bioRxiv preprint doi: https://doi.org/10.1101/600510; this version posted November 6, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Socio-behavioural characteristics and HIV: findings from a graphical modelling analysis of 29 sub-Saharan African countries

Zofia Baranczuk[1,2,3*], Janne Estill[1,4*§], Sara Blough[1,5], Sonja Meier[6], Aziza Merzouki[1], Marloes H. Maathuis[6*], Olivia Keiser[1*]

1 Institute of Global Health, University of Geneva, Geneva, Switzerland
2 Department of Psychology, University of Zurich, Zurich, Switzerland
3 Institute of Mathematics, University of Zurich, Zurich, Switzerland
4 Institute of Mathematical Statistics and Actuarial Science, University of Bern, Bern, Switzerland
5 Rollins School of Public Health, Emory University, Atlanta, GA, United States
6 Seminar for Statistics, ETH Zürich, Zurich, Switzerland

§ Corresponding author: Janne Estill

University of Geneva
9, Chemin des Mines
1202 Geneve, Switzerland
+41 31 631 88 18
janne.estill@unige.ch

*These authors have contributed equally to the work.

E-mail addresses of authors:

ZB: zofia.baranczuk@psychologie.uzh.ch

JE : janne.estill@unige.ch

SB: sblough@elon.edu

SM: sonja.h.meier@gmail.com

AM: fatmaaziza.merzouki@unige.ch

MM: maathuis@stat.math.ethz.ch

OK: olivia.keiser@unige.ch

Keywords: HIV epidemiology, risk factors, Africa, Bayesian network, graphical model, demographic and health surveys (DHS), socio-behavioural factors

bioRxiv preprint doi: https://doi.org/10.1101/600510; this version posted November 6, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Abstract

Introduction: Socio-behavioural factors may contribute to the wide variance in HIV prevalence between and within sub-Saharan African (SSA) countries. We studied the associations between socio-behavioural variables potentially related to the risk of acquiring HIV.

Methods: We used Bayesian network models to study associations between socio-behavioural variables that may be related to HIV. A Bayesian network consists of nodes representing variables, and edges representing the conditional dependencies between variables. We analysed data from Demographic and Health Surveys conducted in 29 SSA countries between 2010 and 2016. We predefined and dichotomized twelve variables, including factors related to age, literacy, HIV knowledge, HIV testing, domestic violence, sexual activity, and women's empowerment. We analysed data on men and women for each country separately and then summarised the results across the countries. We conducted a second analysis including also the individual HIV status in a subset of 23 countries where this information was available. We presented summary graphs showing associations that were present in at least six countries (five in the analysis with HIV status).

Results: We analysed data from 190,273 men (range across countries 2,295–17,359) and 420,198 women (6,621–38,948). The two variables with the highest total number of edges in the summary graphs were literacy and rural/urban location. Literacy was negatively associated with false beliefs about AIDS and, for women, early sexual initiation, in most countries. Literacy was also positively associated with ever being tested for HIV and the belief that women have the right to ask their husband to use condoms if he has a sexually transmitted infection. Rural location was positively associated with false beliefs about HIV and the belief that beating one's wife is justified, and negatively associated with having been tested for HIV. In the analysis including HIV status, being HIV positive was associated with female-headed household, older age and rural location among women, and with no variables among men.

Conclusions: Literacy and urbanity were strongly associated with several factors that are important for HIV acquisition. Since literacy is one of the few variables that can be improved by interventions, this makes it a promising intervention target.

bioRxiv preprint doi: https://doi.org/10.1101/600510; this version posted November 6, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Introduction

In sub-Saharan Africa, about 26 million people were living with HIV in 2018 [1]. HIV prevalence is high overall, but very heterogeneous between countries, ranging from under 1% in Senegal and Niger to up to 25% in Lesotho and eSwatini (previously Swaziland) [2–6]. The HIV prevalence also varies substantially within countries, the overall epidemic being concentrated in clustered micro-epidemics of different geographical scales [7]. The Joint United Nations Programme on HIV/AIDS (UNAIDS) has urged researchers to pay attention to these variations, and to identify geographical areas where the risk of acquiring HIV is higher: "know your epidemic, know your response" [7]. Variation has been attributed to many factors, including socio-demographic, behavioural, and biological [3,8]. Among the factors associated with HIV burden are age, religion, marital status, high levels of HIV-related stigma, sexual coercion, education, occupation, and gender [4,8].

Education can affect HIV prevalence. Even in countries where everyone knows HIV risk is high, the public may not know much about how the disease spreads. For example, those who have not been taught why and how to use condoms are less likely to protect themselves [8]. Attitudes and behaviour can also contribute to HIV risk. If false beliefs about HIV and AIDS are not corrected and replaced with specific, accurate knowledge, this can increase risky sexual behaviour and infection risk [9]. Women with little education are likely to be less informed about interventions that mitigate HIV risk [9], but higher HIV rates among women cannot be fully explained by either lower educational levels or higher biological susceptibility. In countries with high HIV prevalence, gender inequality has been associated with HIV spread along with low levels of general education [10] and sexual education [8]. Women with limited or no income may engage in transactional sex to survive [11], so socioeconomic and living conditions can also be factors in HIV transmission. Urban residence may increase the risk of infection by increasing the probability of high risk sexual contacts [12], but it may also decrease HIV risk because urban dwellers are more likely to be employed and have better access to care and treatment [13].

Most studies about risk factors for HIV or for the uptake of interventions against HIV have used logistic regression models to simultaneously adjust for a range of factors [14,15]. Such an approach, however, does typically not account for the complex network of associations and

bioRxiv preprint doi: https://doi.org/10.1101/600510; this version posted November 6, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

causal pathways between various risk factors and HIV. Understanding this network could help us identify drivers of the epidemic to plan effective interventions and ultimately reach the global treatment targets set by United Nations Member States [2]. We thus explored the associations between socio-behavioural variables that could contribute to the HIV epidemic by analysing data from Demographic and Health surveys from 29 countries in Sub-Saharan Africa using Bayesian network analysis [16].

Methods

Data description

The Demographic and Health Surveys (DHS) Program collects and disseminates high-quality, nationally representative data on fertility, family planning, maternal and child health, gender, HIV/AIDS, malaria, and nutrition in low- and middle-income countries around the world. Typically, it collects data on between 3,300 and 5,300 variables per country every 5 years. The DHS surveys include women of reproductive age (15-49 years) and (depending on the country) men aged 15-49, 15-54 or 15-59 years.

We used data from the DHS and included all countries in sub-Saharan Africa where a DHS survey was administered during or after 2010 [6]. We included the last available survey for each country (as of August 2018): Angola 2015-16; Benin 2011-12; Burkina Faso 2010; Burundi 2010; Cameroon 2011; Chad 2014-15; Congo 2011-12; Democratic Republic of the Congo 2013-14; Côte d'Ivoire 2011-12; Ethiopia 2016; Gabon 2012; The Gambia 2013; Ghana 2014; Kenya 2014; Lesotho 2014; Liberia 2013; Malawi 2015-16; Mali 2012-13; Mozambique 2011; Namibia 2013; Niger 2012; Nigeria 2013; Rwanda 2014-15; Senegal 2010-11; Sierra Leone 2013; Togo 2013-14; Uganda 2011; Zambia 2013-14; and, Zimbabwe 2015. We excluded Tanzania from the analysis because some essential variables were missing from their DHS data.

Two researchers (ZB and OK) preselected variables and dichotomized them before analysis. We included the following variables because they covered topics that could relate to HIV and were available for all selected countries: age 24 or younger (yes, no); rural location (yes, no); female head of household (yes, no); literacy (able to read whole sentence: yes, no/missing); access to media at least once a week (yes, no/missing); sexual initiation before 16 years (yes, no);

bioRxiv preprint doi: https://doi.org/10.1101/600510; this version posted November 6, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

currently working (yes, no/missing); judging wife-beating justified (yes/does not know/missing, no); married or living together either now or previously (yes, no); false beliefs about AIDS (yes/does not know/missing, no); judging it justified for a wife to ask her husband to use a condom if he has a sexually transmitted infection (STI; yes, no/does no know/missing); ever tested for HIV (yes, no/missing); and diagnosed HIV positive at the time of the survey (yes, no/no result) [17].

Data analysis

We conducted exploratory data analyses and Bayesian network analyses separately for men and women in each country and included all variables except HIV status in our main analysis. The information on HIV status was available only for 23 countries, so we performed separate analyses that included HIV status for these countries.

First, we calculated and presented crude odds ratios for each pair of variables by country and sex. Then we conducted a Bayesian network analysis to determine the possible causal relationship between variables for each country, separately for men and women. Bayesian networks are graphical models in which nodes represent random variables and edges represent certain probabilistic dependencies between them. In particular, the presence of an edge means that the corresponding variables are dependent given any subset of the remaining variables. Our analysis assumes that there are no unmeasured confounders. While the presence or absence of an edge is always identifiable under our assumptions, the directions of edges may be unknown, so we can only learn a so-called completed partially directed acyclic graph (CPDAG) [18,19]. The theoretical framework for CPDAG learning is well-developed [20-23]. We used a fully automatic approach with the Incremental Association Markov Blanket (IAMB) algorithm [24] within the bnlearn package in R [25]. To improve the stability of the estimated graphs, we used bootstrap sampling. For each country and sex, we took 10,000 bootstrap samples, using weighted resampling to account for the sampling weights in the DHS data. For each bootstrapped data set, we obtained the corresponding estimated CPDAG. We then constructed a graph for each country as follows: two nodes were connected if the edge was present in at least 90% of the bootstrapped CPDAGs, and edges were oriented if they were oriented in that direction in at least 70% of the bootstrapped CPDAGs.

bioRxiv preprint doi: https://doi.org/10.1101/600510; this version posted November 6, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Finally, we determined an overall summary graph by sex that combined the information for all countries. In this summary graph we only considered edges present in at least six countries for the analysis without HIV status, or five countries for the analysis with HIV status. The orientation of edges in the summary graph was determined as follows: For each edge, the number of countries with an orientation in one direction, and the number of countries with an orientation in the other direction was determined. If the difference between these numbers of countries was at least six, the edge in the summary graph was oriented according to the majority orientation. We estimated the sign of an association conditionally on all other variables using the Cochran-Mantel-Haenszel odds ratio [26] and marginally by calculating the crude odds ratio. We considered associations negative if both the crude and the conditional odds ratio were less than one, and positive if both the crude and the conditional odds ratios were above one.

We used the open source R language, version 3.5.0 for our analysis [25]. The code is available on github (https://github.com/ZofiaBTZ/DAGS_DHS).

Ethical approval

The study was conducted using data from the DHS program, which are freely available on request, and thus no ethical approval was needed. Details of the ethical review process of DHS are available on the programme's website https://www.dhsprogram.com/What-We-Do/Protecting-the-Privacy-of-DHS-Survey-Respondents.cfm.

Results

We included 190,273 men (range across countries 2,295–17,359) and 420,198 women (6,621–38,948) in the analysis. The age distribution and percentage of people living in urban areas were similar for males and females (Table 1). Women were less likely to be able to read (median 46% vs. 68%), to have media access (61% vs. 74%), to currently work (54% vs. 79%), and more likely to have been tested for HIV (49% vs. 31%).

bioRxiv preprint doi: https://doi.org/10.1101/600510; this version posted November 6, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Table 1. Characteristics of persons included in the analysis.

DHS, 2010 - 2016, countries of sub-Saharan Africa		Female median (min, max)	Male median (min, max)
Younger than 25 years	No	60.0% (54.7%-65.5%)	62.9% (56.1%-72.1%)
	Yes	40.0% (34.5%-45.3%)	37.1% (27.9%-43.9%)
Rural setting	No	38.7% (10.4%-88.5%)	40.2% (14.9%-87.1%)
	Yes	61.3% (11.5%-89.6%)	59.8% (12.9%-85.1%)
Household head female	No	72.7% (43.7%-90.8%)	86.8% (71.8%-97.1%)
	Yes	27.3% (9.2%-56.3%)	13.2% (2.9%-28.2%)
Literacy	No (Not able to read whole sentence/missing)	56.6% (9.6%-89.7%)	32.0% (14.0%-74.1%)
	Yes (Able to read whole sentence)	43.4% (10.3%-90.4%)	68.0% (25.9%-86.0%)
Media access	Less than once a week	39.5% (8.0%-81.7%)	25.7% (4.2%-64.8%)
	At least once a week	60.5% (18.3%-92.0%)	74.3% (35.2%-95.8%)
First sex before the age of 16	No	67.4% (50.5%-93.6%)	79.9% (54.6%-97.7%)
	Yes	32.6% (6.4%-49.5%)	20.1% (2.3%-45.4%)
Currently working	No	46.5% (24.8%-79.1%)	21.4% (6.9%-44.0%)
	Currently working / have a job, but on leave during the last 7 days	53.5% (20.9%-75.2%)	78.6% (56%-93.1%)
Married	Never in partnership	28.3% (8.6%-53.5%)	40.8% (28.7%-62.9%)
	Currently or formerly married or living with a partner	71.7% (46.5%-91.4%)	59.2% (37.1%-71.3%)
False beliefs about AIDS	No	55.5% (35.8%-88.4%)	62.3% (44.5%-85.7%)
	Yes / Does not know/ Missing	44.5% (11.6%-64.2%)	37.7% (14.3%-55.5%)
Wife beating justified	No	50.6% (20.6%-83.3%)	88.5% (66.4%-95.5%)
	Yes / Does not know/ Missing	49.4% (16.7%-79.4%)	11.5% (4.5%-33.6%)
Justified asking husband to use condom if he has a sexually transmitted disease	No / Does not know/ Missing	18.2% (2.3%-60.6%)	11.5% (1.5%-29.8%)
	Yes	81.8% (39.4%-97.7%)	88.5% (70.2%-98.5%)
Ever tested for HIV	No/ Missing	50.8% (13.8%-86.7%)	68.8% (19.1%-92.6%)
	Yes	49.2% (13.3%-86.2%)	31.2% (7.4%-80.9%)

The median, minimum and maximum of percentages among all countries are shown.

bioRxiv preprint doi: https://doi.org/10.1101/600510; this version posted November 6, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

We found both differences and similarities in crude odds ratios between variables across the sexes. Figure 1 shows the associations between all pairs of variables, each rectangle showing the crude odds ratios for the 29 countries. For example, for both sexes, false beliefs about AIDS were positively associated with rural location and judging gender-based violence justified (represented by red rectangles in the figure). Additionally, for women false beliefs about AIDS were positively associated with sexual initiation before age 16. Having ever been tested for HIV was positively associated for both sexes with the belief that it is justified for a wife to ask the husband to use a condom if he has a sexually transmitted infection (STI), with being married, with media access, and with literacy. These findings were consistent across countries.

bioRxiv preprint doi: https://doi.org/10.1101/600510; this version posted November 6, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

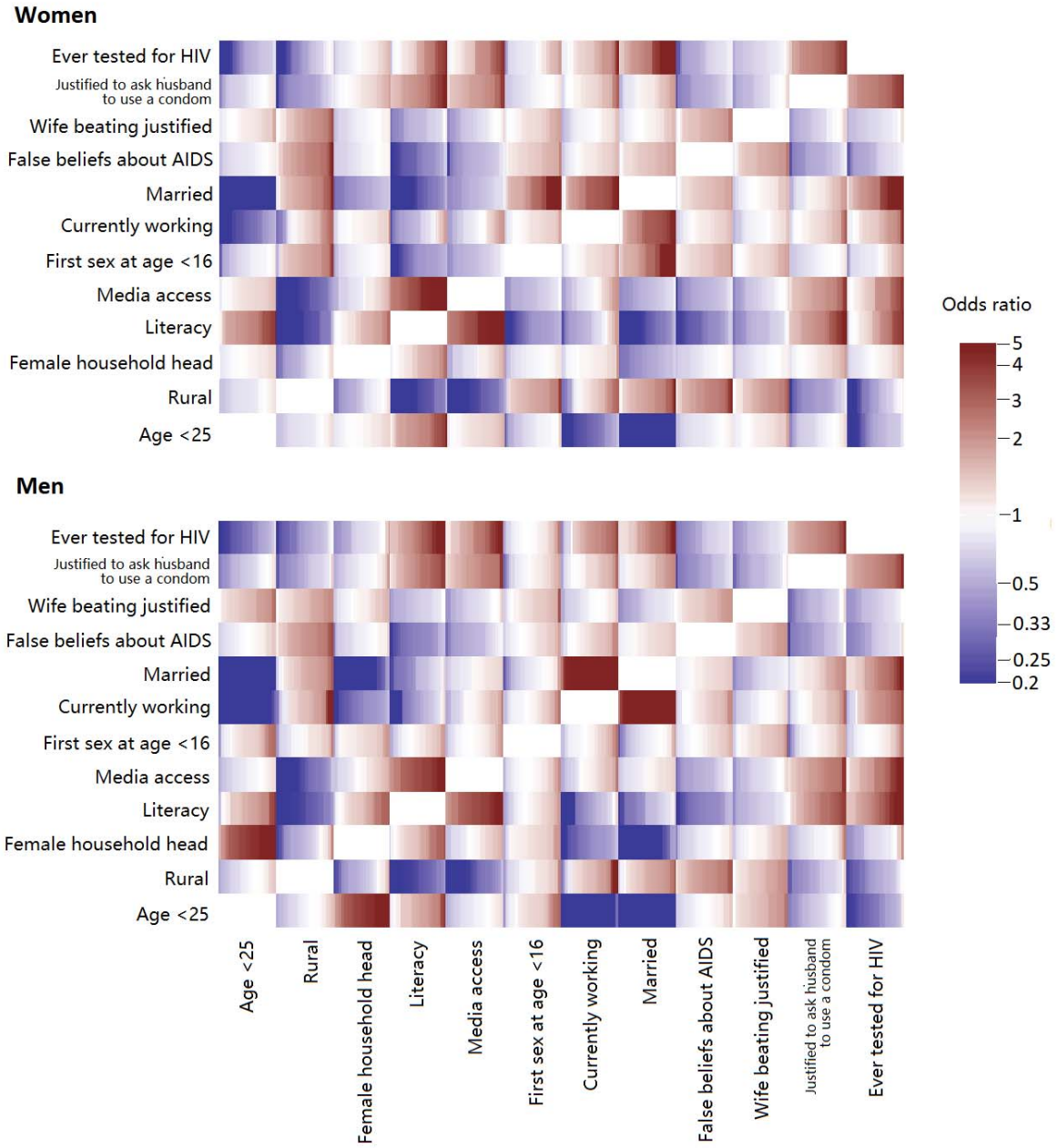


Figure 1. Crude odds ratios for all pairs of selected variables for women (upper panel) and men (lower panel) in the analysis without HIV status (29 countries). Each rectangle shows the odds ratios between the corresponding pair of variables for all countries, sorted from lowest to highest.

bioRxiv preprint doi: https://doi.org/10.1101/600510; this version posted November 6, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Figure 2 shows the estimated summary graphs from the Bayesian network analysis for women and men. The thickness of the edges is proportional to the number of countries for which the given edge was present, and the colour and pattern indicate if the association (i.e. the signs of both crude and conditional log odds ratios) was always negative (blue, dashed), positive (red, solid), or differed between countries (grey, dotted). In both men and women, literacy and urbanity had the largest number of edges in the summary graphs, including many with variables potentially linked with the probability of acquiring HIV. For instance, for women, literacy was negatively associated with sexual initiation before age 16 and false beliefs about AIDS in most countries (21 and 24, respectively). Moreover, for women, literacy was positively associated with the belief that it is justified to ask the husband to use a condom if he has an STI (12 countries) and having ever been tested for HIV (8 countries). For men, literacy was positively associated with having been tested for HIV and negatively associated with having false beliefs about AIDS in almost all countries (20 and 27 countries, respectively). In 13 countries, literacy among men was also positively associated with believing the wife is justified to ask the husband to use a condom if he has an STI. For both men and women, living in a rural area was positively associated with accepting domestic violence against the wife (women 16 countries, men 9 countries) and having false beliefs about AIDS (14 countries for both men and women), and negatively associated with having been tested for HIV (women 12 countries, men 11 countries). For women, the association between rural residence and false beliefs about AIDS was however not present in any country with high (>3%) HIV prevalence. Having been tested for HIV was positively associated with the belief that it is justified to ask the husband to use a condom (women 24 countries, men 8 countries). Having ever been tested for HIV was also associated with having access to media in 8 countries for women and 11 countries for men, all of which were countries with low (<3%) HIV prevalence. Among women, having been tested for HIV also was positively associated with being married in 21 countries. We found only very few directed edges, and will not interpret these in a fully causal way. Detailed results for each country are reported in Additional file 1.

bioRxiv preprint doi: https://doi.org/10.1101/600510; this version posted November 6, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

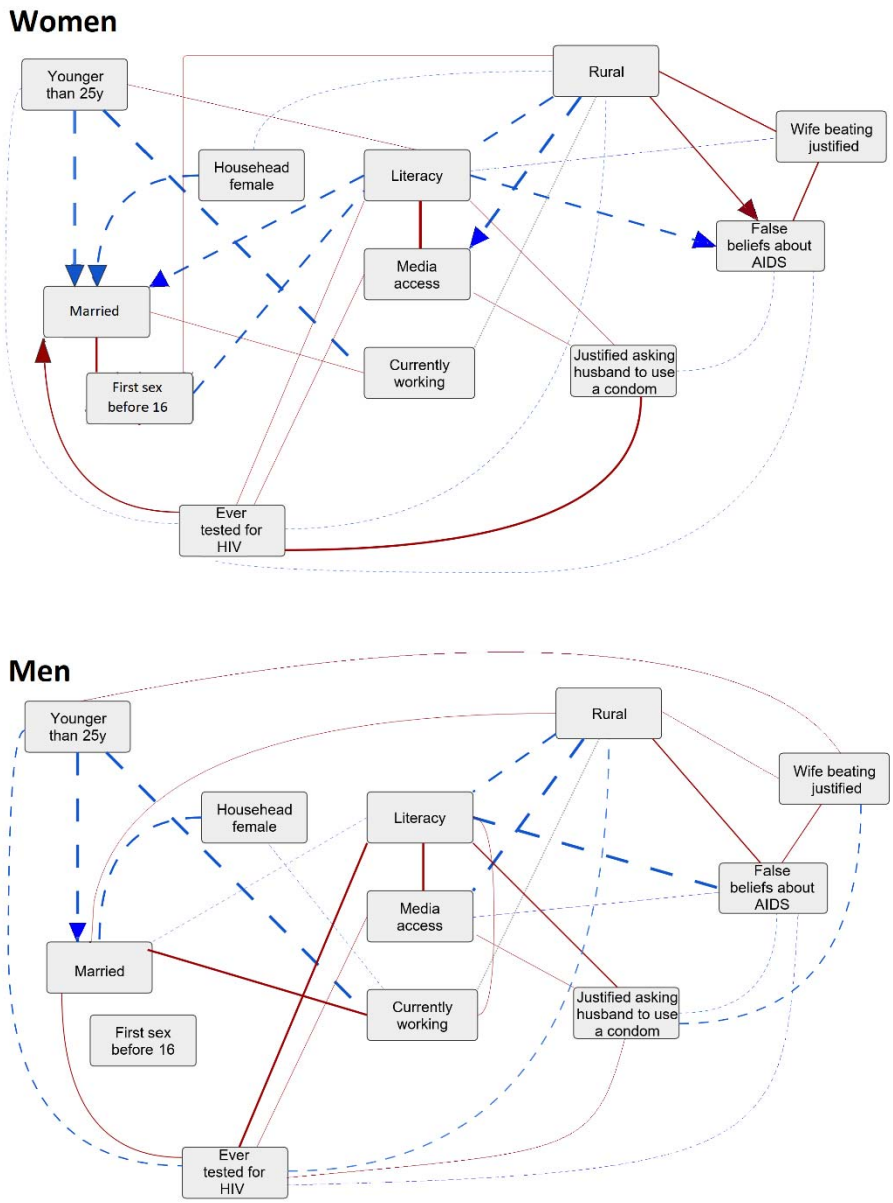


Figure 2. Summary graphs for the Bayesian network analysis without HIV status (29 countries) for women (upper panel) and men (lower panel). Shows associations present in at least six countries. Line thickness is proportional to the number of countries where the edge was present. The edges are oriented (arrowheads) if the number of countries with the shown direction was at least six higher than the number of countries with the opposite direction. Blue dashed lines indicate negative conditional associations, red solid lines indicate positive conditional associations and grey dotted lines indicate that the sign of the association differed across countries.

bioRxiv preprint doi: https://doi.org/10.1101/600510; this version posted November 6, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

HIV data were unavailable for Benin, Congo, Kenya, Mozambique, Nigeria and Uganda. In the analysis of the remaining 23 countries where we included HIV status, testing HIV positive was negatively associated with younger age and living in a rural location for both sexes (Figure 3). Being HIV infected was positively associated with being married, thinking it is justified for a wife to ask a husband to use a condom if he has an STI, and having ever been tested for HIV for both sexes. For women, having a female household head was also positively associated with HIV positivity.

bioRxiv preprint doi: https://doi.org/10.1101/600510; this version posted November 6, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

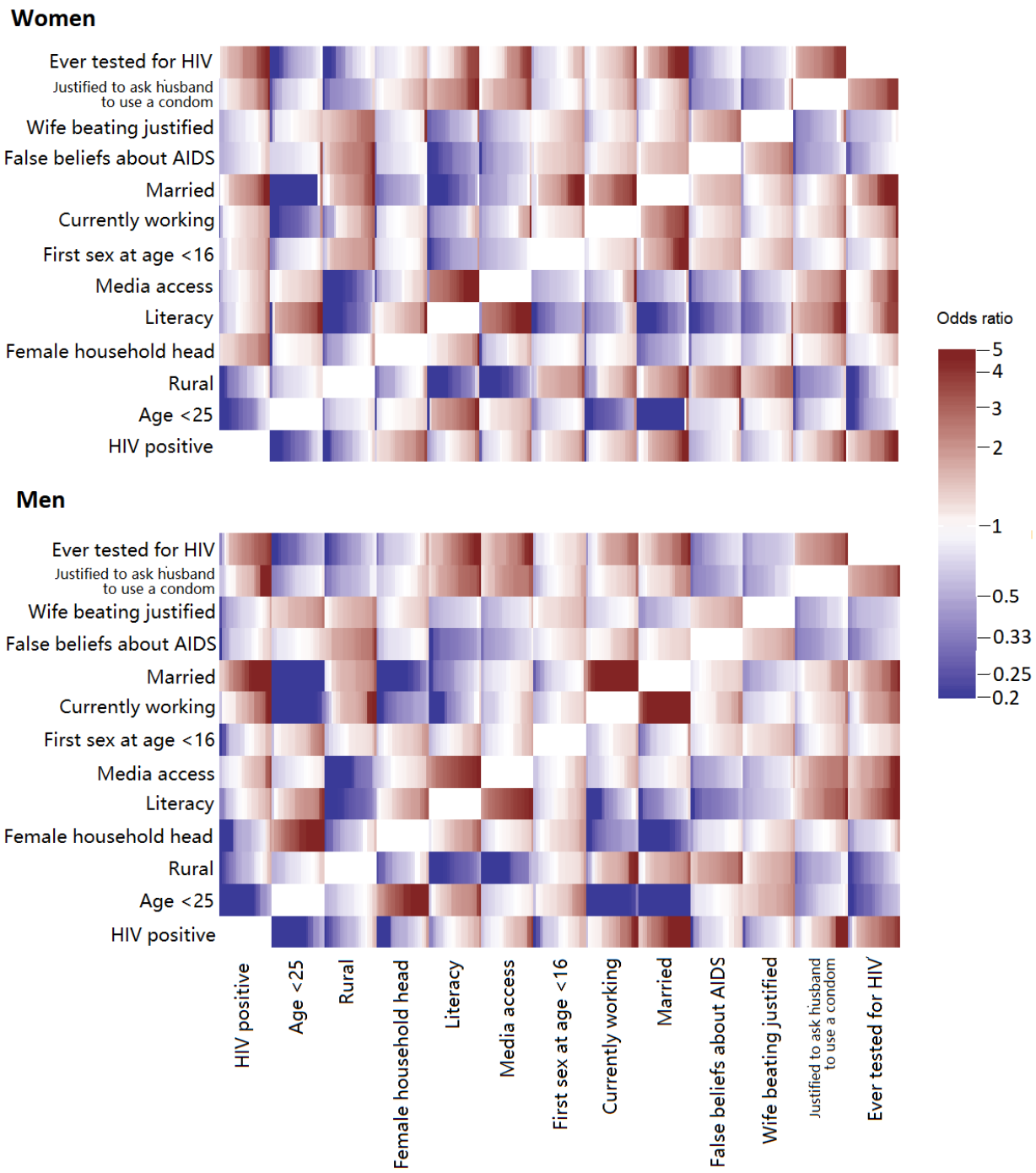


Figure 3. Crude odds ratios for all pairs of the selected variables for women (upper panel) and men (lower panel) in the analysis with HIV status (23 countries). Each rectangle shows the odds ratios between the corresponding pair of variables for all countries, sorted from lowest to highest.

bioRxiv preprint doi: https://doi.org/10.1101/600510; this version posted November 6, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

For men, adding HIV status yielded little difference: HIV status was not connected with any other variable in the summary graph (Figure 4). For women, being HIV positive was positively associated with having a female household head, and negatively associated with younger age and living in a rural area, in at least five countries.

bioRxiv preprint doi: https://doi.org/10.1101/600510; this version posted November 6, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

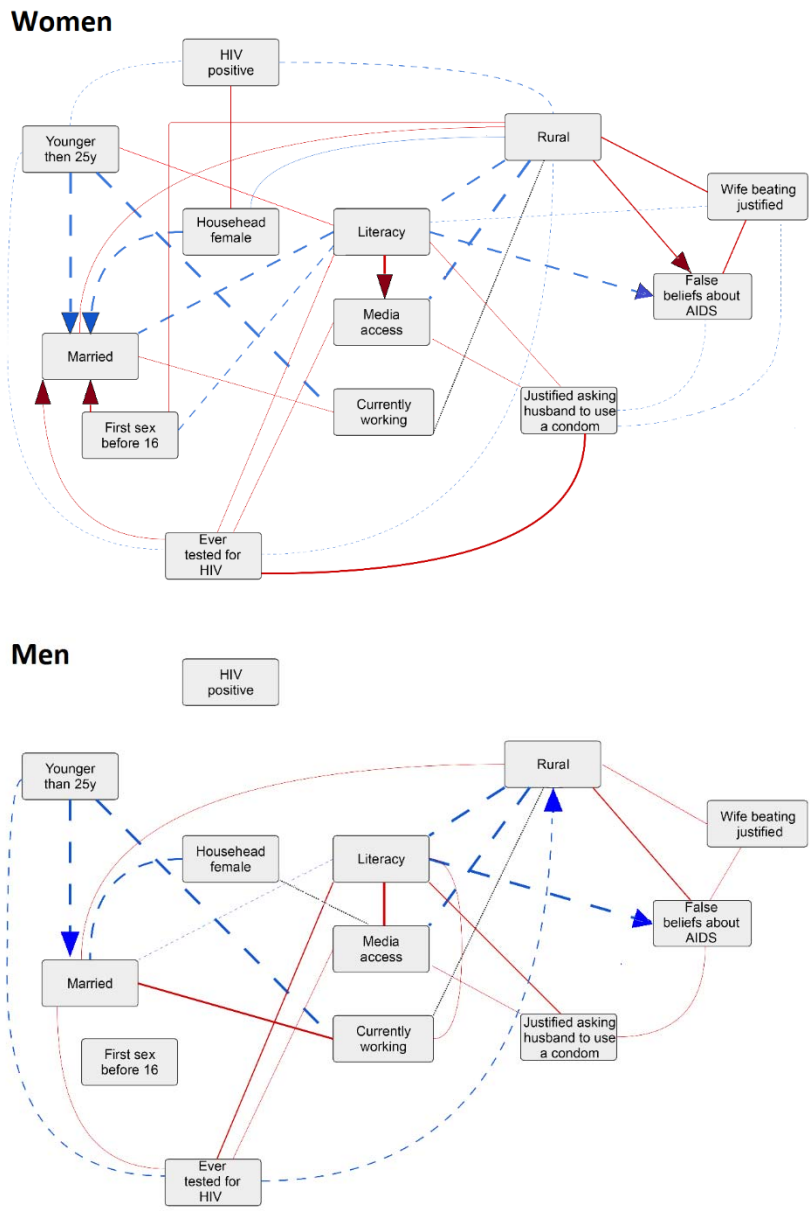


Figure 4. Summary graphs for the Bayesian network analysis with HIV status (23 countries) for women (upper panel) and men (lower panel). Shows associations present in at least five countries. Line thickness is proportional to the number of countries where the edge was present. The edges are oriented (arrowheads) if the number of countries with the shown direction was at least five higher than the number of countries with the opposite direction. Blue dashed lines indicate negative conditional associations, red solid lines indicate positive conditional associations and grey dotted lines indicate that the sign of the association differed across countries.

bioRxiv preprint doi: https://doi.org/10.1101/600510; this version posted November 6, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Discussion

Our analysis revealed a detailed picture of crude and fully adjusted associations between many socio-behavioural variables that may be linked to HIV in sub-Saharan African countries. Literacy and urbanity were the most central variables: they were directly connected to at least half of the remaining variables for both men and women in at least six countries. Access to media, information on HIV and HIV testing differed greatly between urban and rural locations. Having false beliefs about AIDS was strongly connected with rural location, being illiterate and accepting domestic violence against women. The likelihood of ever being tested for HIV was higher among people who were married, literate, or living in urban settings. For women, HIV testing was also positively associated with feeling justified to ask the husband to use a condom if he had an STI. Many associations were also observed in only few countries. Only very few directed edges were found.

Literacy was directly associated with several factors that should protect against acquiring HIV. For example, literate women started sexual relationships later, and literate people were less likely to have false beliefs about AIDS. This is promising, since literacy is one of the few variables in our analysis that can be directly improved by interventions. The overall proportion of literacy in sub-Saharan Africa has been slowly increasing over the last decades, from 63% in men and 42% in women in 1990 to 72% in men and 57% in women in 2016 [27]. The literacy rates among young men and women are considerably higher and likely to continue to rise.

Earlier studies have returned conflicting results about the positive association between literacy and HIV. For example, a study from Mozambique showed that literacy was associated with HIV knowledge among women living in rural settings [28]. In the Gambia, both the degree of literacy and level of education were positively associated with adherence to therapy among patients receiving ART [29]. In our study, illiteracy was also strongly associated with false beliefs about AIDS; both are likely linked to a low level of knowledge about HIV in general. For example, in Malawi, people who knew more about HIV prevention were more likely to reject, rather than endorse, myths and misconceptions about HIV transmission [30,31]. But some studies suggest the opposite. A recent review found that highly educated people were at higher risk of HIV. This

bioRxiv preprint doi: https://doi.org/10.1101/600510; this version posted November 6, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

could stem from a variety of causes, since literacy often correlates with other factors like wealth, type of employment and travel, which may be associated with riskier sexual behaviour. Literacy is also strongly associated with living in an urban setting. Literate people may be more likely to access HIV care and treatment, prolonging survival and thus raising prevalence. Finally, the distribution of HIV burden may depend on the overall wealth and situation in the country. One study showed that in poorer sub-Saharan African countries, wealthier individuals were more likely to be HIV positive, but in wealthier countries, poorer individuals were more likely to be HIV positive [32]. A study from Zimbabwe that used DHS data found that people who could read parts of a sentence had a higher risk of HIV infection than those who could not read at all, or who could read complete sentences [33].

We were slightly surprised by our results on the association between urbanity and variables with protective effects against HIV. HIV prevalence is generally higher in urban than rural settings [34]: for example, in the Democratic Republic of the Congo, distance to a city was negatively associated with HIV [35]. There are several possible explanations for these contradictory findings. Urban residents are likely to be more connected to other people, which may increase the risk of infection, despite other protective factors. HIV may also affect sub-populations differently in different settings; for example, in Kenya the poorest people are more likely to have HIV in urban than rural settings [36]. In Uganda, rural residence was associated with an increase in risky sexual behaviour in people who were on long-term ART [37]. Interestingly, the associations of rural location with less HIV testing and more false beliefs about AIDS were not found in women in high HIV prevalence countries. This may reflect the major role of antenatal care (ANC). The vast majority of women attend at least one ANC visit even in rural settings [38], and in settings with generalized HIV epidemic, HIV testing and counselling is part of routine care.

In most countries in sub-Saharan Africa, HIV testing forms the largest gap in the cascade of care. This gap must be closed to meet the “90-90-90” global treatment target and eventually eradicate the epidemic [39]. We identified many associations between HIV testing uptake and other factors. For women, the strongest associations were with being married and with believing it justified for a woman to ask her husband with an STI to use a condom. These two variables may

bioRxiv preprint doi: https://doi.org/10.1101/600510; this version posted November 6, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

also be directly linked: a study from Zambia showed that married women who knew that their husbands were either sexually promiscuous or had a sexually transmitted infection were more likely to feel justified in refusing sex or requesting he use condoms [40]. Stigma, fear and discrimination have been identified as barriers for utilizing HIV services. For men, literacy was strongly associated with HIV testing. Although the association may not be causal and men are overall more likely to be literate than women, this shows that men should not be neglected in literacy interventions.

Strengths and limitations

Our study analysed the associations between socio-behavioural variables in most sub-Saharan African countries. Other epidemiological studies have used Bayesian networks [41–44] but we believe we are the first to investigate detailed associations between variables related to the spread of HIV in sub-Saharan Africa. Our study was also strengthened by our use of a very large data set from the Demographic Health Surveys of almost 30 countries, which allowed us to identify commonalities and differences between countries, and between men and women.

Our study also had several limitations. The variables were pre-selected by our research team, based on the prior knowledge and some practical constraints. Although we assumed there were no hidden variables in our analysis, we had to exclude potentially relevant variables because they were too heterogeneous or data were missing. Religion was associated with certain factors in many countries, but the diversity of religions prevented us from dichotomizing and including this variable. Other variables like male circumcision, ever paying for sex, or using condoms during the last paid intercourse were excluded because they were not collected in some countries or too many values were missing. For example, we could not include HIV status in the main analysis because too little data on HIV-positive men was available. DHS did not collect data on some factors related to geographic variability of HIV and AIDS in sub-Saharan Africa, including sex work, tuberculosis, and use of injected drugs. We had to dichotomize all variables to compute the CPDAGs with the available data, even though for some variables it would have been better to include more categories. For example, for marital status, widowed and divorced were categorized as married. There may also be a risk of differential misclassification in the data: for example, literate individuals with higher education may be more reluctant to admit accepting

bioRxiv preprint doi: https://doi.org/10.1101/600510; this version posted November 6, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

domestic violence or refusing the use of condom. Finally, we only identified a small number of directed edges. It is possible that most associations were not causal, or it may simply demonstrate how hard it is to prove causal relationships with routine data.

Conclusions

The Bayesian network approach is a valuable technique for studying the structure of associations between variables from surveys and other data sources. Our findings underline the importance of understanding the role of socio-behavioural and contextual factors in the spread of complex diseases like HIV. The associations with variables related to HIV vary across countries, and it is essential to understand and study further the specific situation in each country to identify possible intervention targets. Literacy and urbanity were associated with several factors that are essential for HIV acquisition. Since literacy is one of the few variables that can be improved by interventions, this seems the most promising intervention target.

bioRxiv preprint doi: https://doi.org/10.1101/600510; this version posted November 6, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Competing interests

None declared.

Authors' contributions

ZB, OK and MM designed the study. ZB and OK reviewed and selected the variables to be included in the analysis. ZB wrote the code and performed the statistical analysis with together with JE, SM, AM and MM. SB, ZB, JE and OK reviewed the relevant literature. ZB and JE wrote a first draft of the manuscript, which was revised by OK. All authors contributed to the interpretation of the results and the final version of the manuscript.

Acknowledgements

We thank Gilles Kratzer and Reinhard Furrer for helpful discussions and Kali Tal for carefully editing the paper.

Funding: This project was funded by the Swiss National Science Foundation (grant no. 163878).

Additional files

Additional file 1: Results of the Bayesian network analysis for each country

Text file (pdf).

References

1. UNAIDS AIDSInfo [Internet]. Geneva, Switzerland: UNAIDS; 2019 [cited 2019 Nov 1]. Available from: http://aidsinfo.unaids.org.

bioRxiv preprint doi: https://doi.org/10.1101/600510; this version posted November 6, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

2. UNAIDS. Miles To Go: Closing Gaps, Breaking Barriers, Righting Injustices. Geneva, Switzerland: UNAIDS; 2018 [cited 2018 Dec 7]. Available from: http://www.unaids.org/sites/default/files/media_asset/miles-to-go_en.pdf.

3. Agyei-Mensah S. The HIV/AIDS epidemic in sub-Saharan Africa: Homogeneity or heterogeneity? Nor Geogr Tidsskr - Nor J Geogr. 2005;59(1):14–25.

4. Kharsany ABM, Karim QA. HIV Infection and AIDS in Sub-Saharan Africa: Current Status, Challenges and Opportunities. Open AIDS J. 2016;10(1):34–48.

5. Asamoah-Odei E, Calleja JMG, Boerma JT. HIV prevalence and trends in sub-Saharan Africa: no decline and large subregional differences. Lancet. 2004;364(9428):35–40.

6. ICF. Demographic and Health Surveys. Rockville (MD): ICF; 2010 [cited 2019 Mar 14]. Available from: https://dhsprogram.com.

7. Cuadros DF, Sartorius B, Hall C, Akullian A, Bärnighausen T, Tanser F. Capturing the spatial variability of HIV epidemics in South Africa and Tanzania using routine healthcare facility data. Int J Health Geogr. 2018;17(1):27.

8. Mondal MNI, Shitan M. Factors affecting the HIV/AIDS epidemic: an ecological analysis of global data. Afr Health Sci. 2013;13(2):301–10.

9. Tsafack Temah C. What Drives HIV/AIDS Epidemic in Sub-Saharan Africa? Rev Econ Dev. 2009;17(5):41–70.

10. Magadi MA. Understanding the gender disparity in HIV infection across countries in sub-Saharan Africa: evidence from the Demographic and Health Surveys. Sociol Health Illn. 2011;33(4):522–39.

11. Wamoyi J, Wight D, Plummer M, Mshana GH, Ross D. Transactional sex amongst young people in rural northern Tanzania: an ethnography of young women's motivations and negotiation. Reprod Health. 2010; 7(1):2.

12. Singh RK, Patra S. What Factors are Responsible for Higher Prevalence of HIV Infection among Urban Women than Rural Women in Tanzania? Ethiop J Health Sci. 2015;25(4):321–8.

bioRxiv preprint doi: https://doi.org/10.1101/600510; this version posted November 6, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

13. Hajizadeh M, Sia D, Heymann SJ, Nandi A. Socioeconomic inequalities in HIV/AIDS prevalence in sub-Saharan African countries: evidence from the Demographic Health Surveys. Int J Equity Health. 2014;13(1):18.

14. Johnson K, Way A. Risk Factors for HIV Infection in a National Adult Population: Evidence From the 2003 Kenya Demographic and Health Survey. J Acquir Immune Defic Syndr. 2006;42(5):627–36.

15. Dunkle KL, Jewkes RK, Brown HC, Gray GE, McIntryre JA, Harlow SD. Gender-based violence, relationship power, and risk of HIV infection in women attending antenatal clinics in South Africa. Lancet. 2004;363(9419):1415–21.

16. Korb KB, Nicholson AE. Bayesian Artificial Intelligence. London, United Kingdom: Chapman & Hall/CRC.

17. ICF. Guide to DHS Statistics. Rockville (MD): ICF; 2018 [cited 2018 Nov 12]. Available from: https://dhsprogram.com/data/Guide-to-DHS-Statistics/.

18. Verma T, Pearl J. Equivalence and Synthesis of Causal Models. In: Proceedings of the Sixth Conference on Uncertainty in Artificial Intelligence (UAI 1990); 1990 July 27-29; Cambridge, MA. Amsterdam, The Netherlands: Elsevier; 1990:255-268.

19. Andersson SA, Madigan D, Perlman MD. A characterization of Markov equivalence classes for acyclic graphs. Ann Statist. 1997;25(2):505-41.

20. Drton M, Maathuis MH. Structure learning in graphical modelling. Annual Review of Statistics and Its Application. 2017;4:365-93.

21. Heinze-Deml H, Maathuis MH, Meinshausen N. Causal structure learning. Annual Review of Statistics and Its Application. 2018;5:371-91.

22. Nandy P, Hauser A, Maathuis MH. High-dimensional consistency in score-based and hybrid structure learning. Ann Satist. 2018;46(6A):3151-83.

23. Kalisch M, Mächler M, Colombo D, Maathuis MH, Bühlmann P. Causal inference using graphical models with the R package pcalg. J Stat Softw. 2012;47(11):1-26.

bioRxiv preprint doi: https://doi.org/10.1101/600510; this version posted November 6, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

24. Tsamardinos I, Aliferis CF, Statnikov A. Time and sample efficient discovery of Markov blankets and direct causal relations. In: Proceedings of the Ninth ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD '03), August 24-27, 2003, Washington DC, United States. ACM Press; 2003:673.

25. R Core Team. R: A language and environment for statistical computing [software]. 2013 [cited 2019 Mar 14]. Available from: http://www.r-project.org/.

26. Day NE, Byar DP. Testing Hypotheses in Case-Control Studies-Equivalence of Mantel-Haenszel Statistics and Logit Score Tests. Biometrics. 1979;35(3):623.

27. World Bank Open Data [Internet]. Washington (DC): World Bank; 2019 [cited 2019 Mar 14]. Available from: https://data.worldbank.org.

28. Ciampa PJ, Vaz LME, Blevins M, Sidat M, Rothman RL, Varmund SH, et al. The Association among Literacy, Numeracy, HIV Knowledge and Health-Seeking Behavior: A Population-Based Survey of Women in Rural Mozambique. PLoS One. 2012;7(6):e39391.

29. Hegazi A, Bailey RL, Ahadzie B, Alabi A, Peterson K. Literacy, education and adherence to antiretroviral therapy in The Gambia. AIDS Care. 2010;22(11):1340–5.

30. Sano Y, Antabe R, Atuoye KN, Hussey LK, Bayne J, Galaa SZ, et al. Persistent misconceptions about HIV transmission among males and females in Malawi. BMC Int Health Hum Rights. 2016;16(1):16.

31. Behrman JA. The effect of increased primary schooling on adult women's HIV status in Malawi and Uganda: Universal Primary Education as a natural experiment. Soc Sci Med. 2015;127:108–15.

32. Fox AM. The HIV-poverty thesis re-examined: Poverty, wealth or inequality as a social determinant of HIV infection in sub-Saharan Africa? J Biosoc Sci. 2012;44(4):459–80.

33. Chinomona A, Mwambi HG. Estimating HIV prevalence in Zimbabwe using population-based survey data. PLoS One. 2015;10(12):e0140896.

bioRxiv preprint doi: https://doi.org/10.1101/600510; this version posted November 6, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

34. Global Health Observatory data repository [Internet]. Geneva, Switzerland: World Health Organization; 2014 [cited 2019 Mar 14]. Available from: http://apps.who.int/gho/data/node.main.n247?lang=en.

35. Messina JP, Emch M, Muwonga J, Mwandagalirwa K, Edidi SB, Mama N, et al. Spatial and socio-behavioral patterns of HIV prevalence in the Democratic Republic of Congo. Soc Sci Med. 2010;71(8):1428–35.

36. Dodoo FNA, Zulu EM, Ezec AC. Urban-rural differences in the socioeconomic deprivation-sexual behavior link in Kenya. Soc Sci Med. 2007;64(5):1019–31.

37. Okoboi S, Castelnuovo B, Moore DM, Musaazi J, Kambugu A, Birungi J, et al. Risky sexual behavior among patients on long-term antiretroviral therapy: a prospective cohort study in urban and rural Uganda. AIDS Res Ther. 2018;15(1):15.

38. UNICEF Data: Monitoring the situation of children and women [Internet]. New York (NY): UNICEF; 2019 [cited 2019 Nov 1]. Available from: https://data.unicef.org/topic/maternal-health/antenatal-care/

39. Estill J, Marsh K, Autenrieth C, Ford N. How to achieve the 90-90-90 target by 2020 in sub-Saharan Africa? Trop Med Int Health. 2018;23(11):1223–30.

40. Amoyaw JA, Kuuire VZ, Boateng GO, Asare-Bediako Y, Ung M. Conundrum of Sexual Decision Making in Marital Relationships: Safer-Sex Knowledge, Behavior, and Attitudes of Married Women in Zambia. J Sex Res. 2015;52(8):868–77.

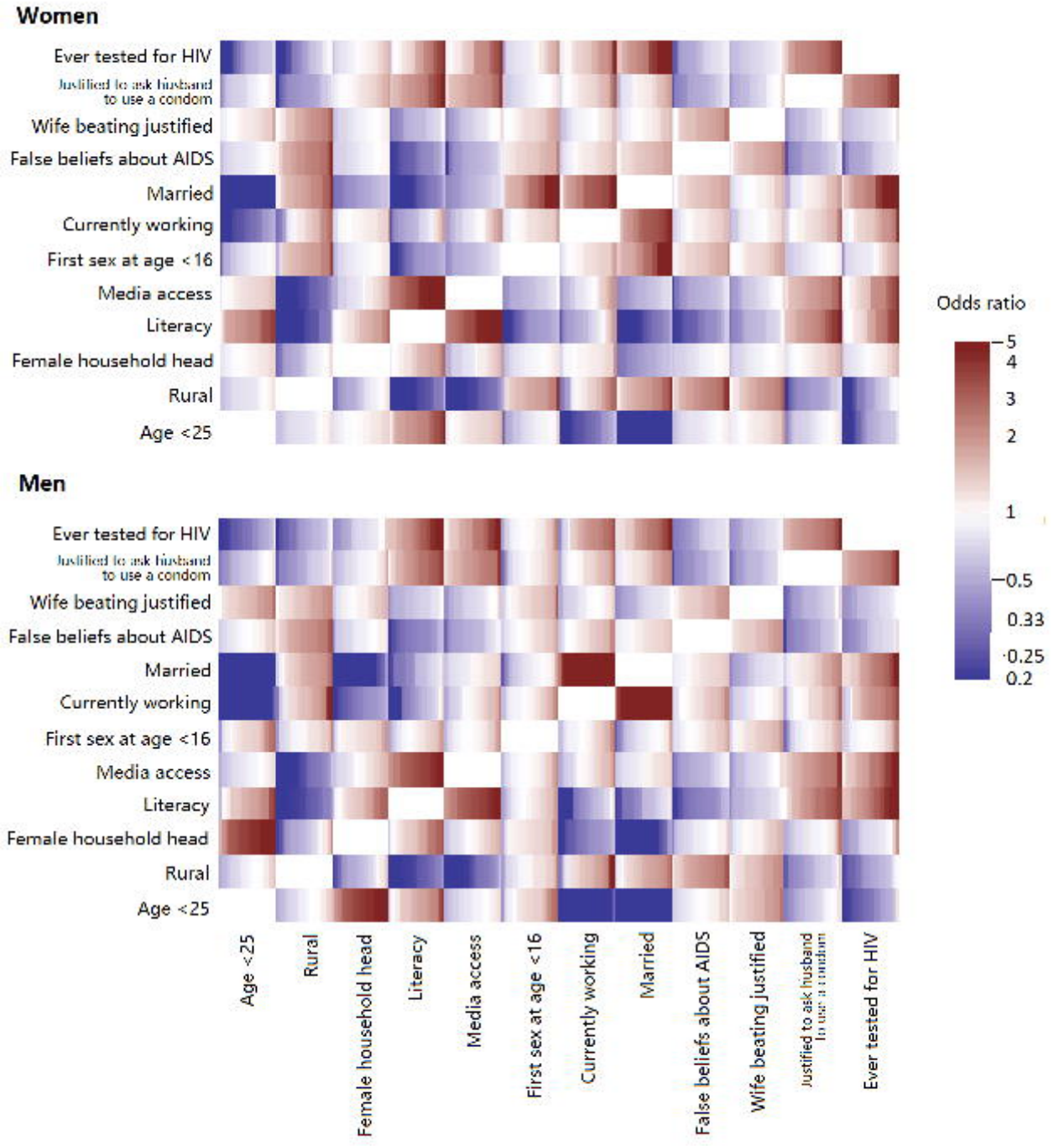
41. Pittavino M, Dreyfus A, Heuer C, Benschop J, Wilson P, Collins-Emerson J, et al. Comparison between generalized linear modelling and additive Bayesian network; identification of factors associated with the incidence of antibodies against Leptospira interrogans sv Pomona in meat workers in New Zealand. Acta Trop. 2017;173:191–9.

42. Lewis FI. Bayesian networks as a tool for epidemiological systems analysis. In: AIP Conference Proceedings; Proceedings of the 9th International Conference on Mathematical Problems in Engineering, Aerospace and Sciences (ICNPAA 2012); 2012 Jul 10-14; Vienna, Austria. Melville (NY): AIP Publishing LLC; 2012:610-7.

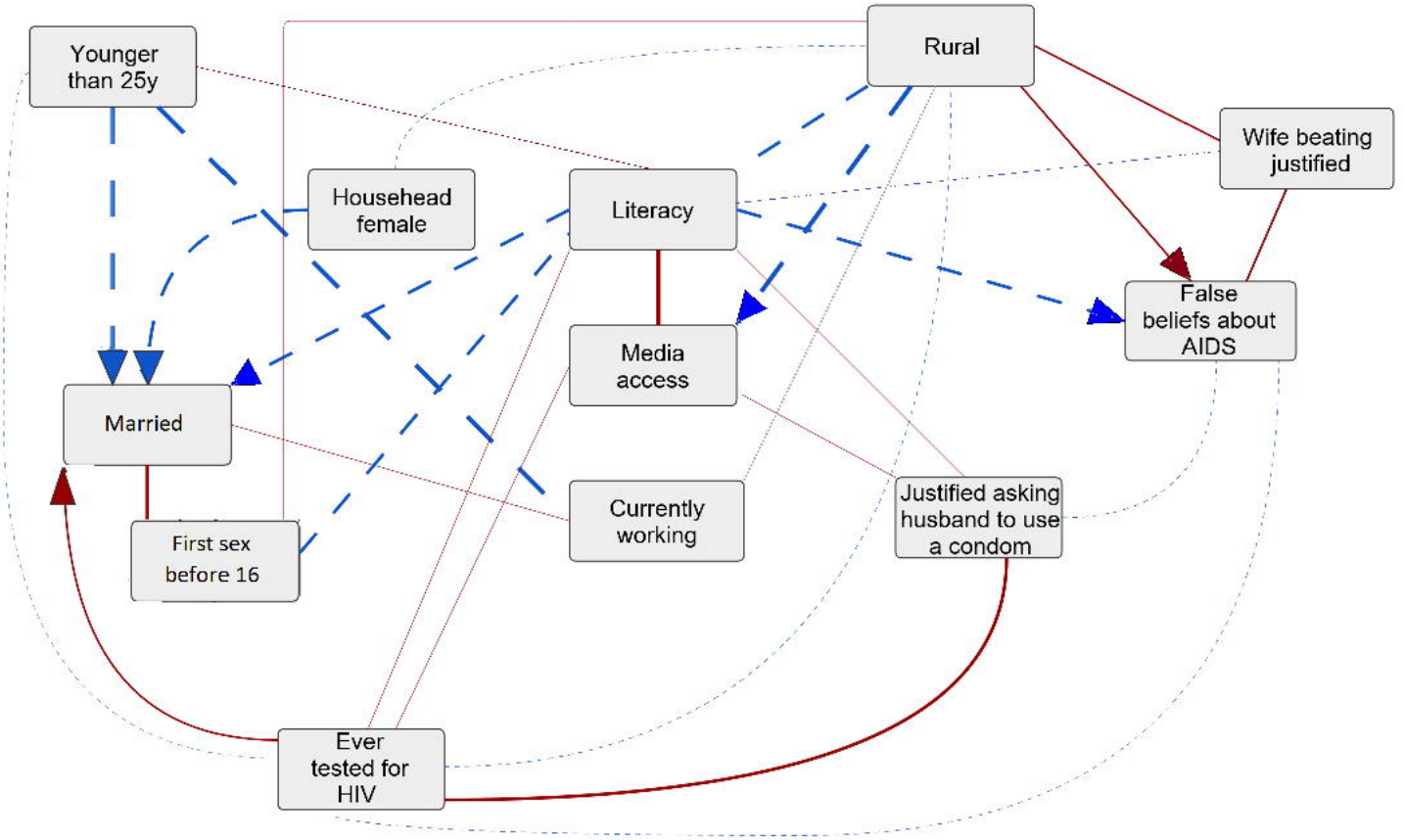
bioRxiv preprint doi: https://doi.org/10.1101/600510; this version posted November 6, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

43. McCormick BJJ, Sanchez-Vazquez MJ, Lewis FI. Using Bayesian networks to explore the role of weather as a potential determinant of disease in pigs. Prev Vet Med. 2013;110(1):54–63.

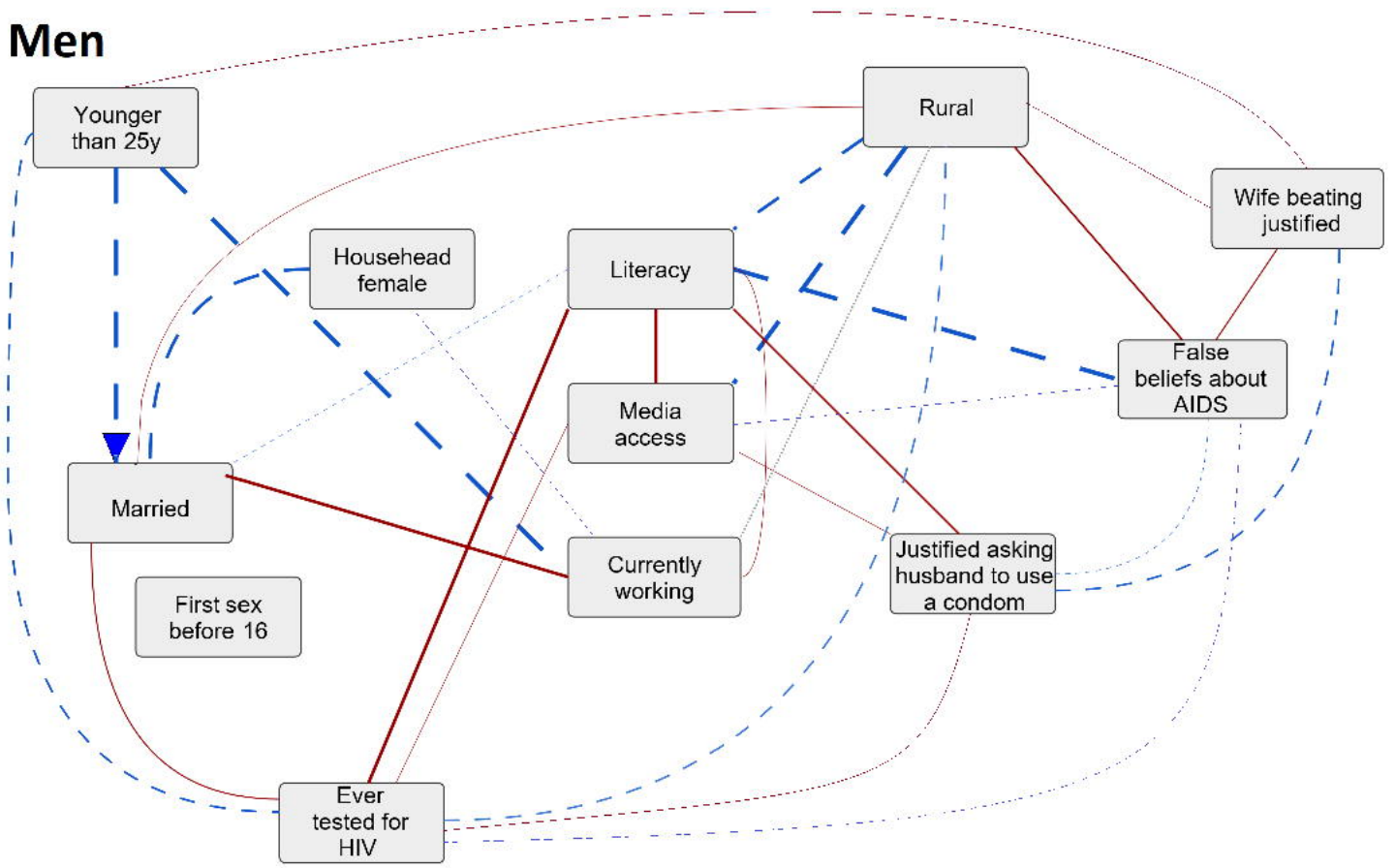
44. Kalisch M, Fellinghauer BA, Grill E, Maathuis MH, Mansmann U, Bühlmann P, et al. Understanding human functioning using graphical models. BMC Med Res Methodol. 2010;10:14.

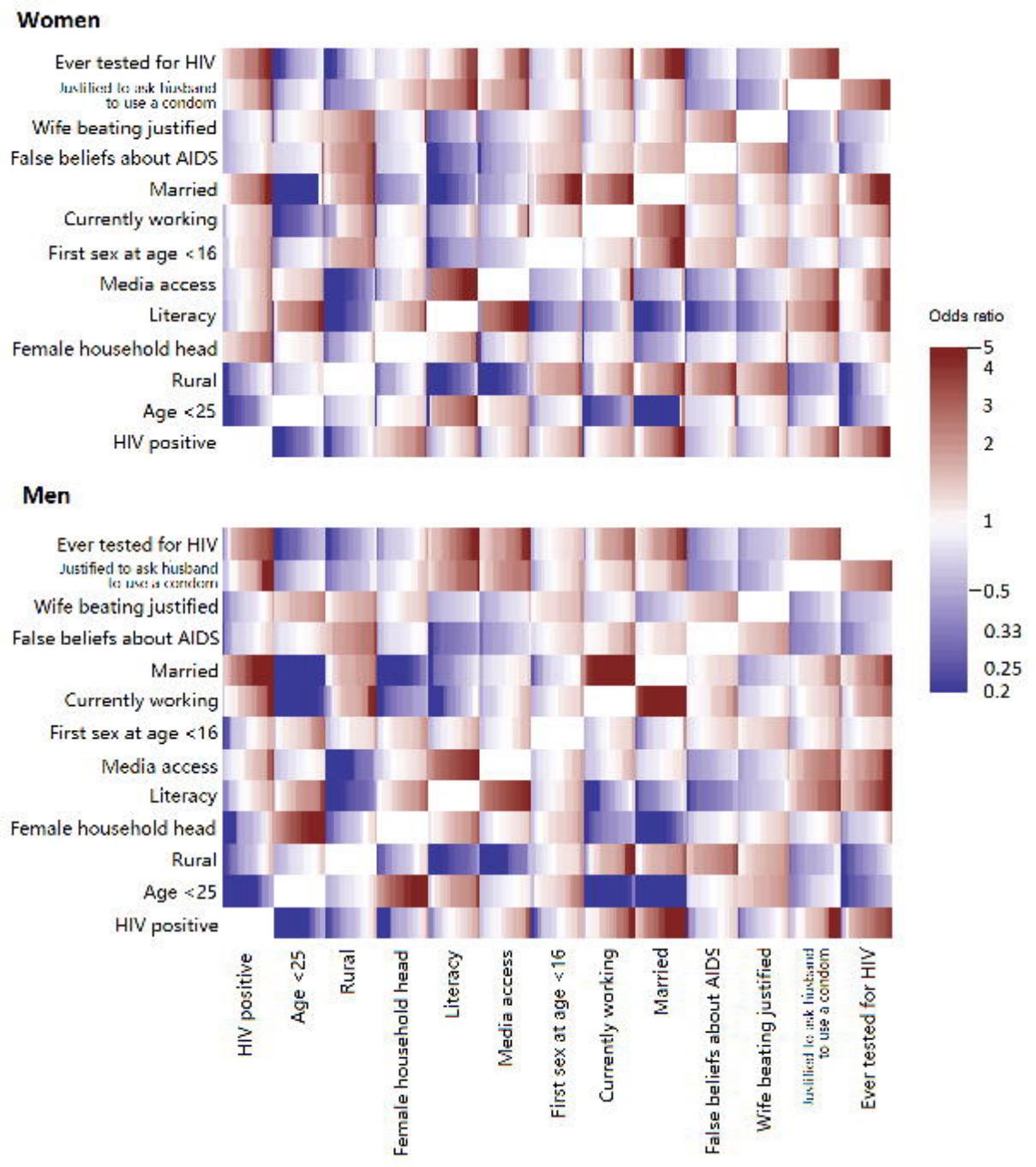

Women
Ever tested for HIV
Justified to ask husband to use a condom
Wife beating justified
False beliefs about AIDS
Married
Currently working
First sex at age <16
Media access
Literacy
Female household head
Rural
Age <25
Men
Ever tested for HIV
Justified to ask husband to use a condom
Wife beating justified
False beliefs about AIDS
Married
Currently working
First sex at age <16
Media access
Literacy
Female household head
Rural
Age <25
Age <25
Rural
Female household head
Literacy
Media access
First sex at age <16
Currently working
Married
False beliefs about AIDS
Wife beating justified
Ever tested for HIV
Odds ratio
5
4
3
2
1
0.5
0.33
0.25
0.2

Women

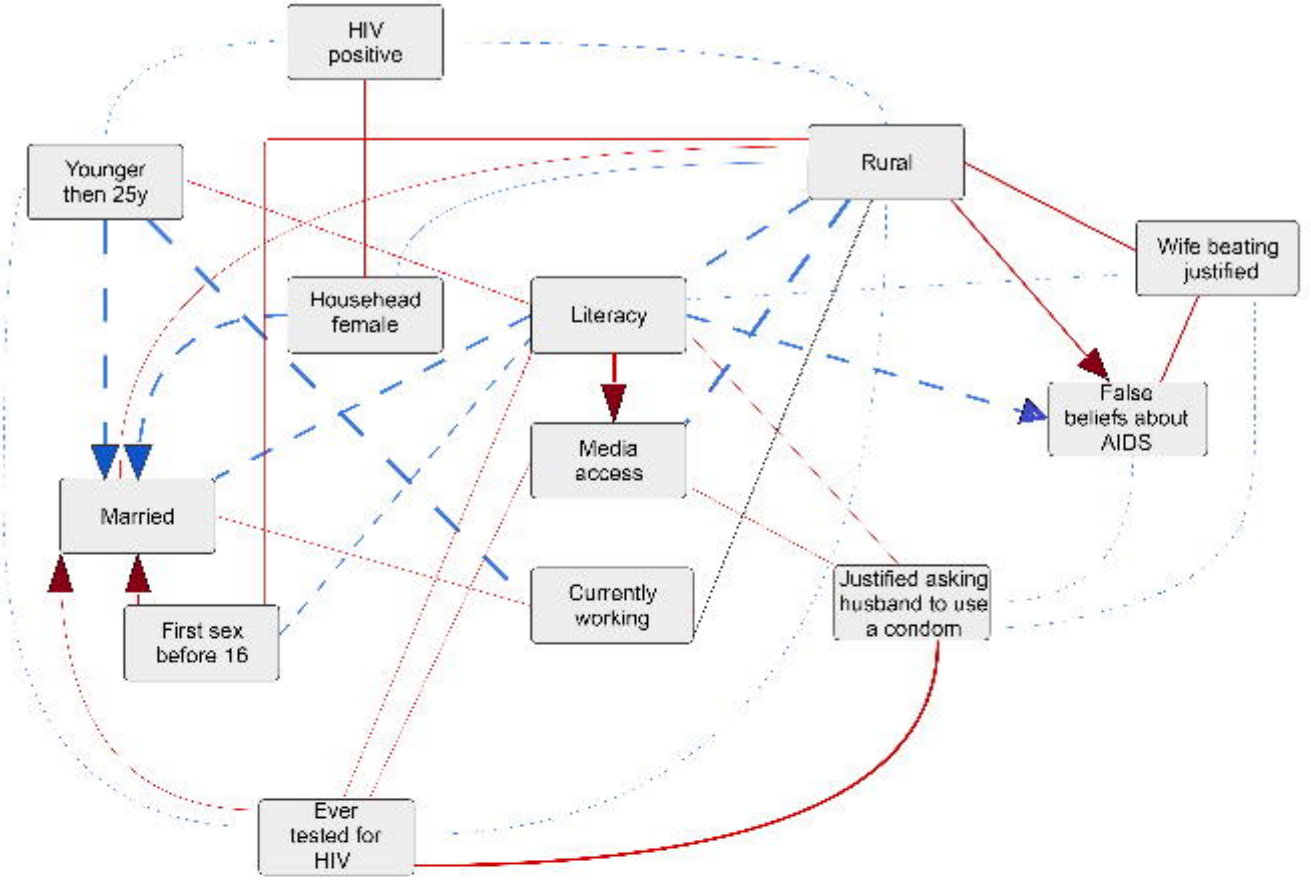

Younger than 25y
Rural
Wife beating justified
Househead female
Literacy
False beliefs about AIDS
Media access
Married
Currently working
Justified asking husband to use a condom
First sex before 16
Ever tested for HIV

Men

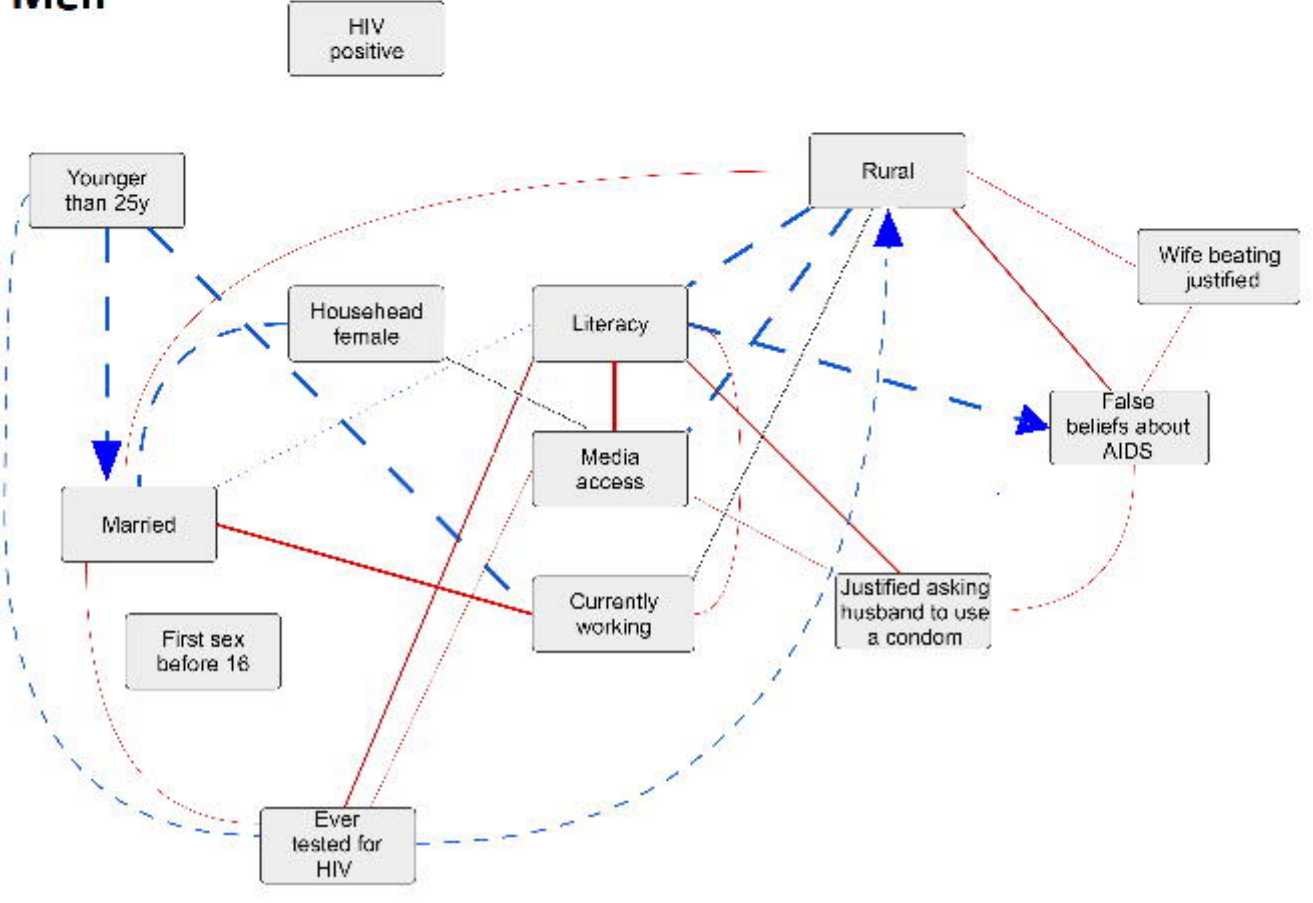

Younger than 25y
Rural
Wife beating justified
Househead female
Literacy
False beliefs about AIDS
Media access
Married
Currently working
Justified asking husband to use a condom
First sex before 16
Ever tested for HIV


Women
Ever tested for HIV
Justified to ask husband to use a condom
Wife beating justified
False beliefs about AIDS
Married
Currently working
First sex at age <16
Media access
Literacy
Female household head
Rural
Age <25
HIV positive
Men
Ever tested for HIV
Justified to ask husband to use a condom
Wife beating justified
False beliefs about AIDS
Married
Currently working
First sex at age <16
Media access
Literacy
Female household head
Rural
Age <25
HIV positive
HIV positive
Age <25
Rural
Female household head
Literacy
Media access
First sex at age <16
Currently working
Married
False beliefs about AIDS
Wife beating justified
Justified to ask husband to use a condom
Ever tested for HIV
Odds ratio
5
4
3
2
1
0.5
0.33
0.25
0.2

Women


HIV positive
Younger then 25y
Rural
Wife beating justified
Househead female
Literacy
False beliefs about AIDS
Media access
Married
Currently working
Justified asking husband to use a condom
First sex before 16
Ever tested for HIV

Men


HIV positive
Younger than 25y
Rural
Wife beating justified
Househead female
Literacy
False beliefs about AIDS
Media access
Married
Currently working
Justified asking husband to use a condom
First sex before 16
Ever tested for HIV