

The Demographic, Social, and Economic Correlates of HIV Infection Status in Sub-Saharan Africa

Chirag J. Patel¹, Kajal Claypool^{1,2}, Eric Chow³, Jake Chung¹, Don Mai⁴, Jessie Chen⁵, Eran Bendavid^{6,7}

1. Department of Biomedical Informatics, Harvard Medical School
2. Lincoln Laboratory, MIT
3. Rosalind Franklin University School of Medicine
4. Alphabet Inc.
5. Loyola University, Chicago
6. Center for Population Health Sciences, Stanford University
7. Center for Health Policy, Department of Medicine, Stanford University

Keywords: X-wide association study, HIV, Sub-Saharan Africa, social determinants of health

Running title: Systematic correlates of HIV in Sub-Saharan Africa

Key points: In a continent-wide search correlating 8,980 variables with HIV+ in >600,000 individuals from 49 countries across Sub-Saharan Africa, we identified hundreds of social determinants, such as education and marital status, whose associations are comparable to established causal factors.

Abstract

Background. Predisposition to HIV+ is influenced by a wide range of correlated economic, environmental, demographic, social, and behavioral factors. While evidence among a candidate handful have strong evidence, there is lack of a consensus among the vast array of variables measured in large surveys.

Methods. We performed a comprehensive data-driven search for correlates of HIV positivity in >600,000 participants of the Demographic and Health Survey (DHS) across 29 sub-Saharan African countries from 2003 to 2017. We associated a total of 7,251 and of 6,288 unique variables with HIV+ in females and males respectively in each of the 50 surveys. We performed a meta-analysis within countries to attain 29 country-specific associations.

Results. We identified 344 (5.4% out possible) and 373 (5.1%) associations with HIV+ in males and females, respectively, with robust statistical support. The identified associations are consistent in directionality across countries and sexes. The association sizes among individual correlates and their predictive capability was low to modest, but comparable to established factors. Among the identified associations, variables identifying being head of household among females was identified in 17 countries with a mean odds ratio (OR) of 2.5 (OR range: 1.1-3.5, $R^2 = 0.01$). Other common associations were identified with marital status, education, age, and ownership of land or livestock.

Conclusions. Our continent-wide search for variables has identified under-recognized variables associated with HIV+ that are consistent across the continent and sex. Many of the association sizes are as high as established risk factors for HIV+, including male circumcision.

Introduction

The number of new HIV infections in Sub-Saharan Africa has been declining for over a decade, but the prevalence of HIV has not declined, over 1 million people are newly infected every year, and over 5 million are unaware of being HIV-positive.[1,2] The global strategy for improving the population burden of HIV calls for intensive identification of HIV-infected individuals in order to pursue high treatment coverage, which, in turn, is anticipated to reduce HIV transmission and incidence.[1] Sub-Saharan Africa -- where the burden of HIV is highest -- remains short of these major targets, including identifying at least 90% of those with HIV.[1] One challenge hindering progress at identifying those with HIV infection is the heterogeneity of HIV risk in terms of modes of transmission and prevalence. In other words, it is hard to know where and whom to target for testing and public health interventions when aiming for very high (>90%) identification rates. This heterogeneity is manifest in the proliferation of inquiries into HIV risk factors, without unifying approaches or systematic studies into HIV risk.[3–7] At the time of this writing, no systematic reviews of HIV risk factors in Africa have been published. In this paper we describe a systematic approach to the identification of HIV risk and across multiple populations and geographic regions of Africa. The goal of this analysis is to help with identifying groups that may benefit from HIV-specific public health efforts.

The epidemiology of risk factor identification commonly proceeds by identifying candidate risk factors. These approaches are grounded in real-world observations but suffer from several limitations: generalizing to populations other than those directly examined is often tenuous; testing of one or several candidate risk factors may spuriously identify candidate risk factors;[8,9] and limited number of candidate risk factors may leave important relationships unobserved.[10]

We address these shortcomings using a large-scale risk factor testing approach across the entire population of individuals that had an HIV test as part of the Demographic and Health Surveys (DHS)

between 2003 and 2017 in Sub-Saharan Africa, a population of over 700,000 individuals extending our previous work to identify factors associated with HIV in Zambian females.[11] We test every variable or candidate risk factor available in DHS against individual-level HIV outcomes, a total of 32,353 variables (27,506 in females and 24,713 in males). This approach allows us to systematically identify variables that have a high degree of generalizability as risk factors for HIV, that play an important role in explaining who does and does not have HIV, over multiple survey waves across region of Sub-Saharan Africa in men and women separately.

Methods

Data Sources

The primary data for this work comes from the Demographic and Health Surveys (DHS). The DHS are nationally representative surveys conducted approximately every 5 years in many low- and middle-income countries. We used data from every individual who had consented and tested positive or negative for HIV-1.[12] There were a total of 50 surveys available for analysis for females and males each across 29 countries.

Selection and preparation of variables per survey

We retained all variables that had at least 90% complete data, a total of 29,092 and 25,980 unique variables in female surveys and male surveys, respectively, and 33,729 unique variables overall in both males and females (see Supplementary Information). All variables with 30 or fewer levels were treated as binary variables. Variables with 30 or more levels were treated as continuous. We retained 7,251 and 6,288 unique variables in females and males, respectively and 8,980 unique variables overall. The number of unique variables per country ranged from 337 (Congo) to 1659 (Malawi) for females and from 348 (Congo) to 1219 (Malawi) for males. See Supplementary Information for the number of surveys available per country.

Survey-Specific Association Models and Meta-Analysis Across Sub-Saharan Africa

We associated each of the variables with HIV status using a weighted logistic regression model:

$$\text{logit}(HIV_{sj}) = \alpha + \beta_s^i X_{sj}^i$$

Where s indexes the survey (including country, year, and female or male survey; e.g., Zimbabwe, 2005, female survey), j indicates the individual observation (all analyses are person-level analyses), and i indexes the variables for that survey. We estimated the Nagelkerke pseudo- R^2 to assess the improved goodness of fit from a logistic model with zero variables (equivalent to the prevalence of HIV) to a model with X . [13] We used survey-weighted logistic regression model to account for the probability-based sampling of DHS, implemented in the *survey* package in R. [14] Last, for each variable, we combined associations across all of the surveys (year and country combination) for males and females with a random effects meta-analysis procedure, estimating the average association and heterogeneity of the associations across the surveys and countries (See Supplementary Information).

We prioritized variables that were the most explanatory and whose associations were statistically non-zero across the surveys after correction for multiple hypotheses. Specifically, we report variables across pan-Saharan-Africa meta-analytic p-value lower than a conservative DHS-wide Bonferroni threshold of 1×10^{-6} (for 7,251 plus 6,288 variables, a Bonferroni threshold would be $0.05/13,539$, or $p < 3.7 \times 10^{-6}$) and whose average R^2 across the surveys were the top 25% of all R^2 for males and females, which was equivalent to a R^2 of 0.001.

For full transparency, a browsable version of all findings are available on a web application here: https://www.chiragipggroup.org/dhs_hiv_meta/. We have placed all of our summary statistics, including the overall and country-specific meta-analytic odds ratios, R^2 , I^2 , standard errors, and p-values in tables located here:

https://github.com/chiragjp/dhs_hiv_meta/blob/master/meta_data/Online%20Supplementary%20Tables.xlsx?raw=true. See Supplementary Information for more information.

Results

Cohort Characteristics

We harmonized 50 DHS surveys from 29 African countries conducted between 2003 and 2017, in which 619,468 participants were tested for HIV (47% male, 6.1% positive) (Table 1). While we only analyzed surveys with HIV prevalence at or above 1% of the adult population, the variability in the prevalence levels ranged from 1% to 25% (Table 1). As described in the methods, there were 7,251 and 6,288 unique variables in females and males respectively (Supplementary Figure 1AB).

Distribution of associations across Sub-Saharan Africa points to consistency in direction but variability in size of associations.

Overall, we found robust but heterogeneous associations between social, environmental, behavioral, and biological variables with HIV+ across the 29 countries of sub-Saharan Africa (Figure 1A-C, Supplementary Figure S2 [for single country]). In the following, we call “identified variables” as those that had a Bonferroni-level p-value less than 1×10^{-6} and a Nagelkerke R^2 of greater than 0.001. Reach of the estimated overall and country-specific odds ratios, standard errors, and p-values are documented in extensive online content (See Supplementary Information).

Many of the identified variables spanned social, behavioral, biological, economic, and environmental domains and exhibited consistent direction of association across Africa, the heterogeneity in risk magnitude and I^2 was large (Figure 1, Table S1). For identified variables with associations in at least 2 surveys, the median odds ratio (in absolute value) was 1.73 (IQR: 1.18 to 73.83) and 2.6 (IQR: 1.24 to 263.09) for females and males respectively. Associations across at least 2

surveys were highly heterogeneous and had a median I^2 of 75% (IQR: 17.6 to 99.3) and 81.9 (IQR: 17.8 to 99.2) for females and males (Table S1). The median Nagelkerke R^2 among females was 3.6×10^{-4} (IQR: 8.6×10^{-5} to 1.17×10^{-3}), and among males the median R^2 was 3.29×10^{-4} (IQR: 8.39×10^{-5} to 3.29×10^{-5}) (Table S1).

Variables identified across sub-Saharan Africa

We identified 344 (5.4% out of 6,288 possible) and 373 (5.1% out of 7,251 possible) variables in their association with HIV+ (R^2 greater than 0.001 and p-value less than 1×10^{-6}) in males and females respectively. For variables that were surveyed among female participants in 11-19 countries (a total of 432 variables), we identified 35 variables (8.1%). Second, for variables that were surveyed among female participants in 20-29 countries (a total of 449 variables), we identified 90 variables (20%). Third, in the male sample, we identified 38 (2.6% out of 1450), 34 (10% out of 334), 36 (11.7% out of 307) variables. A total of 168 and 154 variables were assessed in all 29 countries for females and males. Of these, we identified 31 (18%) and 16 (10%) in females and males respectively. Among the 31 identified variables assessed in all 29 countries in females, the median R^2 was 2.9×10^{-3} , median odds ratio was 1.57, and the I^2 was 69.43%. For the 16 identified variables assessed in all 29 countries in males, the median R^2 was 2.5×10^{-3} , median odds ratio was 1.39, and median I^2 was 60.6%. The distribution of the R^2 were comparable for variables assayed in different number of countries (Figure 1A), while variables with extreme odds ratios diminished when appearing in larger number of countries (Figure 1B). Variables broadly exhibited high heterogeneity in their odds ratios across countries (Figure 1C), with a majority having I^2 greater than 50%.

We highlight several variables that show a striking relationship with HIV+ (Figures 2 and 3). The variable indicating women who are the head of the household was significantly associated with HIV in nearly 60% of study countries, explained 1% of the variation in HIV status on average (and nearly 4% in two countries), and was uniformly associated positively with HIV, with a meta-analytic odds ratio of 2.5 (2.5 increased odds for HIV+ relative to females who were not the head of the household) (min-max

1.1-3.5) for all countries (Figure 2A, first row). The analogous indicator, head of household among men, was identified in 40% of countries, explaining as much as 6% of the variability in HIV status, and was also positively associated with HIV+ status (Figure 2B, 2nd row). As above, the heterogeneity in the odds ratios for females and males was 75%. In other words, there was substantial variability in the odds ratios.

Other indicators of marital status were associated with HIV+ across the subcontinent, including the number of unions, if the participant had been married, and if they were divorced among females. First, females who had been in one union had 60% decreased odds across all countries in the subcontinent to be HIV+ and were identified in slightly less than 40% of the countries (range of the Nagelkerke R^2 was .1% to 4%, Figure 2A, 4th row). Second, if a woman was divorced, they had a consistently increased odds (median OR 2.4) in HIV+ relative to those not divorced. This was consistent throughout the study countries (I^2 estimate of 33%).

Males who were divorced had an inconsistent association with HIV+ (I^2 99%) across countries. Specifically, in 13 out of 28 countries divorced males had a near zero chance of HIV+ (Figure 2B, 14th line from top, see ORs $< 1 \times 10^{-5}$). These countries included Sao Tome and Principe, Kenya, Tanzania, Democratic Republic of Congo, Sierra Leone, Togo, Gabon, Chad, Senegal, Ghana, Guinea, Burkina Faso, and Rwanda. For the other countries, their associations had lower and insignificant association sizes (filled point in figure). Third, males who were in a union (or living with) a woman at the time of the survey had an average odds ratio of 2.2, or greater than 2-fold increase in odds versus relative to those who were not for HIV+, and this was consistent across the study countries, with Nagelkerke R^2 reaching up to 4% (Figure 2B, 5th row). Of note, marital status also stood out in our previous analysis of HIV+ in Zambia. [11]

Complex indicators of education status emerged as a key correlate in numerous countries. For example, we report a meta-analytic OR of 4.4 across 29 countries for the female participant not attending school (Figure 2A). While this indicator may overlap with young age, low educational attainment was also associated with higher risk of HIV among men (Figure 2B). Similarly, the meta-

analytic OR across 28 countries for males who completed primary school versus who did not was extremely small, 0.004 (Figure 2B, 8th row from the top) in 12 out of 28 countries, indicating near zero chance of HIV+. Countries with near zero number of males with HIV+ that have completed primary school included Namibia, Mozambique, Kenya, Gabon, Uganda, Ghana, Angola, Rwanda, Ethiopia, Guinea, Mali, and Congo.

We found several variables that reflected potential biological and co-morbid conditions with HIV+ (Figure 3AB). For example, we identified men who were not anemic had an overall 60% decreased odds of HIV+ (with Nagelkerke R^2 up to 4%). This variable was found in 25% of the countries it was measured (Figure 3B).

Prediction of HIV Status

The use of the top 10 variables improved prediction of HIV status relative to use of prevalence rates for risk prediction across every country in the study (Figure 4AB, see Supplementary Information for difference between AUC and PRAUC). For some countries (e.g. eSwatini) the predicted probability of HIV risk was fairly evenly distributed across the population (Gini=0.15) (Figure 5). On the other hand, the Gini coefficient was upwards of 0.5 in several countries (including Ethiopia, Niger, and females in Rwanda) (Figure 5).

Discussion

Here, we use a data-driven approach for discovery of candidate risk factors for HIV at scale, and in doing so, present extensive correlates of HIV infection in Sub-Saharan Africa that span economic, biological, and environmental domains. Finally, we apply this approach for HIV risk prediction across the sub-continent, and show that risk assessment of HIV status can be improved using this non-invasive epidemiologic approach.

Several identified risk factors deserve further discussion because of their strong association with HIV status, their pervasive presence across the study countries, and their relatively large explanatory power. Being the head of the household stands out for identifying relatively large risk of HIV among females in the majority of our study countries. This is an intuitive correlate of risk, and possibly identifies women whose husbands had HIV and passed away, but its relative ubiquity has not been previously recognized. Complements of this risk factor – women living in households with a male head and who have been in a single marriage – are at lower risk of having HIV. Variables that characterize marital status commonly are also closely associated with HIV status.

Another important identified variable group relates to schooling and educational attainment among both males and females. Not attending school was a consistent correlate of increased risk among females, while completion of primary school was a protective correlate among men. Because school and educational attainment indicators are often closely correlated with one another, additional schooling correlates would be identified if our heuristic (for example, the one hot-encoded complement of attending school among females) did not eliminate them. Finding educational correlates that are in line with causal effect estimates among our top variables supports the importance of schooling in reducing HIV risk. [15]

Several less intuitive variables were also commonly associated with HIV. Non-ownership of livestock is a consistent positive correlate of HIV, among both males and females. We believe that this may be more closely associated with residence in urban environments (non-ownership of agricultural land also carries elevated risk of HIV among both males and females, in fewer surveys). Notably, we do not observe urban residence among the top candidate risk factors for HIV, suggesting livestock ownership carries additional (indirect) linkages to HIV risk.

A notable advantage of our approach is its systematic assessment of the associations of all the variables assessed in DHS, creating for the first time a database of robust (e.g., reproducible association sizes across multiple waves) across all DHS-surveyed sub-Saharan Africa countries. In comparison to “candidate” association studies that examine a few or a handful of associations at a time,

that a systematic approach may lead to more consistent positive, as well as negative, identification of important correlates. This highlights the importance of examining plausible pathways linking the identified variables to HIV: the consistency of the variables we identify lends them statistical strength, and that strength further benefits from real-world context.

A systematic approach also provides a database to contextualize associations. Male circumcision, originally identified in observational studies is associated with a meta-analytic relative risk of 0.52 (an absolute value RR of roughly 2) [7]. This association size would rank in the top 80th percentile of pan-sub-Saharan Africa associations available in this report. In other words, ~1257 variables would meet this threshold among all variables assessed in men and a fraction of those that are robust across the continent (Figure 1B). Further, we emphasize that the DHS aims to collect a representative sample of populations surveyed and we anticipate little sample selection bias among the potential database of correlates and HIV+.

Another strength of this large multi-country person-level analysis is the demonstration of HIV risk distribution in different populations. For example, we can measure the concentration of risk in different populations. We show large variation in HIV risk distribution between countries: HIV risk is highly concentrated among a small portion of the population in some countries, while risk is more distributed in the population in other countries. If all risk is concentrated in a small portion of the population, then identifying at-risk groups may be more feasible from a public health perspective than if risk is more evenly distributed across a large portion of the population. We also note that, in most countries, HIV risk is more concentrated in females than in males, and that HIV risk concentration is very high (upwards of 60%) in a handful of countries (Chad, Ghana, Togo, and Ethiopia).

There are limitations to our data-driven study. First, the associations that we identify emerge may be confounded. Second, we applied a stringent heuristic (Bonferroni pvalue and pseudo-R² thresholds) for identification of associations from a massive database. Therefore, some associations may be “false negative” and fail to be discussed. We have provided all associations for readers to examine the associations in the context of others in the Supplementary Tables. Third, our correlations

are cross-sectional and we rule out potential reverse-causal relationships. Fourth, aside from the HIV test, variables are self-reported and may have differential error rates. As we reported[11] random error or non-differential bias will lead to reduction of association sizes toward the null. In this extensive analysis of risk factors for HIV positivity, we are able to systematically characterize HIV risk factors, including identification of under-appreciated risk factors with meaningful association sizes, as well the extent to which risk factors are shared across countries and over time. This approach can be used to better characterize HIV risk using observational data, to identify hypotheses for HIV interventions, as well as serve as a platform for developing tools for identification risk of important non-HIV outcomes.

Funding

This work was supported by the National Institutes of Allergy and Infectious Diseases at National Institutes of Health [R01AI127250] and National Institute on Drug Abuse [R37DA015612].

References

1. 90-90-90: treatment for all. Available at: <http://www.unaids.org/en/resources/909090>. Accessed 12 April 2019.
2. UNAIDS data 2019. Available at: <https://www.unaids.org/en/resources/documents/2019/2019-UNAIDS-data>. Accessed 11 February 2020.
3. Serwadda D, Wawer MJ, Musgrave SD, Sewankambo NK, Kaplan JE, Gray RH. HIV risk factors in three geographic strata of rural Rakai District, Uganda. *Aids* **1992**; 6:983–990.
4. Cain D, Simbayi LC, Kalichman SC, Cherry C, Jooste S, Mfecane S. Risk factors for HIV-AIDS among youth in Cape Town, South Africa. **2015**;
5. Bulstra CA, Hontelez JAC, Giardina F, et al. Mapping and characterising areas with high levels of HIV transmission in sub-Saharan Africa: A geospatial analysis of national survey data. *PLOS Medicine* **2020**; 17:e1003042. Available at: <https://journals.plos.org/plosmedicine/article?id=10.1371/journal.pmed.1003042>. Accessed 3 July 2020.
6. Gersovitz M. The HIV epidemic in four African countries seen through the Demographic and Health Surveys. *Journal of African Economies* **2005**; 14:191–246.

7. Weiss HA, Quigley MA, Hayes RJ. Male circumcision and risk of HIV infection in sub-Saharan Africa: a systematic review and meta-analysis. *Aids* **2000**; 14:2361–2370.
8. Ioannidis JP. Why most published research findings are false. *Chance* **2005**; 18:40–47.
9. Ioannidis JP, Tarone R, McLaughlin JK. The false-positive to false-negative ratio in epidemiologic studies. *Epidemiology* **2011**; 22:450–456.
10. Patel CJ, Ioannidis JP. Studying the elusive environment in large scale. *Jama* **2014**; 311:2173–2174.
11. Patel CJ, Bhattacharya J, Bendavid E. Systematic Identification of Social, Behavioral, Environmental, and Economic Correlates of HIV Infection in Zambia in 2013-2014: An X-Wide Association Study. *Aids* **2019**; 32(7): 933–943.
12. Staveteig SE, Bradley SEK, Nybro E, Wang S. Demographic Patterns of HIV Testing Uptake in Sub-Saharan Africa. DHS Comparative Reports No. 30. ICF International. 2013.
13. Hosmer D, Lemeshow S, Sturdivant RX. Applied Logistic Regression. John Wiley & Sons, 2013.
14. Lumley T. Analysis of complex survey samples. *Journal of Statistical Software* **2004**; 9:1–19.
15. Duflo E, Dupas P, Kremer M. Education, HIV, and Early Fertility: Experimental Evidence from Kenya. *American Economic Review* **2015**; 105:2757–2797. Available at: <https://www.aeaweb.org/articles?id=10.1257/aer.20121607>. Accessed 11 February 2020.
16. Sorokin A, Forsyth D. Utility data annotation with Amazon Mechanical Turk. In: 2008 IEEE Computer Society Conference on Computer Vision and Pattern Recognition Workshops. 2008: 1–8.
17. Borenstein M, Hedges LV, Higgins JPT, Rothstein HR. Introduction to Meta-Analysis. 1st Edition. Wiley, 2011.
18. Viechtbauer W. Conducting Meta-Analyses in R with the metafor Package. *Journal of Statistical Software* **2010**; 36:1–48. Available at: <https://www.jstatsoft.org/index.php/jss/article/view/v036i03>. Accessed 6 October 2020.
19. ineq: Measuring Inequality, Concentration, and Poverty version 0.2-13 from CRAN. Available at: <https://rdr.io/cran/ineq/>. Accessed 13 October 2020.
20. Milanovic B. A simple way to calculate the Gini coefficient, and some implications. *Economics Letters* **1997**; 56:45–49. Available at: <http://www.sciencedirect.com/science/article/pii/S0165176597001018>. Accessed 6 October 2020.

Figure and Table Legends

Table 1. Sample sizes across sub-Saharan Africa

Table 2. Distributions of odds ratios, Nagelkerke R^2 , and I^2 (heterogeneity) estimates across countries.

Table S1. Distributions of odds ratios, Nagelkerke R^2 , and I^2 (heterogeneity) estimates for variables appearing in 1 country and at least 2 countries.

Figure 1. Volcano plots of $-\log_{10}(\text{pvalue})$ versus R^2 and Odds Ratios. (A) Nagelkerke R^2 (B) Odds Ratios (C) Heterogeneity I^2 of identified variables across all surveys. OR capped at 0.01 and 100 for visualization purposes. In (A) and (B), Dotted line denotes Bonferroni-level of significance ($1e-6$). In C, dotted line corresponds to I^2 of 50%. f: females. m: males.

Figure 2. Top 15 variables identified across 21-29 countries. (A) Top 15 variables in females and (B) Top 15 variables in males. Left most panel is the variable name and code name, 2nd panel from left is the percent of countries variable was identified, 3rd panel from left is the Nagelkerke R^2 per country (red dot is the average; blue points are countries with OR > 1, dark blue are countries with OR < 1), 4th panel shows the odds ratios (overall meta-analytic estimate in the red dot), and 5th panel shows the I^2 (heterogeneity).

Figure 3. Top 15 variables identified across 11-20 countries. (A) Top 15 variables in females and (B) Top 15 variables in males. Left most panel is the variable name and code name, 2nd panel from left is the percent of countries variable was identified, 3rd panel from left is the Nagelkerke R^2 per country (red dot is the average; blue points are countries with OR > 1, dark blue are countries with OR < 1), 4th panel shows the odds ratios (overall meta-analytic estimate in the red dot), and 5th panel shows the I^2 (heterogeneity).

Figure 4. Predictive capability of top 10 variables (A) Area under the receiver-operator-curve (ROC) and (B) Are under the precision-recall curve. Left panels are prediction for males and right panel for females. Red points in (B) denote the prevalence of HIV+.

Figure 5. Concentration of HIV predicted risk (Gini coefficient). Gini coefficient of the predicted probability per country in Sub-Saharan Africa

Figure S1. Number of Variables assessed as a function of number of surveys and countries.

Figure S2. Empirical CDF of (A) Nagelkerke R^2 , (B) $\exp(\text{absolute value}(\beta))$ or OR, and (C) Heterogeneity (I^2). Red line depicts CDF for those not identified, blue line identified (e.g., $\text{pvalue} < 1e-6$ and $R^2 > 0.001$).

Tables

Table 1. Sample sizes across sub-Saharan Africa

Country	Sample # Size	Cases	HIV	HIV	HIV	Mean Age (SE)	% Female (SE)	% Urban (SE)
			Prevalence (SE)	Prevalence Female (SE)	Prevalence Male (SE)			
Angola	12260	271	1.95 (0.2)	2.57 (0.28)	1.21 (0.22)	27.71 (0.16)	54.21 (0.57)	70.77 (2.41)
Burkina_Faso	23434	295	1.17 (0.09)	1.29 (0.11)	1.04 (0.13)	29.1 (0.1)	53.93 (0.33)	26.43 (1.85)
Burundi	8790	167	1.43 (0.15)	1.72 (0.19)	1.11 (0.17)	27.75 (0.15)	52.78 (0.54)	12.55 (1.62)
Cameroon	10348	565	5.35 (0.3)	6.75 (0.44)	3.91 (0.31)	28.63 (0.12)	50.5 (0.49)	56.23 (2.68)
Chad	11034	172	1.56 (0.17)	1.8 (0.23)	1.29 (0.21)	28.11 (0.16)	52.73 (0.53)	25.74 (2.36)
Congo_Democratic_Republic	27530	313	1.17 (0.13)	1.62 (0.2)	0.7 (0.11)	28.96 (0.13)	51.76 (0.39)	39.74 (2.36)
Congo	12382	379	3.16 (0.3)	4.12 (0.46)	2.06 (0.28)	28.95 (0.18)	53.16 (0.42)	63.07 (3.91)
Cote_d'Ivoire	18070	721	4.33 (0.24)	5.49 (0.34)	3.11 (0.28)	27.82 (0.16)	51.15 (0.52)	49.44 (3.19)
Ethiopia	41165	787	1.43 (0.12)	1.86 (0.17)	0.98 (0.1)	28.31 (0.09)	51.48 (0.28)	20.79 (1.97)
Gabon	11264	502	4.24 (0.38)	5.81 (0.6)	2.69 (0.36)	28.65 (0.18)	49.67 (0.64)	87.65 (1.74)
Ghana	18795	351	1.96 (0.13)	2.56 (0.2)	1.31 (0.15)	29.89 (0.11)	51.79 (0.38)	49.62 (2.04)
Guinea	15407	272	1.69 (0.14)	2.03 (0.2)	1.27 (0.14)	29.71 (0.13)	55.12 (0.39)	36.11 (2.42)
Kenya	13359	902	6.54 (0.39)	8.32 (0.48)	4.63 (0.4)	28.8 (0.12)	51.78 (0.5)	24.71 (2.15)
Lesotho	18730	4308	23.67 (0.47)	27.57 (0.59)	18.99 (0.57)	28.66 (0.1)	54.52 (0.45)	29.57 (1.76)
Liberia	20362	320	1.65 (0.15)	1.88 (0.18)	1.39 (0.2)	29.37 (0.15)	54.04 (0.42)	48.69 (2.74)
Malawi	34409	3514	10.25 (0.31)	12.13 (0.39)	8.24 (0.33)	28.37 (0.08)	51.61 (0.28)	18.92 (1.18)
Mali	17978	199	1.16 (0.12)	1.36 (0.15)	0.93 (0.14)	29.58 (0.12)	53.25 (0.38)	30.21 (2.33)
Mozambique	16976	1466	7.48 (0.47)	8.75 (0.56)	5.97 (0.47)	32.49 (0.22)	54.43 (0.4)	31.03 (3.53)
Namibia	9309	1297	14.33 (0.6)	16.84 (0.75)	11.45 (0.81)	32.69 (0.23)	53.51 (0.72)	54.62 (2.64)
Niger	7897	70	0.63 (0.1)	0.58 (0.12)	0.7 (0.15)	30.31 (0.16)	57.32 (0.6)	22.55 (2.38)
Rwanda	41044	1206	2.8 (0.11)	3.38 (0.15)	2.17 (0.11)	28.8 (0.07)	52.28 (0.21)	17.15 (1.24)
Sao_Tome_and_Principe	4860	78	1.54 (0.25)	1.29 (0.29)	1.79 (0.35)	29.77 (0.21)	50.49 (0.92)	52.15 (5.62)
Senegal	18122	123	0.52 (0.07)	0.57 (0.11)	0.43 (0.12)	28.37 (0.11)	53.66 (0.73)	52.62 (2.53)
Sierra Leone	21619	328	1.46 (0.12)	1.68 (0.15)	1.22 (0.16)	29.69 (0.13)	52.87 (0.29)	36.29 (2.36)

Swaziland	13008	2481	18.89 (0.58)	22.25 (0.69)	15 (0.67)	26.89 (0.12)	53.69 (0.53)	23.02 (2.47)
Tanzania	44896	2180	5.8 (0.19)	6.69 (0.24)	4.75 (0.22)	28.45 (0.06)	54.18 (0.26)	26.84 (1.54)
Togo	9321	200	2.5 (0.22)	3.11 (0.3)	1.85 (0.24)	29.34 (0.14)	51.65 (0.52)	44.81 (3.05)
Uganda	31644	1592	5.2 (0.21)	6.02 (0.27)	4.26 (0.22)	30.65 (0.11)	53.4 (0.31)	18.05 (2.08)
Zambia	40821	5634	13.68 (0.37)	15.34 (0.45)	11.97 (0.39)	28.68 (0.07)	50.7 (0.26)	45.19 (2.04)
Zimbabwe	43835	7158	15.71 (0.3)	18.35 (0.37)	12.76 (0.33)	28.01 (0.06)	52.8 (0.31)	35.71 (1.61)

Table 2. Distributions of odds ratios, Nagelkerke R², and I² (heterogeneity) estimates across countries. Num Assoc. denotes the number of associations for that category.

gender/number	Num Assoc.	25th OR	Median OR	75th OR	25th R2	Median R2	75th R2	25th I2	Median I2	75th I2
1 Country										
Female	4543	1.2	1.62	3.69	8.56 × 10 ⁻⁵	3.58 × 10 ⁻⁴	1.17 × 10 ⁻³	.	.	.
Male	4089	1.24	1.78	12.67	8.39 × 10 ⁻⁵	3.29 × 10 ⁻⁴	1.12 × 10 ⁻³	.	.	.
2-10 Countries										
Female	1702	1.19	1.67	136.4	2.56 × 10 ⁻⁴	6.30 × 10 ⁻⁴	1.39 × 10 ⁻³	0	62.1	99.01
Male	1488	1.24	2.3	526.3	2.42 × 10 ⁻⁴	5.91 × 10 ⁻⁴	1.26 × 10 ⁻³	0	64.85	99.14
11-20 Countries										
Female	467	1.15	3.93	110.95	4.10 × 10 ⁻⁴	7.67 × 10 ⁻⁴	1.22 × 10 ⁻³	49.8	98.59	99.39
Male	368	1.17	3.44	189.43	4.09 × 10 ⁻⁴	7.70 × 10 ⁻⁴	1.57 × 10 ⁻³	47.49	98.38	99.41
20-29 Countries										
Female	539	1.17	1.65	9.88	4.77 × 10 ⁻⁴	8.42 × 10 ⁻⁴	1.42 × 10 ⁻³	56.04	97.57	99.35
Male	343	1.26	3.68	56.47	4.71 × 10 ⁻⁴	7.10 × 10 ⁻⁴	1.22 × 10 ⁻³	67.7	99.14	99.48

Figure 1.

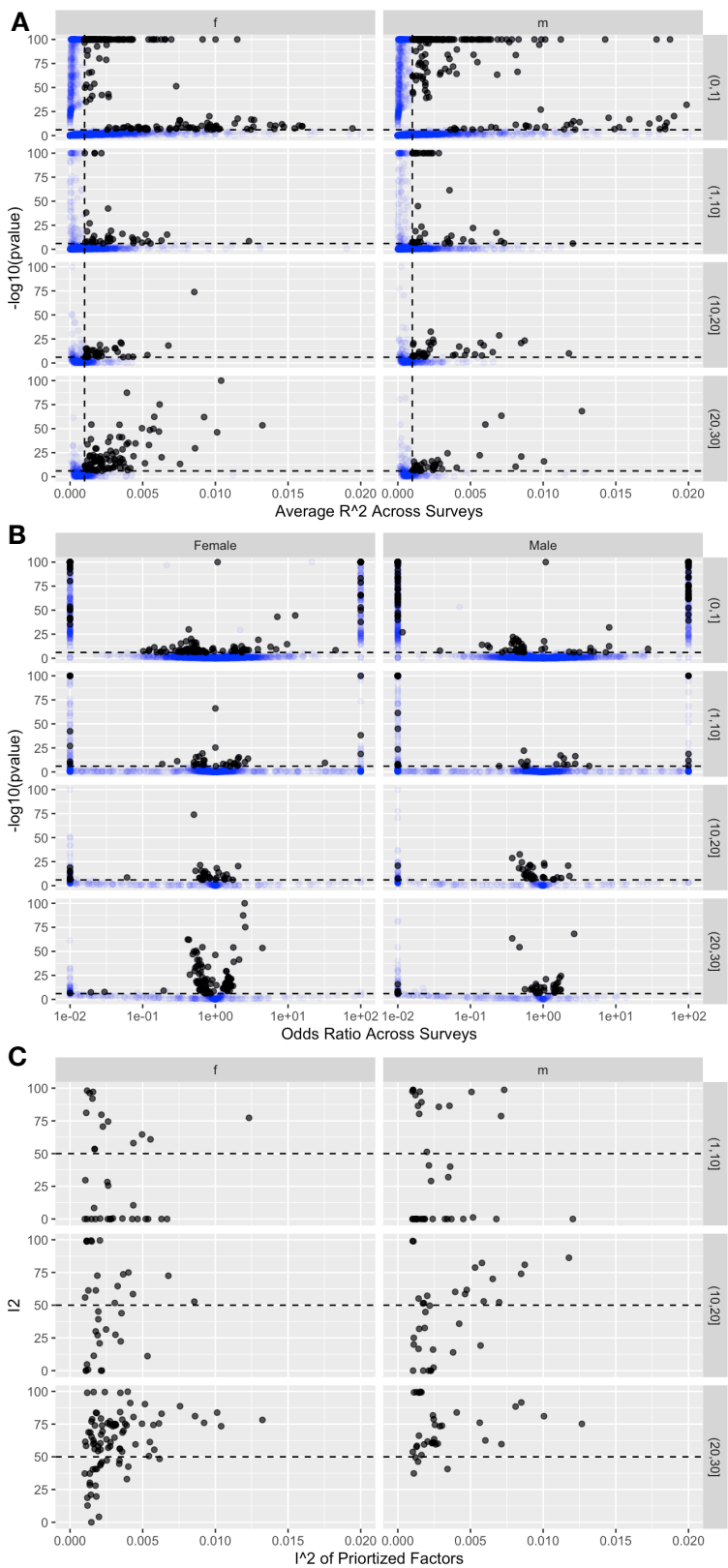


Figure 2 (2A and 2B).

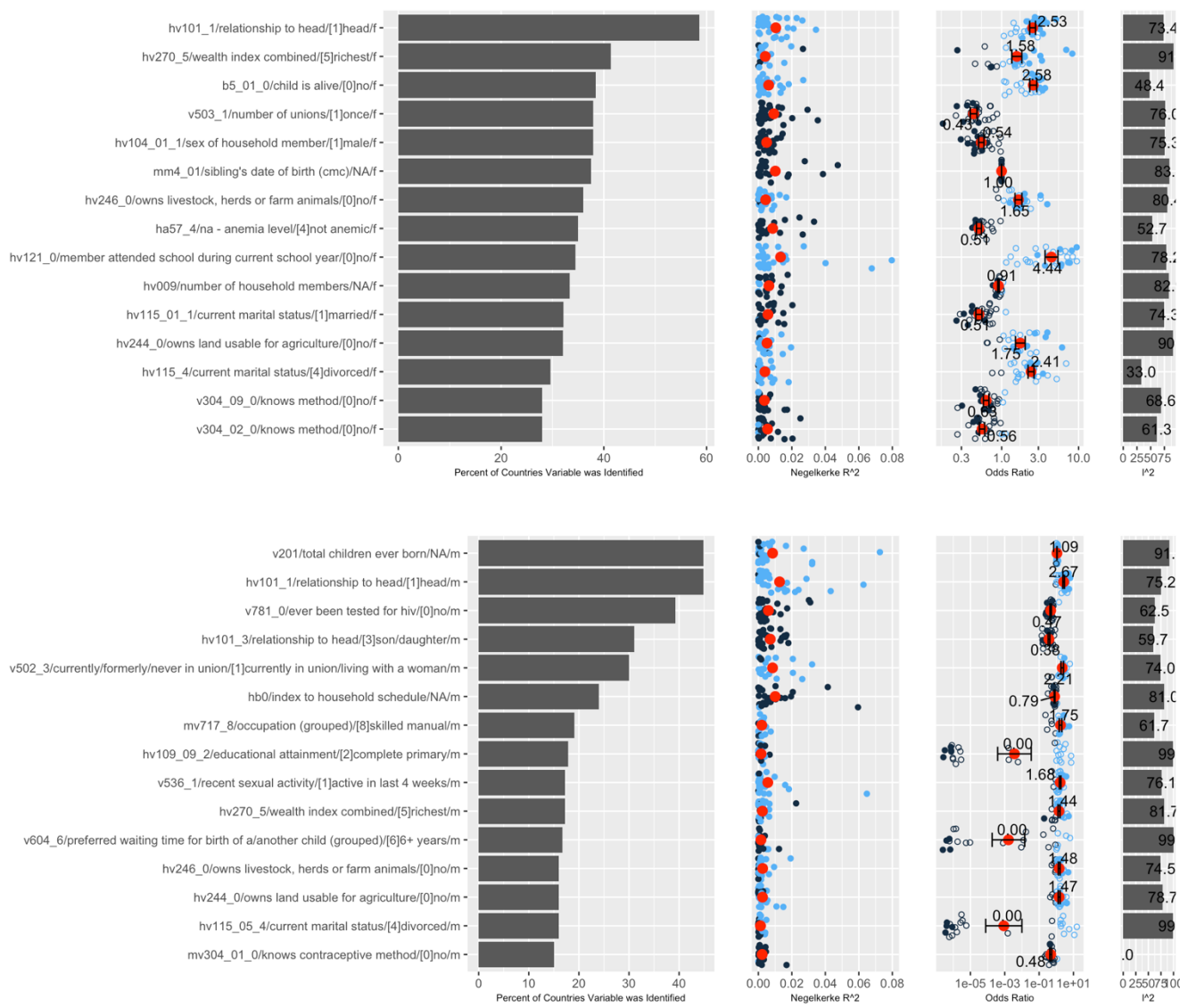


Figure 3 (3A and 3B).

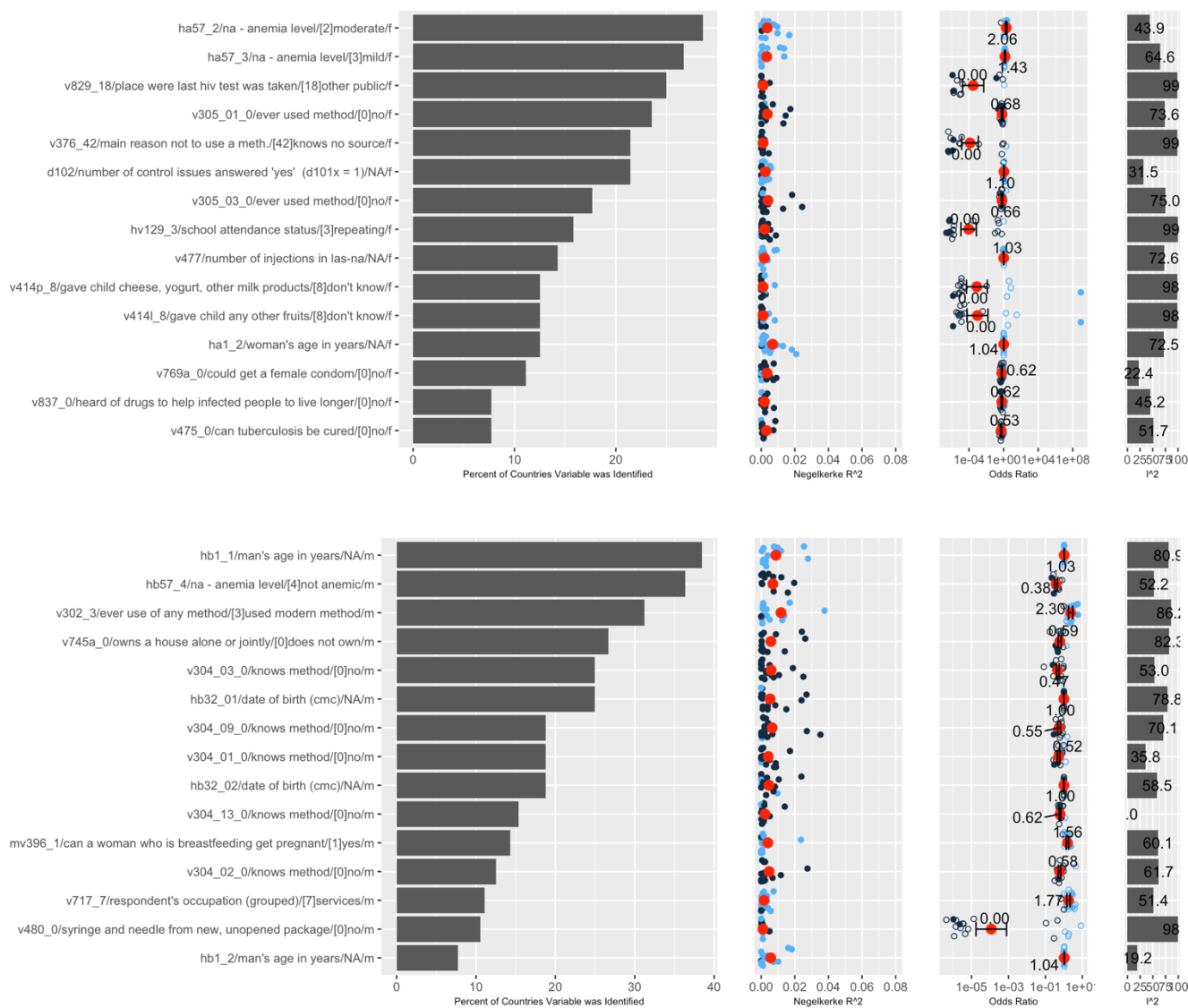


Figure 4.

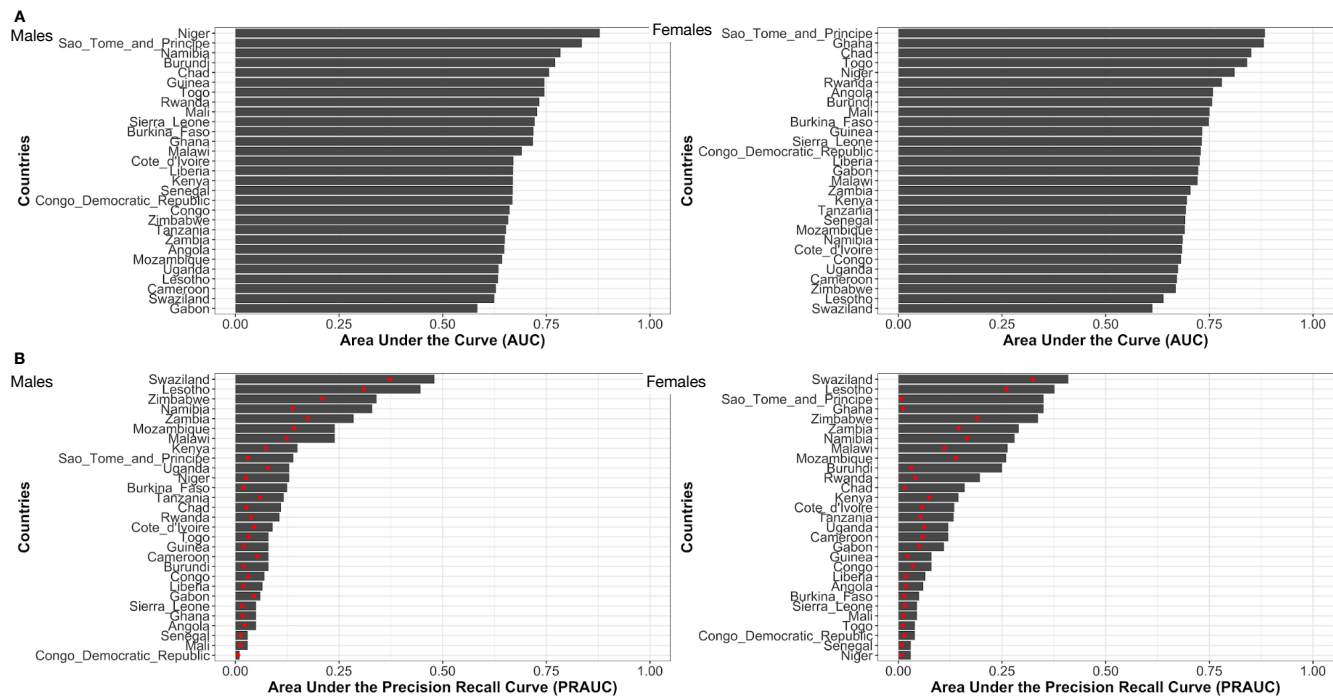


Figure 5.

