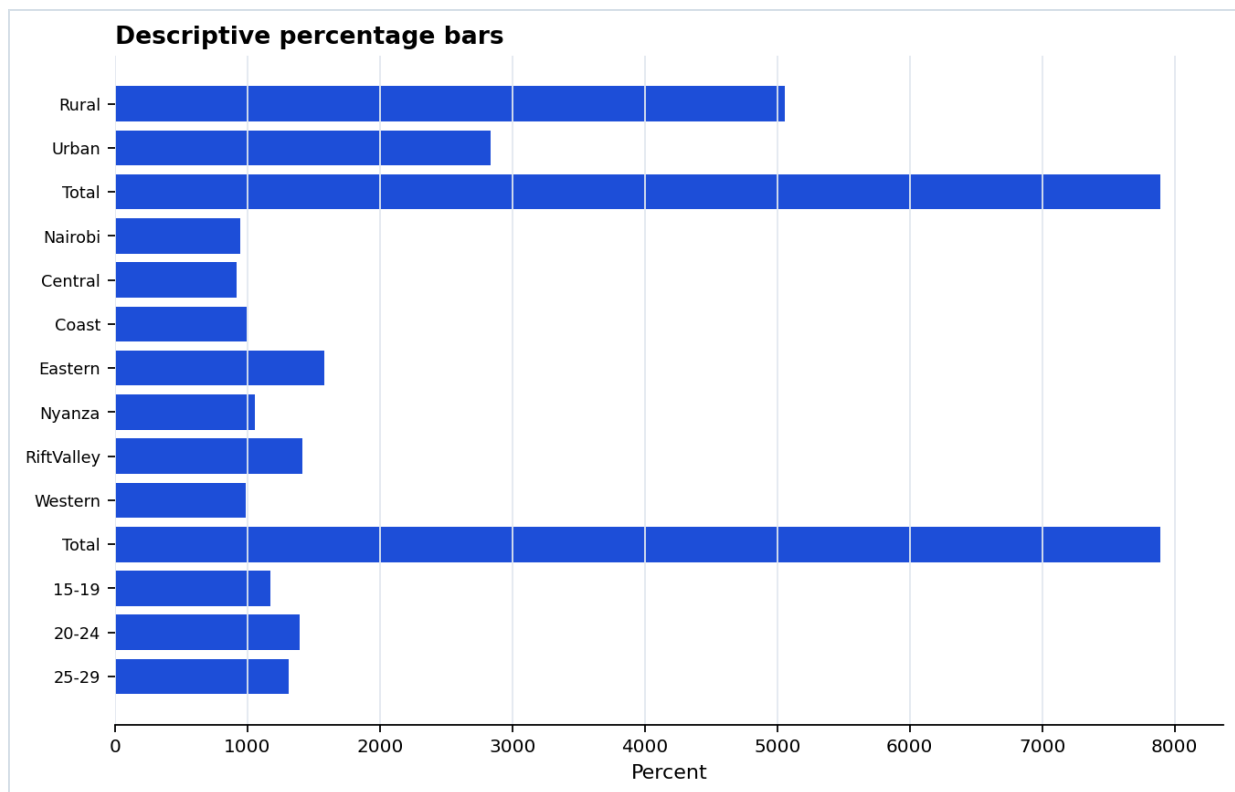


Figure A3. Odds-ratio forest plot from the saved simple survey logistic regression table.

