Prevalence of HIV among women in Malawi: Identifying the most-at-risk groups for targeted and cost-effective interventions

Journal:	*BMJ Open*
Manuscript ID:	bmjopen-2012-002459
Article Type:	Research
Date Submitted by the Author:	06-Dec-2012
Complete List of Authors:	Emina, Jacques; University of Kinshasa, ; CEPS/ INSTEAD, Madise, Nyovani; University of Southampton, Faculty of Social and Human Sciences Kuepie, Mathias; CEPS/INSTEAD, AFRILUX Zulu, Eliya; African Institute for Development Policy, Yazoume , Ye ; ICF International,
Primary Subject Heading:	HIV/AIDS
Secondary Subject Heading:	Epidemiology, Health policy, Sociology, Public health
Keywords:	HIV & AIDS < INFECTIOUS DISEASES, Epidemiology < INFECTIOUS DISEASES, Public health < INFECTIOUS DISEASES

Prevalence of HIV among women in Malawi: Identifying the most-at-risk groups for targeted and cost-effective interventions

Jacques B. O. Emina [1,2], Nyovani Madise [3], Mathias Kuepie [2], Eliya M. Zulu [4], Yazoume Ye [5]

1 Department of Population and Development Studies, University of Kinshasa, P.O.BOX 176 Kinshasa XI, Democratic Republic of Congo

2 CEPS/INSTEAD 3, avenue de la Fonte, L-4364 Esch-sur-Alzette Luxembourg

3 Faculty of Social and Human Sciences, Room 4103 Building 58, University of Southampton, Southampton, SO17 1BJ.

4 African Institute for Development Policy (AFIDEP), P. O. Box 14688-00800, Westlands, Nairobi, Kenya.

5 ICF International 11785 Beltsville Drive, Suite 300 Calverton, MD 20705, USA.

Correspondence to: Dr Jacques B.O. Emina; jacques.emina@gmail.com

Abstract

Objectives: To identify HIV socioeconomic predictors as well as identify the most-at-risk groups of women in Malawi.

Design: consecutive cross-sectional surveys

Setting: Malawi

Participants: The study relies on a sample of 8,596 age 15-49 from the 2004 and 2010 Malawi Health and Demographic Surveys (MDHS).

I**nterventions**: N/A

Primary and secondary outcome measures: Whether individual is HIV positive.

Results: Findings from Pearson Chi-square and Chi-square Automatic Interaction Detector (CHAID) analyses revealed that marital status is the most significant predictor of HIV. Women who are no longer in union and living in rich households constitute the most-at-risk population; whereas the less-at-risk group includes nulliparous never married women living in the rural areas.

Conclusion: In the light of these findings, this study recommends: (1) design and implementation of targeted interventions taking into account HIV prevalence and the demographic size of different groups at risk groups. With 45 percent of the study population, among whom the HIV prevalence is estimated at 17 percent on average, couples (males and women in union) living in the Southern region and those living in the urban areas of the Central and the Northern should be the first targets of HIV interventions in Malawi; (2) reinforcement of integration of family planning and HIV/AIDS services through community health workers, households based campaign, reproductive health services and reproductive health courses at school.

Keywords: HIV/AIDS, Malawi, CHAID methods, decision analysis, most-at-risk groups, targeted interventions.

Article summary

1. Article focus

- Targeted interventions and evidence based prevention programmes have been advocated as cost-effective strategy to combat HIV/AIDS.
- whereas in countries with concentrated AIDS epidemics (Latina America, East Asia and Eastern Europe), the most-at-risk populations including commercial sex workers (CSWs), long distance truck drivers, men who have sex with men, and unmarried youth account for a large proportion of new infections, in countries with high prevalence, they account only for a smaller share of new infections.
- Who are the most-at-risk populations regarding HIV prevalence in Malawi? With HIV prevalence of about 13 percent among women of reproductive age, HIV/AIDS constitutes a drain on the labor force and government expenditures in Malawi.

2. Key messages

- We use data from the Malawi 2004 and 2010 Demographic and Health Surveys to profile HIV most-at-risk groups of women in Malawi where about 14 percent of women are HIV positive.
- Our findings revealed that richest and formerly in union women are the most-at-risk population.
- We suggested targeted interventions considering the groups HIV prevalence and size. With 45 percent of the study population, among whom the HIV prevalence is estimated at 17 percent on average, couples (males and women in union) living in the Southern region and those living in the urban areas of the Central and the Northern should be the first targets of HIV interventions in Malawi.

Strengths and limitations of this study

- Our study seems the first in Malawi that has attempted to profile HIV most-at-risk groups of women in Malawi. The most-at-risk population refers to a combination of several factors because socioeconomics factors associated with HIV are not mutually exclusive.
- The major strength is the use of the Chi-square Automatic Interaction Detector (CHAID) to identify HIV predictors and the most at risk groups of women for intervention. CHAID uses regression and classification algorithms and offers a non-algebraic method for partitioning data that lends itself to graphical displays.
- The major limitation of this study is the cross- sectional nature of the data, which does not permit one to draw causal association between HIV status and the associated factors. For instance, we do not know whether HIV infection has occurred before, during or after the union.

Introduction

In 2000, the United Nations' Millennium summit identified the reduction of HIV prevalence as one of the eight fundamental goals for furthering human development. Though global HIV/ AIDS incidence is declining, HIV/AIDS has remained the leading cause of death in women of reproductive age in low- and middle-income countries, particularly in sub-Sahara Africa (SSA) [1]. The gap between the state of HIV/AIDS and the UNAIDS goals of three zero (zero new HIV infections, zero discrimination, and zero AIDS related deaths) remains important. With barely two years remaining to the end-date of the MDG target, HIV/AIDS remains a long-term global challenge [1].

Based on the current costs of HIV treatment (US $ 4,707 over lifetime) (International HIV / AIDS Alliance, 2010), targeted interventions and evidence based prevention programmes have been advocated as cost-effective strategy to combat HIV/AIDS. Such a strategy reduces levels of vulnerability and risk as well as allowing HIV interventions to optimize coverage, reducing costs and lowering the number of new infections [2].

Despite growing literature in health and social sciences on factors associated with HIV/AIDS during the last three decades, important questions are still poorly answered in the context of Malawi [3, 4, 5, 6, 7, 8]. For instance, who are the most-at-risk populations regarding HIV prevalence in Malawi? Indeed, whereas in countries with concentrated AIDS epidemics (Latina America, East Asia and Eastern Europe), the most-at-risk populations including commercial sex workers (CSWs), long distance truck drivers, men who have sex with men, and unmarried youth [9,4,10] account for a large proportion of new infections, in countries with high prevalence, they account only for a smaller share of new infections [10].

Against this background, this study aims to identify HIV socioeconomic predictors as well as identify the most-at-risk groups of women in Malawi. With HIV prevalence of about 13 percent among women of reproductive age [11], HIV/AIDS constitutes a drain on the labor force and government expenditures in Malawi. Most-at-risk populations refer to a combination of several factors because socioeconomics factors associated with HIV are not mutually exclusive.

Data and Methods

Study setting

The Republic of Malawi is a landlocked country in southeast Africa. Malawi is over 118,000 km^2 with an estimated population of about 16 millions [11]. Its capital is Lilongwe, which is also Malawi's largest city; the second largest is Blantyre and the third is Mzuzu.

Malawi is among the world's least-developed countries. The economy is heavily based on agriculture, with a largely rural population. The country GNI PPP per Capita is estimated at $860 while the world average is estimated at $10,780 [11]. Ninety-one percent of Malawians live below 2 dollars (US) per day. The country's Human Development Index is estimated at 0.400, which gives the country a rank of 171 out of 187 countries with comparable data [12].

Malawi has a low life expectancy (53 years) and high infant mortality (66 deaths per 1,000 live births) compared to the world' average (41 deaths per 1,000 live births). There is a high prevalence of HIV/AIDS, especially among women, among whom about 14 percent are HIV positive [11]. Malawi is divided into 28 districts within three regions: Southern, Central and Northern regions.

Data sources

This study uses data from two independently pooled cross-section data, the 2004 and 2010 Malawi Health and Demographic Surveys (MDHS). The inclusion of HIV testing in the 2004 and 2010 MDHS offers the opportunity to identify socioeconomic profile of women age 15-49 who were living with HIV. Participation in HIV testing was voluntary. To ensure confidentiality, case numbers (and not names) were used in linking the HIV test results to individual and household characteristics.

In both surveys, a subsample of one-third of the households was selected to conduct HIV testing for eligible women age 15-49. Overall, 4,071 women age 15-49 were identified as eligible for testing in the 2004 MDHS. Of these, testing was successfully conducted on 2,686 women, resulting in a response rate of 70 percent for women. Ninety percent of all 2010 MDHS women who were eligible (8,174) for testing were interviewed and consented to HIV testing. Details on the sample design are provided elsewhere [13,14].

Variables

The dependent variable for this analysis is HIV status, characterized as positive or negative depending on blood tests carried out during the surveys. The independent variables encompass 12 main variables grouped into two major types:

1. Demographic and reproductive behavior variables: age, age at first sex, marital status, age at first birth, number of children ever born, Experience in premarital childbearing, and relationship to the head of household.
2. Socioeconomic and contextual variables: religion, region of residence, place of residence, education, and household wealth index.

The choice of these variables is guided by the literature on factors associated with HIV in sub-Saharan Africa [3-8].

Statistical analyses

Statistical analyses rely on Pearson Chi-square and Chi-square Automatic Interaction Detector (CHAID) using SPSS version 16. We used data weights to take into account the complexity of the DHS design.

We performed Pearson chi-square to identify associations between the HIV status (positive, negative) and demographic and reproductive behavior variables as well as socioeconomic and contextual variables.

We used CHAID to identify HIV predictors and the most at risk groups of women for intervention [15]. This is a nonparametric technique that makes no distributional assumptions and on outliers, collinearities, heteroskedasticity, or distributional error structures. The dependent variable and predictor variables can be nominal (categorical), ordinal (ordered categories ranked from small to large), or interval (a "scale").

CHAID uses regression and classification algorithms and offers a non-algebraic method for partitioning data that lends itself to graphical displays. The method is a sequential fitting algorithm and its statistical tests are sequential with later effects being dependent upon earlier ones, and not simultaneous as would be the case in a regression model or analysis of variance where all effects are fit simultaneously. At each step, CHAID chooses the independent (predictor) variable that has the strongest interaction with the HIV status (dependent variable).

CHAID analysis goes one step further, and identifies as the important elements of HIV prevalence those variables that most differentiate the likelihood to be HIV positive. The CHAID procedure begins by finding variables that have a significant association with HIV status. It then assesses the category groupings, or interval breaks to pick the most significant combination of variables. The variable having the strongest association with HIV status becomes the first branch in a tree with a leaf for each category that is significantly different relative to be HIV positive. The process is repeated to find the predictor variable on each leaf most significantly related to HIV status, until no significant predictors remain.

The developed model is a classification tree (or data partitioning tree) that shows how major "types" formed from the independent (predictor or splitter) variables differentially predict a criterion or dependent variable. The method permits also identification of women who are likely to be members of a particular group (Segmentation), and assign cases into one of several categories, such as high-, medium-, and low-risk groups (stratification). Selecting a useful subset of predictors from a large set of variables for use in building a formal parametric model (Data reduction and variable screening); Identify relationships that pertain only to specific subgroups and specify these in a formal parametric model (Interaction identification); and recoding group predictor categories and continuous variables with minimal loss of information. Categories of each predictor are merged if they are not significantly different with respect to the dependent variable (Category merging and discretizing continuous variables).

Like other decision trees methods, CHAID's its output is highly visual and easy to interpret. It allows identifying homogeneous groups with high or low risk; and constructing rules for making predictions about individual cases.

However, CHAID needs large sample sizes to work effectively because it uses multi-way splits. Indeed, with small sample sizes the respondent groups can quickly become too small for reliable analysis.

Results

Sampling description

Table 1 presents the main characteristics of the population under study. Since the principal mode of HIV transmission in Malawi is heterosexual contact, our analyses

focus on women who ever had sexual intercourse. The distribution of the sample by age shows that more than half (56 percent) of the populations are age less than 30 years old. The average age of the sample is estimated at 29 years old. Women who are in union (i.e., currently married or living with a man) constitute about 80 percent. The proportion of women who have never been married is estimated at 14 percent. Regarding the relationship to the head of household, the majority of women are spouse (64 percent). Eighteen percent of the studied population are head of household.

Since the principal mode of HIV transmission in Malawi is heterosexual contact, our analyses focus on women who ever had sexual intercourse.

[Table 1, about here]

Table 1 also shows that the majority of women (82 percent) live in rural areas. By region, the majority of women live in the Central and Southern Regions, while 12 percent of women live in the Northern Region. Furthermore, although the majority of respondents have had some education, 20 percent of women never attended school, while 63 percent have attended only primary school.

Considering reproductive behavior, a large majority of women had their first sexual intercourse before 20 years (average 16 years old).

Factors associated with HIV prevalence: bivariate analysis

Table 2 describes HIV prevalence in Malawi by women's selected background characteristics. Overall, 14 percent of studied women are HIV positive. Except for religion, all independent variables are statistically associated with HIV infection status.

[Table 2, about here]

HIV infection prevalence was high (20 percent) among women aged 30-39 years. Regarding marital status, women who are no longer in union (widowed, divorced and separated) had significantly higher prevalence (30 percent) compared to those who have never been in a marital union (10 percent). HIV prevalence was high among heads of household.

Furthermore, while 25 percent of women in urban area were HIV positive, the prevalence was less than half (12 percent) compared to their counterparts from the rural areas. The HIV epidemic shows regional heterogeneity with a higher prevalence

(20 percent) observed in the Southern region. Women with secondary education had higher HIV prevalence compared to those who never attended school (18 percent versus 14 percent). Regarding the household wealth quintiles the prevalence of HIV infection is higher among the women from the highest quintiles. With reference to sexual and reproductive behavior, HIV prevalence was higher among women who had their first sexual intercourse before the 15th birthday and /or who have experienced a premarital childbearing.

HIV prevalence in Malawi risk groups: results from CHAID analysis

Table 3 shows some broad information about the specifications used to build the Chi-square Automatic Interaction Detector (CHAID) model and the resulting model (the model summary).

[Table 3, about here]

Eleven independent variables were specified, but only seven were included in the final model. The variables such as age at first sex, age at first birth and female education did not make a significant contribution to the model, so they were automatically dropped from the final model.

Overall, there are 27 nodes among which 16 terminal nodes. Parent nodes include at least 100 cases whereas child nodes account for 50 cases in minimum. The tree diagram shows that "Marital status" (Chi-square = 323.1, P-value<0.000) is the best predictor of HIV status among women in Malawi (Figure 1).

[Figure 1, about here]

The tree is split into 3 branches: (1) Node 1 including women in union; (2) Node 2 encompassing women formerly in union; and (3) Node 3 comprising never married women.

For women in union (married or living together), Figure 1a reveals that region of residence is the following best predictor (Chi-square = 132.21, P-value<0.000). Women living in the Southern region (Node 4) are opposed to those living in the Central and the Northern regions (Node 5).

[Figure 1a, about here]

The model includes one more predictor for women belonging to Node 4 (Southern region) and Node 5 (Central and Northern region). For women living in the Southern

region, the next best predictor is women's age at the survey (Chi-square=55.9, P-value <0.000). Women in union, living in the Southern region are divided into three groups considering their age at the survey:

- Women aged from 30 to 44 years old, among whom 23 percent are HIV-positive (Node 12);
- Women aged 25-29 years old (Node 14). HIV prevalence is estimated at 18 percent among age;
- Women aged from 15 to 24 and from 45 to 49, among which 11 percent are HIV positive.

Considering women in union and living in the Central and the Northern regions (Node 5), Place of residence is the next best predictor of HIV prevalence (Chi-square=86.6, P-value <0.000). Among these women prevalence of HIV varies between 6 percent in rural areas (Node 15) and 18 percent in urban area (Node 16).

Figure 1b shows that for women formerly in union (Node 2) including divorced, widowed and not living together, household wealth index is the second best predictor of HIV prevalence (Chi-square=92.8, P-value <0.000).

[Figure 1b, about here]

With reference to women living in poorer households (Node 7), wealth index is the only significant predictor of HIV prevalence. Proportion of HIV positive is estimated at 27.5 percent. Since there are no child nodes below it, this is considered a terminal node.

For women living in Poorest (Node 6) as well as in Richer and Middle (Node 8) the next best predictor of HIV prevalence is the region of residence. HIV prevalence is estimated at 21 percent among women formerly in union living in the poorest households of Southern province (Node 17), while proportion of HIV positive is estimated at 10 percent for poorest women living in the Central and Northern regions (Node 18).

Among women living in Middle and Richer households (Node 8), about 47 percent are HIV positive in the Southern region (Node 19), whereas HIV prevalence is estimated at 22 percent for those living in other regions (Node 20).

Figure 1c reveals also that age at first sex is the third best predictor (Chi-square=12.9, P-value <0.002) for HIV prevalence among women formerly in union and living in the richest households (Node 9). Among those women HIV prevalence is estimated at 72.7 percent for women who experienced their first sex before the age of 15 or from 25 years old (Node 22). The corresponding HIV prevalence is estimated at 45.3 percent for women who had their first sex experience from 15 to 24 years old (Node 21).

Considering women who have been never in union (Figure 1c), place of residence is the second best predictor (Chi-square=20.5, P-value<0.000).

[Figure 1c, about here]

Among women living in rural area (Node 10) the model includes one more variable "Whether the woman ever gave birth" (Chi-square=13.9, P-value<0.000). The region of residence (Chi-square=15.3, P-value<0.000) is the additional significant variable for never married women living in urban areas.

Among never married living in rural areas, HIV prevalence is estimated at 2.3 percent if the woman never gave birth (Node 24) and at 11.4 percent if the woman experienced childbearing (Node 23). For never married women living in urban area 23 percent are HIV positive in the Southern and the Northern regions (Node 25), while proportion of women living with HIV is estimated at 2.7 percent among never married women living in urban areas of the Central region (Node 26).

Findings also allow dividing the study population into four major groups taking into account interaction between the most statistically significant variables: very high, high, intermediate, and low HIV prevalence. Table 4 describes composition of each group.

- The first group (very high prevalence) represents 5.7 percent of the sample. HIV prevalence is estimated at 54.6 percent on average, varying between 45.3 percent and 73 percent. This category encompasses three subgroups: a) women in union disruption, living in a richest household and who had their first sex from 25 years old; b) women in union disruption and live in richer or middle households in the Southern region; c) women in union disruption living in richest household and who had their first sex between 15 and 24 years old.

- The second group (high prevalence) includes 21 percent of the sample. HIV prevalence is estimated at 23.3 percent (range 21 percent - 28 percent) in this group. This group comprises 5 sub-groups: Women in union disruption living in poorer households; Never married women who live in urban area of the Southern or Northern region; Women in union living in the Southern region and who are age 30-44 years old; women formerly in union who are living in richer or middle households of the Central or Northern region; and women in union disruption living in the poorest households of the Southern region.

- The third group (intermediate prevalence) comprises about 33 percent of women under study. HIV prevalence varies between 10 and 19 percents (13.8 percent on average) among women belonging to this group. This category could be divided into 5 subgroups: (a) women in union, living in urban areas of the Central or Northern region; b) women in union who are living in the Southern region and aged 25-29; c) never married women, living in rural area and who have experienced childbearing; d) women in union, living in the Southern region and age 15-24 or 45-49; e) women who are in union disruption, living in poorest households of the Central or the Northern region.

- The last group (low prevalence) include three subgroups: a) women in union, living in the rural areas of the Northern or of the Central region; b) women who never married living in urban areas of the Central region; c) nulliparous never married women living in rural area. HIV prevalence is estimated at 3.7 percent, ranges between 2.3 percent and 6 percent. This group accounts for 40 percent of the sample.

Discussions and Conclusion

This paper aimed to describe and profile HIV prevalence among women in Malawi. The study relied on data from the Malawi 2004 and 2004 DHS using Chi-square and CHAID techniques. CHAID offers a useful alternative to traditional logistic regression and allows identifying population subgroups that share similar characteristics [16].

Analyses suggested three keys findings that could be summarized as follows. First, consistent with previous studies [6, 17], findings from bivariate analysis and chi-square

test showed high HIV prevalence among women in union dissolution, among those living in wealthy households and/or among women living in urban areas, as well as region heterogeneity in HIV prevalence.

Second, results from CHAID models reported that marital status is the best predictor of HIV status among women in Malawi followed by the household wealth index. Women who are no longer in union (widowed and divorced or separated) and living in less poor households have significantly higher HIV prevalence. This probably because: (1) a rich husband or a male partner may have more access to transactional sex and other risk behaviors such as polygamy which may increase women's vulnerability to HIV; (2) wealthier HIV positive widowed may have better quality of life as well as better access to treatment and survive longer [18]. Furthermore, divorced and separated are more frequent among the most educated women with economic autonomy [19]. Their causes (polygyny and/or infidelity) as well as consequences (multiple sexual partnerships) are also factors associated with HIV prevalence [20,21].

Last, CHAID models depicted also different interactions between risk factors and profiled HIV risk groups in Malawi. For instance, whilst HIV prevalence is higher among women living in urban areas (25 percent) compared to those living in rural areas (12 percent), only 3 percent of never married women living in urban areas of the Central region are HIV positive compared to 11 percent observed among single mothers living in the rural areas. Likewise, while overall HIV prevalence is low among never married women (9 percent), CHAID results revealed a higher HIV prevalence (23 percent) among never married women who live in urban areas of the Southern or Northern region compared to women in union who reside in urban areas of the Central or Northern (18 percent) as well as to women in union dissolution who live in poorest households of the Central or Northern region (10 percent).

In the light of these findings, it is noteworthy that to reduce number of new infection, interventions should be targeted and prioritized according to the prevalence and demographic size of different risk groups. These interventions should reinforce integration of family planning and HIV/AIDS services through community health workers; household based campaigns, reproductive health services and reproductive health courses at school.

Couples (males and women in union) living in the Southern region and those living in the urban areas of the Central and the Northern should be the first targets. Indeed, this group includes 45 percent of the study population, among whom the HIV prevalence is estimated at 17 percent on average.

Unmarried women, including never married women and those in union disruption, could be considered as the second target using “Abstinence”, “Be faithful” and “use condom” campaign. Indeed, though women in union dissolution represent only about 13 percent of women of reproductive age in Malawi, they have the higher HIV prevalence in Malawi. Similarly, despite low HIV prevalence among never married women, findings show relatively high HIV prevalence among single mothers. Therefore, zero new infection among single women can have a significant effect in achieving the MDG 6.

In conclusion, this study recommends: (1) design and implementation of targeted interventions taking into account HIV prevalence and the demographic size of different groups at risk; (2) reinforcement of integration of family planning and HIV/AIDS services through community health workers, households based campaign, reproductive health services and reproductive health courses at school.

Acknowledgements: *This research was supported by the Luxembourg National Research Fund (FNR). The authors thank Macro international for providing free the Malawi 2004 and 2010 DHS data-sets including HIV data. We thank Prof. Guillaume Wunsch for his comments.*

Contributors: JBOE participated in conception and design, literature review, data analysis and interpretation, drafting the article, critical revisions for important intellectual content and approval of final article for submission. MK, NJM, EMZ and YY participated in conception and design, interpretation of results, critical revisions for important intellectual content and approval of final article for submission.

Data sharing statement: This study is based on the Malawi 2004 and 2010 Demographic and Health Surveys. These data are available on www.measuredhs.com. Access to individual HIV status as well as individual background information required authorization from MACRO ORC (www.measuredhs.com). The Website depicts the process.

Funding

The Luxembourg National Research Fund (FNR).

Competing Interests

None

References

[1] United Nations.*The Millennium Development Goals. 2012 Report* . New York: United Nations, 2012.

[2] UAIDS. *World AIDS day report 2011. How to get to zero: Faster, smarter, Better.* Geneva: UNAIDS, 2011.

[3] Durevall, D, & Lindskog, A. *HIV/AIDS, Adult Mortality and Fertility: Evidence from Malawi. Working Papers in Ecomics, No 284.* Göteborg: School of Business, Economics and Law, University ofGothenburg, 2007.

[4] Rombo, D. *Marital risk factors and HIV infection among women: A comparison between Ghana and Kenya,* Ph.D. Dissertation. Minneapolis: University of Minnesota, 2009.

[5] Öjteg, K. *Socio-Economic determinants of HIV in Zambia. A district level Analysis*. Ph.Dissertation. Lund: Department of Economics at the University of Lund, 2009.

[6] Magadi, M., & Desta , M. A multilevel analysis of the determinants and cross-national variations of HIV seropositivity in sub-Saharan Africa: Evidence from the DHS. *Health & Place*, 2011; 17 (5): 1067–83.

[7] Magadi, M. Understanding the gender disparity in HIV infection across countries in sub-Saharan Africa: evidence from the Demographic and Health Surveys. *Sociology of Health & Illness,* 2011; 33 (4): 522-39.

[8] Asiedu, C., Asiedu, E., & Owusu, F. The Socio-Economic Determinants of HIV/AIDS Infection Rates in Lesotho, Malawi, Swaziland and Zimbabwe. *Development Policy Review ,* 2012; 30 (3): 305-326 .

[9] Green, E. *Rethinking AIDS Prevention. Learning from Successes in Developing Countries* . Westport, CT: Praeger publishers, 2004

[10] International HIV / AIDS Alliance. *The cost efficiency of HIV prevention for vulnerable and most-at-risk populations and the reality of funding. What's Preventing Prevention Campaign Briefing 2.* Hove: International HIV / AIDS Alliance, 2010.

[11] Population Reference Bureau, *2012 World Population Data sheet*, Washington DC, PRB, 2012.

[12] United Nations Development Programme/UNDP. Sustainability and Equity: A better future for all, Human Development Report 2011. New York, UNDP, 2011.

[13] Malawi National Statistical Office (NSO); ORC Macro. *Malawi Demographic and Health Survey 2004.* Calverton, Maryland: NSO and ORC Macro, 2005.

[14] Malawi National Statistical Office (NSO); ICF Macro. *Malawi Demographic and Health Survey 2010.* Zomba, Malawi, and Calverton, Maryland, USA: NSO and ICF Macro, 2011.

[15] Kass, G. V. An Exploratory Technique for Investigating Large Quantitaties of Categorical Data. *Applied Statistics,* 1980; 29 (2): 119-127.

[16] Kitsantas, P., Hollander M., & Li, L. Using classification trees to assess low birth weight outcomes. *Artificial Intelligence in Medicine*. 2006; 38 (3): 275-289.

[17] Adair, T. *HIV Status and Age at First Marriage among Women in Cameroon.* DHS Working Paper No 33, Calverton, Maryland, USA: Macro International Inc, 2007.

[18] Rodrigo, C., & Rajapakse, S. (2010). HIV, poverty and women. *International Health*, 2(1): 9-16.

[19] Takyi, B., & Broughton, C. Marital Stability in Sub-Saharan Africa: Do Women's Autonomy and Socioeconomic Situation Matter? *Journal of Family and Economic Issues* , 2006; 27 (1): 113-132.

[20] Reniers, G. Divorce and remarriage in rural Malawi. *Demographic Research*, 2003; Special Collection 1, article 6: 175-206. www.demographic-research.org/special/1/6/S1-6.pdf.

[21] Reniers, G. Marital strategies for regulating exposure to HIV. *Demography*, 2008; 45 (2): 417-438.

Table 1 – Description of the sample

Socioeconomic and demographic Characteristics	Weight			Unweight		
	2004	2010	Total	2004	2010	Total
Age						
15-19	11.4	11.5	11.5	12.1	11.6	11.8
20-24	26.3	21.5	22.8	25.7	20.8	22.2
25-29	19.3	22.7	21.7	20.0	21.9	21.4
30-34	15.7	15.2	15.3	14.9	15.7	15.4
35-39	10.6	13.0	12.3	10.6	12.7	12.1
40-44	9.7	8.6	8.9	9.4	9.1	9.2
45-49	7.1	7.5	7.4	7.3	8.3	8.0
Average	29.2	29.6	29.5	29.1	29.8	29.6
Age at first sex						
<15	18.5	19.1	18.9	20.4	19.2	19.6
15-19	70.1	68.5	68.9	68.4	68.4	68.4
20-24	10.0	11.2	10.8	9.9	11.1	10.7
25&+	1.4	1.3	1.3	1.3	1.3	1.3
Average	16.6	16.6	16.6	16.5	16.6	16.6
Marital status						
Single	6.1	7.5	7.1	6.3	7.6	7.2
In union	81.6	77.4	78.6	80.3	77.1	78.0
Ever married	12.3	15.1	14.3	13.4	15.4	14.8
Number of ever born children						
0	9.7	9.9	9.9	10.3	9.6	9.8
1&+	90.3	90.1	90.1	89.7	90.4	90.2
Age at first birth						
Never give birth	11.0	10.3	10.5	10.2	10.6	10.5
< 20 years old	65.0	64.8	64.9	64.6	64.2	64.3
20 & +	24.0	24.9	24.6	25.2	25.2	25.2
Ever had premarital child						
No	87.2	88.7	88.3	87.9	88.4	88.2
Yes	12.8	11.3	11.7	12.1	11.6	11.8
Relationship to the head of household						
Head of household	16.9	19.4	18.7	17.9	19.0	18.6
Spouse	68.1	62.6	64.1	67.0	62.4	63.7
Daughter & Grand daughter	10.0	11.0	10.7	10.2	11.7	11.3
Others	5.1	7.1	6.5	5.0	6.9	6.3
Province of residence						
Northern	13.7	11.1	11.8	14.3	17.5	16.6
Central	38.0	42.2	41.0	33.8	34.1	34.0
Southern	48.3	46.7	47.1	51.9	48.4	49.4
Place of residence						
Urban	14.4	19.2	17.8	12.2	13.1	12.8
Rural	85.6	80.8	82.2	87.8	86.9	87.2
Religion						
Catholic	23.0	21.2	21.7	21.6	20.6	20.9
Protestant	25.9	24.3	24.7	24.9	25.2	25.1
Other Christians	38.6	39.7	39.4	37.1	42.3	40.8
Muslim	11.6	13.5	12.9	15.5	10.9	12.2
Others	0.9	1.3	1.2	0.9	1.0	1.0
Education						
None	25.7	17.5	19.8	25.4	16.6	19.1
Primary	61.4	63.8	63.1	61.8	66.4	65.1
Secondary & +	12.8	18.7	17.1	12.8	17.0	15.8
Household wealth Index						
Poorest	17.1	17.6	17.5	17.6	19.0	18.6
Poorer	21.2	20.1	20.4	20.8	20.6	20.7
Middle	21.9	19.7	20.3	22.9	20.9	21.4
Richer	22.4	19.3	20.2	22.3	20.7	21.2
Richest	17.4	23.3	21.6	16.4	18.8	18.1
Total	2605	6395	9000	2605	6395	9000

Table 2 – Factors associated with HIV prevalence: Descriptive analyses

Socioeconomic and demographic Characteristics	HIV+	N	Chi-Square	P-value
Age				
15-19	6.1	1,060		
20-24	8.9	1,996		
25-29	14.1	1,922		
30-34	20.0	1,389	205.10	0.000
35-39	22.1	1,091		
40-44	19.7	824		
45-49	15.3	718		
Age at first sex				
<15	18.1	1,762		
15-19	13.8	6,157	21.69	0.000
20-24	12.7	967		
25&+	14.9	114		
Marital status				
Single	9.0	649		
In union	11.9	7,021	331.20	0.000
Ever married	31.3	1,330		
Number of ever born children				
0	10.4	886	12.91	0.000
1&+	14.9	8,114		
Age at first birth				
Never give birth	11.5	946		
< 20 years old	14.8	5,837	7.080	0.029
20 & +	14.9	2,217		
Ever experience premarital childbearing				
No	13.7	7,941	29.54	0.000
Yes	20.2	1,059		
Relationship to the head of household				
Head of household	25.3	1,678		
Spouse	11.7	5,737	197.80	0.000
Daughter & Grand daughter	11.1	1,014		
Others	16.4	571		
Region of residence				
Northern	10.0	1,494		
Central	9.5	3,062	184.90	0.000
Southern	20.0	4,444		
Place of residence				
Urban	24.7	1,156	157.00	0.000
Rural	12.3	7,844		
Religion				
Catholic	13.1	1,879		
Protestant	15.1	2,258		
Other Christians	14.3	3,674	7.84	0.090
Muslim	16.4	1,100		
Others	10.6	89		
Education				
None	14.3	1,722		
Primary	13.4	5,855	27.92	0.000
Secondary & +	18.8	1,423		
Household wealth Index				
Poorest	10.3	1,673		
Poorer	10.5	1,862		
Middle	12.4	1,930	148.60	0.000
Richer	15.7	1,904		
Richest	22.5	1,631		
Year of survey				
2004	14.4	2,605	0.01	0.935
2007	14.5	6,395		
Total	14.5	9,000		

Table 3 - Model Summary

Model components	Model specification	Results
Growing Method:	CHAID	-
Dependent variable	HIV Marital status	HIV+=14.7%
Independent Variables	Age, Age at first sex, Marital status, Age at first union, Ever had a child, Experience premarital childbearing, age at first birth, Region of residence, Place of residence, Education, Wealth Index, Religion, Relationship to the head of household.	**Marital status**, Region of residence, Age, Place of residence, Wealth Index, Age at first sex, Ever had a child
Maximum Tree Depth	3	3
Minimum Cases in Parent Node	100	100
Minimum Cases in Child Node	50	50
Number of Nodes	-	27
Number of Terminal Nodes	-	16

Table 4 – Prevalence of HIV by groups

Node	Group description	Population %	Population N	HIV Prevalence
	Group 1			
22	Formerly in union-richest-had first sex from 25 years old	0.7	62	72.7
19	Formerly in union-richer or middle households- Southern region	2.7	242	45.7
21	Formerly in union-richest-had first sex between 15 and 24 years old	2.3	206	45.3
Total	*Group 1*	*5.7*	*510*	*54.6*
	Group 2			
7	Formerly in union-poorer households	2.8	251	27.5
25	Never married, living in urban area-Southern or Northern region	1.6	144	23.5
12	In union – living in Southern region – age 30-44	13.2	1,185	22.9
20	Formerly in union-richer or middle households-Central or Northern region	1.6	144	21.9
17	Formerly in union-poorest households-Southern region	2.3	206	20.9
Total	*Group 2*	*21.5*	*1,930*	*23.3*
	Group 3			
16	In union – living in Central or Northern region – urban area	6.1	549	18.0
14	In union – living in Southern region – age 25-29	9.0	810	18.0
23	Never married, living in rural area and ever gave birth	1.4	126	11.4
13	In union – living in Southern region – age 15-24/ 45-49	14.3	1,286	11.2
18	Formerly in union-poorest households-Central or Northern region	2.1	189	10.3
Total	*Group 3*	*32.9*	2,960	*13.8*
	Group 4			
15	In union, living in Northern or Central province – rural areas	35.9	3,231	6.1
26	Never married, living in urban area in Central region	0.9	81	2.7
24	Never married, living in rural area and never gave birth	3.2	288	2.3
Total	*Group 4*	*40.0*	3,600	*3.7*
	Total (Overall)	100	9,000	14.7

Figure 1 - HIV prevalence in Malawi: tree diagram

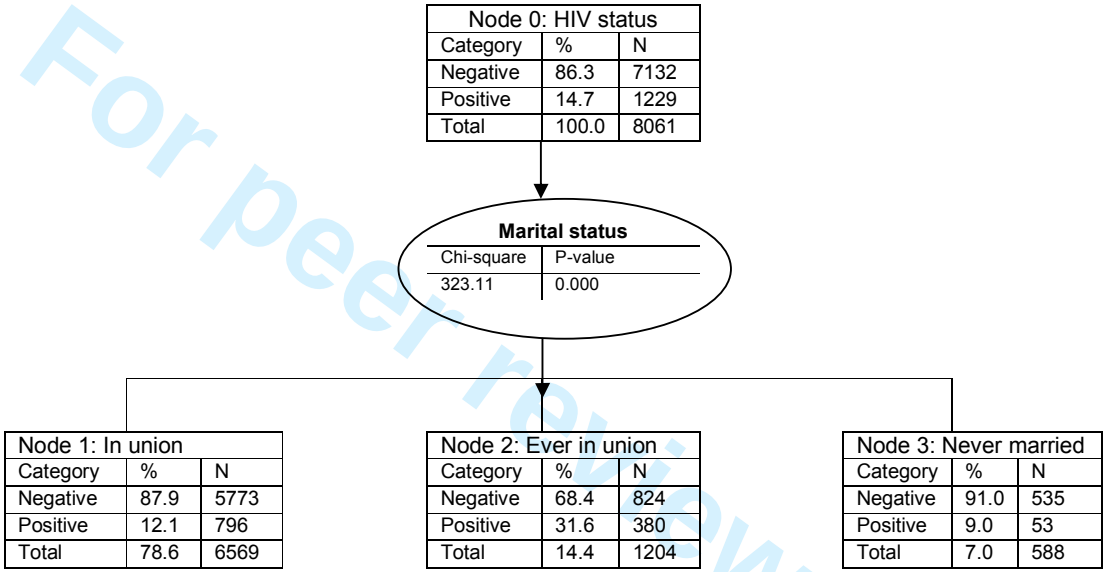


Figure 1a- HIV prevalence in Malawi: tree diagram for women in union

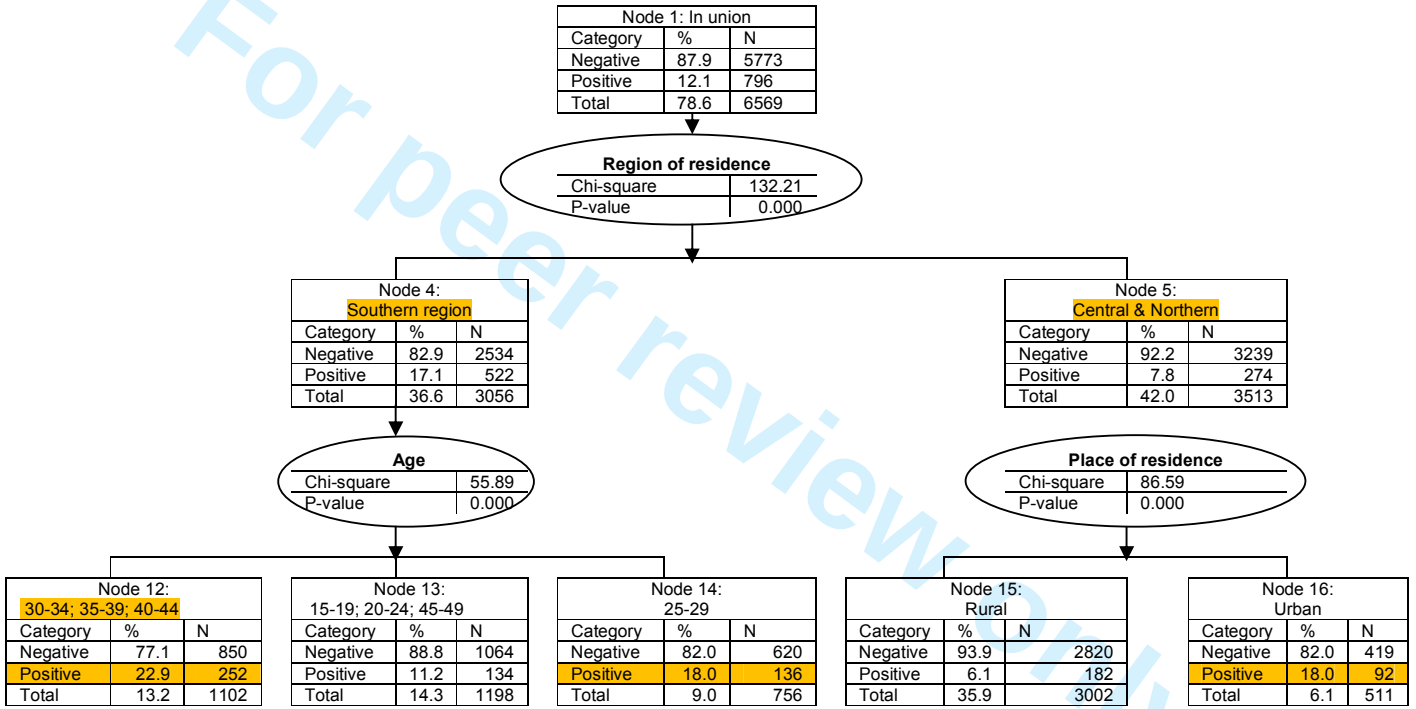


Figure 1b - HIV prevalence in Malawi: tree diagram for women in union disruption

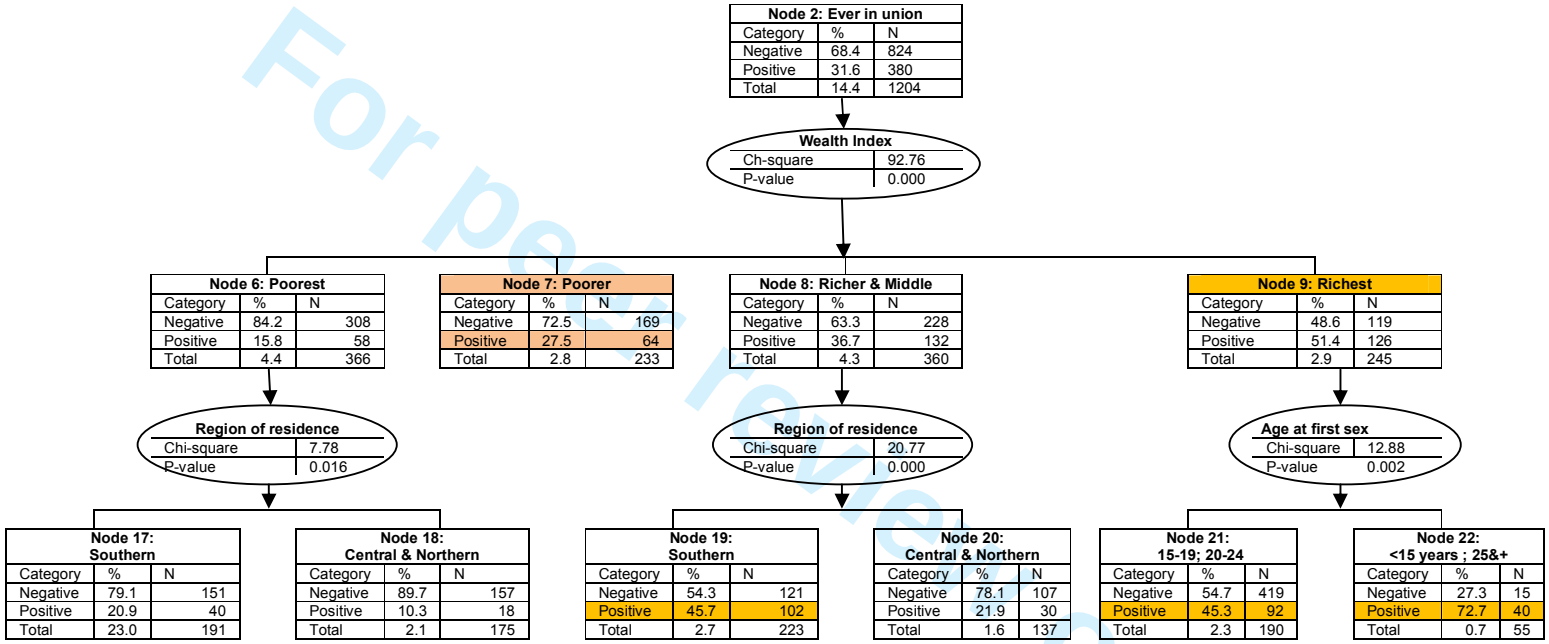
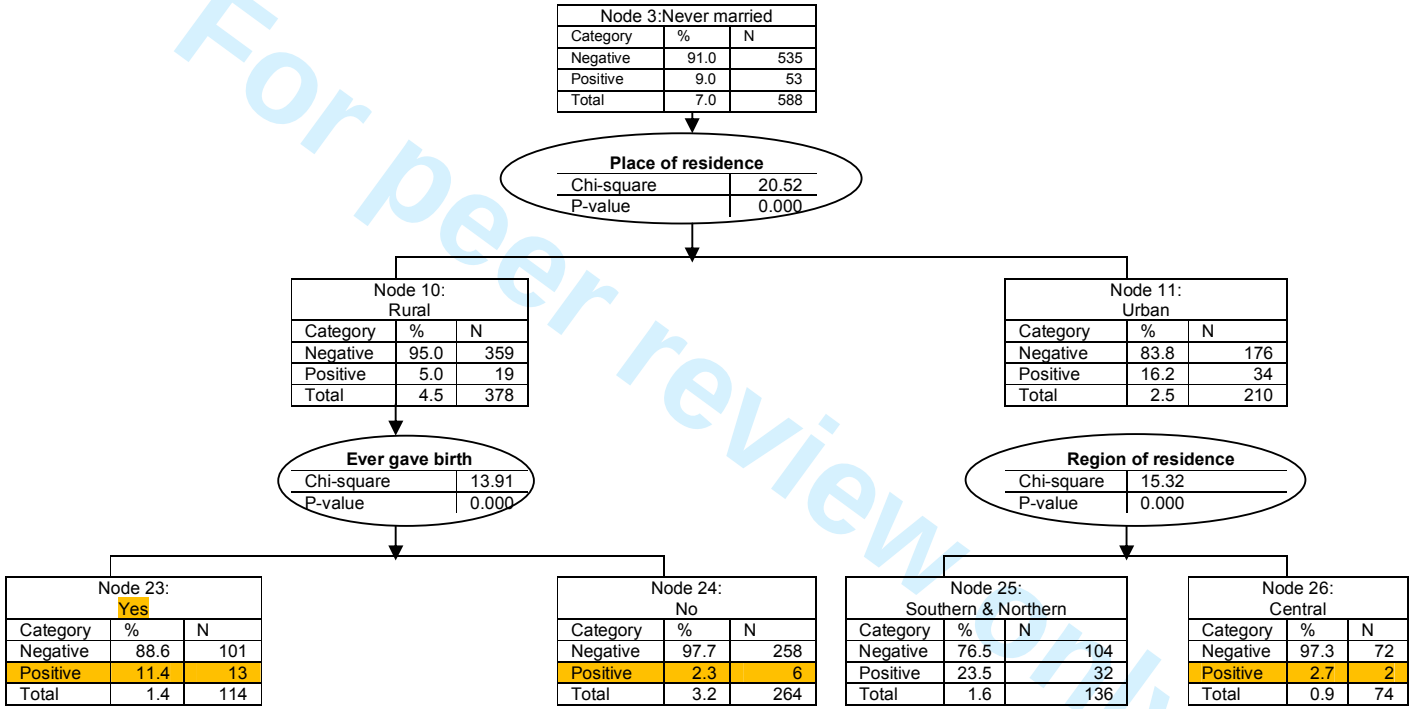


Figure 1c - HIV prevalence in Malawi: tree diagram for never married women



Article summary

1. Article focus

- Targeted interventions and evidence based prevention programmes have been advocated as cost-effective strategy to combat HIV/AIDS.
- Whereas in countries with concentrated AIDS epidemics (Latina America, East Asia and Eastern Europe), the most-at-risk populations including commercial sex workers (CSWs), long distance truck drivers, men who have sex with men, and unmarried youth account for a large proportion of new infections, in countries with high prevalence, they account only for a smaller share of new infections.
- Who are the most-at-risk populations regarding HIV prevalence in Malawi? With HIV prevalence of about 13 percent among women of reproductive age, HIV/AIDS constitutes a drain on the labor force and government expenditures in Malawi.

2. Key messages

- We use data from the Malawi 2004 and 2010 Demographic and Health Surveys to profile HIV most-at-risk groups of women in Malawi where about 14 percent of women are HIV positive.
- Our findings revealed that richest and formerly in union women are the most-at-risk population.
- We suggested targeted interventions considering the groups HIV prevalence and size. With 45 percent of the study population, among whom the HIV prevalence is estimated at 17 percent on average, couples (males and women in union) living in the Southern region and those living in the urban areas of the Central and the Northern should be the first targets of HIV interventions in Malawi.

Strengths and limitations of this study

- Our study seems the first in Malawi that has attempted to profile HIV most-at-risk groups of women in Malawi. The most-at-risk population refers to a combination of several factors because socioeconomics factors associated with HIV are not mutually exclusive.
- The major strength is the use of the Chi-square Automatic Interaction Detector (CHAID) to identify HIV predictors and the most at risk groups of women for intervention. CHAID uses regression and classification algorithms and offers a non-algebraic method for partitioning data that lends itself to graphical displays.
- The major limitation of this study is the cross- sectional nature of the data, which does not permit one to draw causal association between HIV status and the associated factors. For instance, whether HIV infection has occurred before, during or after the union.

STROBE 2007 (v4) Statement—Checklist of items that should be included in reports of *cross-sectional studies*

Section/Topic	Item #	Recommendation	Reported on page #
Title and abstract	1	(*a*) Indicate the study's design with a commonly used term in the title or the abstract	1
		(*b*) Provide in the abstract an informative and balanced summary of what was done and what was found	2
Introduction			
Background/rationale	2	Explain the scientific background and rationale for the investigation being reported	4
Objectives	3	State specific objectives, including any prespecified hypotheses	4
Methods			
Study design	4	Present key elements of study design early in the paper	5
Setting	5	Describe the setting, locations, and relevant dates, including periods of recruitment, exposure, follow-up, and data collection	5
Participants	6	(*a*) Give the eligibility criteria, and the sources and methods of selection of participants	5
Variables	7	Clearly define all outcomes, exposures, predictors, potential confounders, and effect modifiers. Give diagnostic criteria, if applicable	6
Data sources/ measurement	8*	For each variable of interest, give sources of data and details of methods of assessment (measurement). Describe comparability of assessment methods if there is more than one group	5-6
Bias	9	Describe any efforts to address potential sources of bias	*5-7*
Study size	10	Explain how the study size was arrived at	5
Quantitative variables	11	Explain how quantitative variables were handled in the analyses. If applicable, describe which groupings were chosen and why	6
Statistical methods	12	(*a*) Describe all statistical methods, including those used to control for confounding	6-7
		(*b*) Describe any methods used to examine subgroups and interactions	6-7
		(*c*) Explain how missing data were addressed	6-7
		(*d*) If applicable, describe analytical methods taking account of sampling strategy	6-7
		(*e*) Describe any sensitivity analyses	6-7
Results			

Participants	13*	(a) Report numbers of individuals at each stage of study—eg numbers potentially eligible, examined for eligibility, confirmed eligible, included in the study, completing follow-up, and analysed	5
		(b) Give reasons for non-participation at each stage	5
		(c) Consider use of a flow diagram	5
Descriptive data	14*	(a) Give characteristics of study participants (eg demographic, clinical, social) and information on exposures and potential confounders	7-8
		(b) Indicate number of participants with missing data for each variable of interest	7-8
Outcome data	15*	Report numbers of outcome events or summary measures	7-8
Main results	16	(*a*) Give unadjusted estimates and, if applicable, confounder-adjusted estimates and their precision (eg, 95% confidence interval). Make clear which confounders were adjusted for and why they were included	8-12
		(*b*) Report category boundaries when continuous variables were categorized	8-12
		(*c*) If relevant, consider translating estimates of relative risk into absolute risk for a meaningful time period	8-12
Other analyses	17	Report other analyses done—eg analyses of subgroups and interactions, and sensitivity analyses	8-12
Discussion			
Key results	18	Summarise key results with reference to study objectives	12-14
Limitations	19	Discuss limitations of the study, taking into account sources of potential bias or imprecision. Discuss both direction and magnitude of any potential bias	12-14
Interpretation	20	Give a cautious overall interpretation of results considering objectives, limitations, multiplicity of analyses, results from similar studies, and other relevant evidence	12-14
Generalisability	21	Discuss the generalisability (external validity) of the study results	12-14
Other information			
Funding	22	Give the source of funding and the role of the funders for the present study and, if applicable, for the original study on which the present article is based	15

*Give information separately for cases and controls in case-control studies and, if applicable, for exposed and unexposed groups in cohort and cross-sectional studies.

Note: An Explanation and Elaboration article discusses each checklist item and gives methodological background and published examples of transparent reporting. The STROBE checklist is best used in conjunction with this article (freely available on the Web sites of PLoS Medicine at http://www.plosmedicine.org/, Annals of Internal Medicine at http://www.annals.org/, and Epidemiology at http://www.epidem.com/). Information on the STROBE Initiative is available at www.strobe-statement.org.

Identifying HIV most-at-risk groups in Malawi for targeted interventions

Journal:	*BMJ Open*
Manuscript ID:	bmjopen-2012-002459.R1
Article Type:	Research
Date Submitted by the Author:	11-Mar-2013
Complete List of Authors:	Emina, Jacques; University of Kinshasa, ; CEPS/ INSTEAD, Madise, Nyovani; University of Southampton, Faculty of Social and Human Sciences Kuepie, Mathias; CEPS/INSTEAD, AFRILUX Zulu, Eliya; African Institute for Development Policy, Yazoume , Ye ; ICF International,
Primary Subject Heading:	HIV/AIDS
Secondary Subject Heading:	Epidemiology, Health policy, Sociology, Public health
Keywords:	HIV & AIDS < INFECTIOUS DISEASES, Epidemiology < INFECTIOUS DISEASES, Public health < INFECTIOUS DISEASES

Prevalence of HIV among women in Malawi: Identifying the most-at-risk groups for targeted and cost-effective interventions

Jacques B. O. Emina [1,2], Nyovani Madise [3], Mathias Kuepie [2], Eliya M. Zulu [4], Yazoume Ye [5]

1 Department of Population and Development Studies, University of Kinshasa, P.O.BOX 176 Kinshasa XI, Democratic Republic of Congo

2 CEPS/INSTEAD 3, avenue de la Fonte, L-4364 Esch-sur-Alzette Luxembourg

3 Faculty of Social and Human Sciences, Room 4103 Building 58, University of Southampton, Southampton, SO17 1BJ.

4 African Institute for Development Policy (AFIDEP), P. O. Box 14688-00800, Westlands, Nairobi, Kenya.

5 ICF International 11785 Beltsville Drive, Suite 300 Calverton, MD 20705, USA.

Correspondence to: Dr Jacques B.O. Emina; jacques.emina@gmail.com

Abstract

Objectives: To identify HIV socioeconomic predictors as well as identify the most-at-risk groups of women in Malawi.

Design: consecutive cross-sectional surveys

Setting: Malawi

Participants: The study relies on a sample of 8,596 age 15-49 from the 2004 and 2010 Malawi Health and Demographic Surveys (MDHS).

I**nterventions**: N/A

Primary and secondary outcome measures: Whether individual is HIV positive.

Results: Findings from Pearson Chi-square and Chi-square Automatic Interaction Detector (CHAID) analyses revealed that marital status is the most significant predictor of HIV. Women who are no longer in union and living in rich households constitute the most-at-risk population; whereas the less-at-risk group includes nulliparous never married women living in the rural areas.

Conclusion: In the light of these findings, this study recommends: (1) design and implementation of targeted interventions taking into account HIV prevalence and the demographic size of different groups at risk groups. With 45 percent of the study population, among whom the HIV prevalence is estimated at 17 percent on average, couples (males and women in union) living in the Southern region and those living in the urban areas of the Central and the Northern should be the first targets of HIV interventions in Malawi; (2) reinforcement of integration of family planning and HIV/AIDS services through community health workers, households based campaign, reproductive health services and reproductive health courses at school.

Keywords: HIV/AIDS, Malawi, CHAID methods, decision analysis, most-at-risk groups, targeted interventions.

Article summary

1. Article focus

- Targeted interventions and evidence based prevention programmes have been advocated as cost-effective strategy to combat HIV/AIDS.
- whereas in countries with concentrated AIDS epidemics (Latina America, East Asia and Eastern Europe), the most-at-risk populations including commercial sex workers (CSWs), long distance truck drivers, men who have sex with men, and unmarried youth account for a large proportion of new infections, in countries with high prevalence, they account only for a smaller share of new infections.
- Who are the most-at-risk populations regarding HIV prevalence in Malawi? With HIV prevalence of about 13 percent among women of reproductive age, HIV/AIDS constitutes a drain on the labor force and government expenditures in Malawi.

2. Key messages

- We use data from the Malawi 2004 and 2010 Demographic and Health Surveys to profile HIV most-at-risk groups of women in Malawi where about 14 percent of women are HIV positive.
- Our findings revealed that richest and formerly in union women are the most-at-risk population.
- We suggested targeted interventions considering the groups HIV prevalence and size. With 45 percent of the study population, among whom the HIV prevalence is estimated at 17 percent on average, couples (males and women in union) living in the Southern region and those living in the urban areas of the Central and the Northern should be the first targets of HIV interventions in Malawi.

Strengths and limitations of this study

- Our study seems the first in Malawi that has attempted to profile HIV most-at-risk groups of women in Malawi. The most-at-risk population refers to a combination of several factors because socioeconomics factors associated with HIV are not mutually exclusive.
- The major strength is the use of the Chi-square Automatic Interaction Detector (CHAID) to identify HIV predictors and the most at risk groups of women for intervention. CHAID uses regression and classification algorithms and offers a non-algebraic method for partitioning data that lends itself to graphical displays.
- The major limitation of this study is the cross- sectional nature of the data, which does not permit one to draw causal association between HIV status and the associated factors. For instance, we do not know whether HIV infection has occurred before, during or after the union.

Introduction

In 2000, the United Nations' Millennium summit identified the reduction of HIV prevalence as one of the eight fundamental goals for furthering human development. Though global HIV/ AIDS incidence is declining, HIV/AIDS has remained the leading cause of death in women of reproductive age in low- and middle-income countries, particularly in sub-Sahara Africa (SSA) [1]. The gap between the state of HIV/AIDS and the UNAIDS goals of three zero (zero new HIV infections, zero discrimination, and zero AIDS related deaths) remains important. With barely two years remaining to the end-date of the MDG target, HIV/AIDS remains a long-term global challenge [1].

Based on the current costs of HIV treatment (US $ 4,707 over lifetime) (International HIV / AIDS Alliance, 2010), targeted interventions and evidence based prevention programmes have been advocated as cost-effective strategy to combat HIV/AIDS. Such a strategy reduces levels of vulnerability and risk as well as allowing HIV interventions to optimize coverage, reducing costs and lowering the number of new infections [2].

Despite growing literature in health and social sciences on factors associated with HIV/AIDS during the last three decades, important questions are still poorly answered in the context of Malawi [3, 4, 5, 6, 7, 8]. For instance, who are the most-at-risk populations regarding HIV prevalence in Malawi? Indeed, whereas in countries with concentrated AIDS epidemics (Latina America, East Asia and Eastern Europe), the most-at-risk populations including commercial sex workers (CSWs), long distance truck drivers, men who have sex with men, and unmarried youth [9,4,10] account for a large proportion of new infections, in countries with high prevalence, they account only for a smaller share of new infections [10].

Against this background, this study aims to identify HIV socioeconomic predictors as well as identify the most-at-risk groups of women in Malawi. With HIV prevalence of about 13 percent among women of reproductive age [11], HIV/AIDS constitutes a drain on the labor force and government expenditures in Malawi. Most-at-risk populations refer to a combination of several factors because socioeconomics factors associated with HIV are not mutually exclusive.

Data and Methods

Study setting

The Republic of Malawi is a landlocked country in southeast Africa. Malawi is over 118,000 km^2 with an estimated population of about 16 millions [11]. Its capital is Lilongwe, which is also Malawi's largest city; the second largest is Blantyre and the third is Mzuzu.

Malawi is among the world's least-developed countries. The economy is heavily based on agriculture, with a largely rural population. The country GNI PPP per Capita is estimated at $860 while the world average is estimated at $10,780 [11]. Ninety-one percent of Malawians live below 2 dollars (US) per day. The country's Human Development Index is estimated at 0.400, which gives the country a rank of 171 out of 187 countries with comparable data [12].

Malawi has a low life expectancy (53 years) and high infant mortality (66 deaths per 1,000 live births) compared to the world' average (41 deaths per 1,000 live births). There is a high prevalence of HIV/AIDS, especially among women, among whom about 14 percent are HIV positive [11]. Malawi is divided into 28 districts within three regions: Southern, Central and Northern regions.

Data sources

This study uses data from two independently pooled cross-section data, the 2004 and 2010 Malawi Health and Demographic Surveys (MDHS). The inclusion of HIV testing in the 2004 and 2010 MDHS offers the opportunity to identify socioeconomic profile of women age 15-49 who were living with HIV. Participation in HIV testing was voluntary. To ensure confidentiality, case numbers (and not names) were used in linking the HIV test results to individual and household characteristics.

In both surveys, a subsample of one-third of the households was selected to conduct HIV testing for eligible women age 15-49. Overall, 4,071 women age 15-49 were identified as eligible for testing in the 2004 MDHS. Of these, testing was successfully conducted on 2,686 women, resulting in a response rate of 70 percent for women. Ninety percent of all 2010 MDHS women who were eligible (8,174) for testing were interviewed and consented to HIV testing. Details on the sample design are provided elsewhere [13,14].

Variables

The dependent variable for this analysis is HIV status, characterized as positive or negative depending on blood tests carried out during the surveys. The independent variables encompass 12 main variables grouped into two major types:

1. Demographic and reproductive behavior variables: age, age at first sex, marital status, age at first birth, number of children ever born, Experience in premarital childbearing, and relationship to the head of household.
2. Socioeconomic and contextual variables: religion, region of residence, place of residence, education, and household wealth index.

The choice of these variables is guided by the literature on factors associated with HIV in sub-Saharan Africa [3-8].

Statistical analyses

Statistical analyses rely on Pearson Chi-square and Chi-square Automatic Interaction Detector (CHAID) using SPSS version 16. We used data weights to take into account the complexity of the DHS design.

We performed Pearson chi-square to identify associations between the HIV status (positive, negative) and demographic and reproductive behavior variables as well as socioeconomic and contextual variables.

We used CHAID to identify HIV predictors and the most at risk groups of women for intervention [15]. This is a nonparametric technique that makes no distributional assumptions and on outliers, collinearities, heteroskedasticity, or distributional error structures. The dependent variable and predictor variables can be nominal (categorical), ordinal (ordered categories ranked from small to large), or interval (a "scale").

CHAID uses regression and classification algorithms and offers a non-algebraic method for partitioning data that lends itself to graphical displays. The method is a sequential fitting algorithm and its statistical tests are sequential with later effects being dependent upon earlier ones, and not simultaneous as would be the case in a regression model or analysis of variance where all effects are fit simultaneously. At each step, CHAID chooses the independent (predictor) variable that has the strongest interaction with the HIV status (dependent variable).

CHAID analysis goes one step further, and identifies as the important elements of HIV prevalence those variables that most differentiate the likelihood to be HIV positive. The CHAID procedure begins by finding variables that have a significant association with HIV status. It then assesses the category groupings, or interval breaks to pick the most significant combination of variables. The variable having the strongest association with HIV status becomes the first branch in a tree with a leaf for each category that is significantly different relative to be HIV positive. The process is repeated to find the predictor variable on each leaf most significantly related to HIV status, until no significant predictors remain.

The developed model is a classification tree (or data partitioning tree) that shows how major "types" formed from the independent (predictor or splitter) variables differentially predict a criterion or dependent variable. The method permits also identification of women who are likely to be members of a particular group (Segmentation), and assign cases into one of several categories, such as high-, medium-, and low-risk groups (stratification). Selecting a useful subset of predictors from a large set of variables for use in building a formal parametric model (Data reduction and variable screening); Identify relationships that pertain only to specific subgroups and specify these in a formal parametric model (Interaction identification); and recoding group predictor categories and continuous variables with minimal loss of information. Categories of each predictor are merged if they are not significantly different with respect to the dependent variable (Category merging and discretizing continuous variables).

Like other decision trees methods, CHAID's its output is highly visual and easy to interpret. It allows identifying homogeneous groups with high or low risk; and constructing rules for making predictions about individual cases.

However, CHAID needs large sample sizes to work effectively because it uses multi-way splits. Indeed, with small sample sizes the respondent groups can quickly become too small for reliable analysis.

Results

Sampling description

Table 1 presents the main characteristics of the population under study. Since the principal mode of HIV transmission in Malawi is heterosexual contact, our analyses

focus on women who ever had sexual intercourse. The distribution of the sample by age shows that more than half (56 percent) of the populations are age less than 30 years old. The average age of the sample is estimated at 29 years old. Women who are in union (i.e., currently married or living with a man) constitute about 80 percent. The proportion of women who have never been married is estimated at 14 percent. Regarding the relationship to the head of household, the majority of women are spouse (64 percent). Eighteen percent of the studied population are head of household.

Since the principal mode of HIV transmission in Malawi is heterosexual contact, our analyses focus on women who ever had sexual intercourse.

[Table 1, about here]

Table 1 also shows that the majority of women (82 percent) live in rural areas. By region, the majority of women live in the Central and Southern Regions, while 12 percent of women live in the Northern Region. Furthermore, although the majority of respondents have had some education, 20 percent of women never attended school, while 63 percent have attended only primary school.

Considering reproductive behavior, a large majority of women had their first sexual intercourse before 20 years (average 16 years old).

Factors associated with HIV prevalence: bivariate analysis

Table 2 describes HIV prevalence in Malawi by women's selected background characteristics. Overall, 14 percent of studied women are HIV positive. Except for religion, all independent variables are statistically associated with HIV infection status.

[Table 2, about here]

HIV infection prevalence was high (20 percent) among women aged 30-39 years. Regarding marital status, women who are no longer in union (widowed, divorced and separated) had significantly higher prevalence (30 percent) compared to those who have never been in a marital union (10 percent). HIV prevalence was high among heads of household.

Furthermore, while 25 percent of women in urban area were HIV positive, the prevalence was less than half (12 percent) compared to their counterparts from the rural areas. The HIV epidemic shows regional heterogeneity with a higher prevalence

(20 percent) observed in the Southern region. Women with secondary education had higher HIV prevalence compared to those who never attended school (18 percent versus 14 percent). Regarding the household wealth quintiles the prevalence of HIV infection is higher among the women from the highest quintiles. With reference to sexual and reproductive behavior, HIV prevalence was higher among women who had their first sexual intercourse before the 15th birthday and /or who have experienced a premarital childbearing.

HIV prevalence in Malawi risk groups: results from CHAID analysis

Table 3 shows some broad information about the specifications used to build the Chi-square Automatic Interaction Detector (CHAID) model and the resulting model (the model summary).

[Table 3, about here]

Eleven independent variables were specified, but only seven were included in the final model. The variables such as age at first sex, age at first birth and female education did not make a significant contribution to the model, so they were automatically dropped from the final model.

Overall, there are 27 nodes among which 16 terminal nodes. Parent nodes include at least 100 cases whereas child nodes account for 50 cases in minimum. The tree diagram shows that "Marital status" (Chi-square = 323.1, P-value<0.000) is the best predictor of HIV status among women in Malawi (Figure 1).

[Figure 1, about here]

The tree is split into 3 branches: (1) Node 1 including women in union; (2) Node 2 encompassing women formerly in union; and (3) Node 3 comprising never married women.

For women in union (married or living together), Figure 1a reveals that region of residence is the following best predictor (Chi-square = 132.21, P-value<0.000). Women living in the Southern region (Node 4) are opposed to those living in the Central and the Northern regions (Node 5).

[Figure 1a, about here]

The model includes one more predictor for women belonging to Node 4 (Southern region) and Node 5 (Central and Northern region). For women living in the Southern

region, the next best predictor is women's age at the survey (Chi-square=55.9, P-value <0.000). Women in union, living in the Southern region are divided into three groups considering their age at the survey:

- Women aged from 30 to 44 years old, among whom 23 percent are HIV-positive (Node 12);
- Women aged 25-29 years old (Node 14). HIV prevalence is estimated at 18 percent among age;
- Women aged from 15 to 24 and from 45 to 49, among which 11 percent are HIV positive.

Considering women in union and living in the Central and the Northern regions (Node 5), Place of residence is the next best predictor of HIV prevalence (Chi-square=86.6, P-value <0.000). Among these women prevalence of HIV varies between 6 percent in rural areas (Node 15) and 18 percent in urban area (Node 16).

Figure 1b shows that for women formerly in union (Node 2) including divorced, widowed and not living together, household wealth index is the second best predictor of HIV prevalence (Chi-square=92.8, P-value <0.000).

[Figure 1b, about here]

With reference to women living in poorer households (Node 7), wealth index is the only significant predictor of HIV prevalence. Proportion of HIV positive is estimated at 27.5 percent. Since there are no child nodes below it, this is considered a terminal node.

For women living in Poorest (Node 6) as well as in Richer and Middle (Node 8) the next best predictor of HIV prevalence is the region of residence. HIV prevalence is estimated at 21 percent among women formerly in union living in the poorest households of Southern province (Node 17), while proportion of HIV positive is estimated at 10 percent for poorest women living in the Central and Northern regions (Node 18).

Among women living in Middle and Richer households (Node 8), about 47 percent are HIV positive in the Southern region (Node 19), whereas HIV prevalence is estimated at 22 percent for those living in other regions (Node 20).

Figure 1c reveals also that age at first sex is the third best predictor (Chi-square=12.9, P-value <0.002) for HIV prevalence among women formerly in union and living in the richest households (Node 9). Among those women HIV prevalence is estimated at 72.7 percent for women who experienced their first sex before the age of 15 or from 25 years old (Node 22). The corresponding HIV prevalence is estimated at 45.3 percent for women who had their first sex experience from 15 to 24 years old (Node 21).

Considering women who have been never in union (Figure 1c), place of residence is the second best predictor (Chi-square=20.5, P-value<0.000).

[Figure 1c, about here]

Among women living in rural area (Node 10) the model includes one more variable "Whether the woman ever gave birth" (Chi-square=13.9, P-value<0.000). The region of residence (Chi-square=15.3, P-value<0.000) is the additional significant variable for never married women living in urban areas.

Among never married living in rural areas, HIV prevalence is estimated at 2.3 percent if the woman never gave birth (Node 24) and at 11.4 percent if the woman experienced childbearing (Node 23). For never married women living in urban area 23 percent are HIV positive in the Southern and the Northern regions (Node 25), while proportion of women living with HIV is estimated at 2.7 percent among never married women living in urban areas of the Central region (Node 26).

Findings also allow dividing the study population into four major groups taking into account interaction between the most statistically significant variables: very high, high, intermediate, and low HIV prevalence. Table 4 describes composition of each group.

- The first group (very high prevalence) represents 5.7 percent of the sample. HIV prevalence is estimated at 54.6 percent on average, varying between 45.3 percent and 73 percent. This category encompasses three subgroups: a) women in union disruption, living in a richest household and who had their first sex from 25 years old; b) women in union disruption and live in richer or middle households in the Southern region; c) women in union disruption living in richest household and who had their first sex between 15 and 24 years old.

- The second group (high prevalence) includes 21 percent of the sample. HIV prevalence is estimated at 23.3 percent (range 21 percent - 28 percent) in this group. This group comprises 5 sub-groups: Women in union disruption living in poorer households; Never married women who live in urban area of the Southern or Northern region; Women in union living in the Southern region and who are age 30-44 years old; women formerly in union who are living in richer or middle households of the Central or Northern region; and women in union disruption living in the poorest households of the Southern region.

- The third group (intermediate prevalence) comprises about 33 percent of women under study. HIV prevalence varies between 10 and 19 percents (13.8 percent on average) among women belonging to this group. This category could be divided into 5 subgroups: (a) women in union, living in urban areas of the Central or Northern region; b) women in union who are living in the Southern region and aged 25-29; c) never married women, living in rural area and who have experienced childbearing; d) women in union, living in the Southern region and age 15-24 or 45-49; e) women who are in union disruption, living in poorest households of the Central or the Northern region.

- The last group (low prevalence) include three subgroups: a) women in union, living in the rural areas of the Northern or of the Central region; b) women who never married living in urban areas of the Central region; c) nulliparous never married women living in rural area. HIV prevalence is estimated at 3.7 percent, ranges between 2.3 percent and 6 percent. This group accounts for 40 percent of the sample.

Discussions and Conclusion

This paper aimed to describe and profile HIV prevalence among women in Malawi. The study relied on data from the Malawi 2004 and 2004 DHS using Chi-square and CHAID techniques. CHAID offers a useful alternative to traditional logistic regression and allows identifying population subgroups that share similar characteristics [16].

Analyses suggested three keys findings that could be summarized as follows. First, consistent with previous studies [6, 17], findings from bivariate analysis and chi-square

test showed high HIV prevalence among women in union dissolution, among those living in wealthy households and/or among women living in urban areas, as well as region heterogeneity in HIV prevalence.

Second, results from CHAID models reported that marital status is the best predictor of HIV status among women in Malawi followed by the household wealth index. Women who are no longer in union (widowed and divorced or separated) and living in less poor households have significantly higher HIV prevalence. This probably because: (1) a rich husband or a male partner may have more access to transactional sex and other risk behaviors such as polygamy which may increase women's vulnerability to HIV; (2) wealthier HIV positive widowed may have better quality of life as well as better access to treatment and survive longer [18]. Furthermore, divorced and separated are more frequent among the most educated women with economic autonomy [19]. Their causes (polygyny and/or infidelity) as well as consequences (multiple sexual partnerships) are also factors associated with HIV prevalence [20,21].

Last, CHAID models depicted also different interactions between risk factors and profiled HIV risk groups in Malawi. For instance, whilst HIV prevalence is higher among women living in urban areas (25 percent) compared to those living in rural areas (12 percent), only 3 percent of never married women living in urban areas of the Central region are HIV positive compared to 11 percent observed among single mothers living in the rural areas. Likewise, while overall HIV prevalence is low among never married women (9 percent), CHAID results revealed a higher HIV prevalence (23 percent) among never married women who live in urban areas of the Southern or Northern region compared to women in union who reside in urban areas of the Central or Northern (18 percent) as well as to women in union dissolution who live in poorest households of the Central or Northern region (10 percent).

In the light of these findings, it is noteworthy that to reduce number of new infection, interventions should be targeted and prioritized according to the prevalence and demographic size of different risk groups. These interventions should reinforce integration of family planning and HIV/AIDS services through community health workers; household based campaigns, reproductive health services and reproductive health courses at school.

Couples (males and women in union) living in the Southern region and those living in the urban areas of the Central and the Northern should be the first targets. Indeed, this group includes 45 percent of the study population, among whom the HIV prevalence is estimated at 17 percent on average.

Unmarried women, including never married women and those in union disruption, could be considered as the second target using “Abstinence”, “Be faithful” and “use condom” campaign. Indeed, though women in union dissolution represent only about 13 percent of women of reproductive age in Malawi, they have the higher HIV prevalence in Malawi. Similarly, despite low HIV prevalence among never married women, findings show relatively high HIV prevalence among single mothers. Therefore, zero new infection among single women can have a significant effect in achieving the MDG 6.

In conclusion, this study recommends: (1) design and implementation of targeted interventions taking into account HIV prevalence and the demographic size of different groups at risk; (2) reinforcement of integration of family planning and HIV/AIDS services through community health workers, households based campaign, reproductive health services and reproductive health courses at school.

Acknowledgements: *This research was supported by the Luxembourg National Research Fund (FNR). The authors thank Macro international for providing free the Malawi 2004 and 2010 DHS data-sets including HIV data. We thank Prof. Guillaume Wunsch for his comments.*

Contributors: JBOE participated in conception and design, literature review, data analysis and interpretation, drafting the article, critical revisions for important intellectual content and approval of final article for submission. MK, NJM, EMZ and YY participated in conception and design, interpretation of results, critical revisions for important intellectual content and approval of final article for submission.

Data sharing statement: This study is based on the Malawi 2004 and 2010 Demographic and Health Surveys. These data are available on www.measuredhs.com. Access to individual HIV status as well as individual background information required authorization from MACRO ORC (www.measuredhs.com). The Website depicts the process.

Funding

The Luxembourg National Research Fund (FNR).

Competing Interests

None

References

[1] United Nations.*The Millennium Development Goals. 2012 Report* . New York: United Nations, 2012.

[2] UAIDS. *World AIDS day report 2011. How to get to zero: Faster, smarter, Better.* Geneva: UNAIDS, 2011.

[3] Durevall, D, & Lindskog, A. *HIV/AIDS, Adult Mortality and Fertility: Evidence from Malawi. Working Papers in Ecomics, No 284.* Göteborg: School of Business, Economics and Law, University ofGothenburg, 2007.

[4] Rombo, D. *Marital risk factors and HIV infection among women: A comparison between Ghana and Kenya,* Ph.D. Dissertation. Minneapolis: University of Minnesota, 2009.

[5] Öjteg, K. *Socio-Economic determinants of HIV in Zambia. A district level Analysis*. Ph.Dissertation. Lund: Department of Economics at the University of Lund, 2009.

[6] Magadi, M., & Desta , M. A multilevel analysis of the determinants and cross-national variations of HIV seropositivity in sub-Saharan Africa: Evidence from the DHS. *Health & Place*, 2011; 17 (5): 1067–83.

[7] Magadi, M. Understanding the gender disparity in HIV infection across countries in sub-Saharan Africa: evidence from the Demographic and Health Surveys. *Sociology of Health & Illness,* 2011; 33 (4): 522-39.

[8] Asiedu, C., Asiedu, E., & Owusu, F. The Socio-Economic Determinants of HIV/AIDS Infection Rates in Lesotho, Malawi, Swaziland and Zimbabwe. *Development Policy Review ,* 2012; 30 (3): 305-326 .

[9] Green, E. *Rethinking AIDS Prevention. Learning from Successes in Developing Countries* . Westport, CT: Praeger publishers, 2004

[10] International HIV / AIDS Alliance. *The cost efficiency of HIV prevention for vulnerable and most-at-risk populations and the reality of funding. What's Preventing Prevention Campaign Briefing 2.* Hove: International HIV / AIDS Alliance, 2010.

[11] Population Reference Bureau, *2012 World Population Data sheet*, Washington DC, PRB, 2012.

[12] United Nations Development Programme/UNDP. Sustainability and Equity: A better future for all, Human Development Report 2011. New York, UNDP, 2011.

[13] Malawi National Statistical Office (NSO); ORC Macro. *Malawi Demographic and Health Survey 2004.* Calverton, Maryland: NSO and ORC Macro, 2005.

[14] Malawi National Statistical Office (NSO); ICF Macro. *Malawi Demographic and Health Survey 2010.* Zomba, Malawi, and Calverton, Maryland, USA: NSO and ICF Macro, 2011.

[15] Kass, G. V. An Exploratory Technique for Investigating Large Quantitaties of Categorical Data. *Applied Statistics,* 1980; 29 (2): 119-127.

[16] Kitsantas, P., Hollander M., & Li, L. Using classification trees to assess low birth weight outcomes. *Artificial Intelligence in Medicine*. 2006; 38 (3): 275-289.

[17] Adair, T. *HIV Status and Age at First Marriage among Women in Cameroon.* DHS Working Paper No 33, Calverton, Maryland, USA: Macro International Inc, 2007.

[18] Rodrigo, C., & Rajapakse, S. (2010). HIV, poverty and women. *International Health*, 2(1): 9-16.

[19] Takyi, B., & Broughton, C. Marital Stability in Sub-Saharan Africa: Do Women's Autonomy and Socioeconomic Situation Matter? *Journal of Family and Economic Issues* , 2006; 27 (1): 113-132.

[20] Reniers, G. Divorce and remarriage in rural Malawi. *Demographic Research*, 2003; Special Collection 1, article 6: 175-206. www.demographic-research.org/special/1/6/S1-6.pdf.

[21] Reniers, G. Marital strategies for regulating exposure to HIV. *Demography*, 2008; 45 (2): 417-438.

Identifying HIV most-at-risk groups in Malawi for targeted interventions

Jacques B. O. Emina [1,2], Nyovani Madise [3], Mathias Kuepie [2], Eliya M. Zulu [4], Yazoume Ye [5]

1 Department of Population and Development Studies, University of Kinshasa, P.O.BOX 176 Kinshasa XI, Democratic Republic of Congo

2 CEPS/INSTEAD 3, avenue de la Fonte, L-4364 Esch-sur-Alzette Luxembourg

3 Faculty of Social and Human Sciences, Room 4103 Building 58, University of Southampton, Southampton, SO17 1BJ.

4 African Institute for Development Policy (AFIDEP), P. O. Box 14688-00800, Westlands, Nairobi, Kenya.

5 ICF International 11785 Beltsville Drive, Suite 300 Calverton, MD 20705, USA.

Correspondence to: Jacques B.O. Emina; jacques.emina@gmail.com

Abstract

Objectives: To identify HIV socioeconomic predictors as well as identify the most-at-risk groups of women in Malawi.

Design: cross-sectional survey

Setting: Malawi

Participants: The study used a sample of 6,395 women age 15-49 years from the 2010 Malawi Health and Demographic Surveys (MDHS).

Interventions: N/A

Primary and secondary outcome measures: Individual HIV status: positive or not.

Results: Findings from Pearson Chi-square and Chi-square Automatic Interaction Detector (CHAID) analyses revealed that marital status is the most significant predictor of HIV. Women who are no longer in union and living households in the highest wealth quintiles constitute the most-at-risk group; whereas the less-at-risk group includes young women (15-24) never married or in union and living in rural area.

Conclusion: In the light of these findings, this study recommends: (1) design and implementation of targeted interventions taking into account HIV prevalence and the demographic size of different groups at risk groups. Preventive interventions should prioritize couples and never married people age 25-49 living in rural areas because this group accounts for 49 percent of the study population and 40 percent of women living with HIV in Malawi; (2) With reference to treatment and care, higher priority must be given to promoting HIV test, monitoring and evaluation of equity in access to treatment among women in union disruption and never married or women in union age 30-49 living in urban areas; (3) Community health workers, households based campaign, reproductive health services and reproductive health courses at school could be used as canon to achieve universal prevention strategy, testing, counseling and treatment.

Keywords: HIV/AIDS, Malawi, CHAID methods, decision analysis, most-at-risk groups, targeted interventions.

Article summary

1. Article focus

- Targeted interventions and evidence based prevention programmes have been advocated as cost-effective strategy to combat HIV/AIDS.
- Who are the most-at-risk populations regarding HIV prevalence in Malawi? With HIV prevalence of about 14 percent among women of reproductive age, HIV/AIDS constitutes a drain on the labor force and government expenditures in Malawi.

2. Key messages

- We use data from the Malawi 2010 Demographic and Health Surveys to profile HIV most-at-risk groups of women in Malawi where about 14 percent of women are HIV positive.
- Our findings revealed that richest and formerly in union women is the most-at-risk group.
- To achieve zero new infection as part of MDG 6, there is need of more comprehensive policy to combat HIV because of the complexity of HIV socioeconomic profile in Malawi. There are several groups built from several socioeconomic categories depending on individual marital status, wealth index, age, place of residence, and relationship to the head of household.

Strengths and limitations of this study

- From our knowledge this study may be the first in Malawi to attempt to profile HIV most-at-risk groups of women in Malawi. The most-at-risk population refers to a combination of several factors because factors associated with HIV are not mutually exclusive.
- The major strength is the use of the Chi-square Automatic Interaction Detector (CHAID) to identify HIV predictors and the most-at-risk groups among women for intervention. CHAID uses regression and classification algorithms and offers a non-algebraic method for partitioning data that lends itself to graphical displays.
- This study has two major limitations. First, this study used cross-sectional data from the Demographic and Health Surveys, which does not permit one to draw causal association between HIV status and the associated factors. For instance, whether HIV infection has occurred before, during or after the union. Last, CHAID model ignores the hierarchical structure of the Demographic and Health Survey data and need large sample size.

Introduction

In 2000, the United Nations' Millennium summit identified the reduction of HIV prevalence as one of the eight fundamental goals for furthering human development. Though global HIV/ AIDS incidence is declining, HIV/AIDS has remained the

leading cause of death in women of reproductive age in low-and middle-income countries, particularly in sub-Sahara Africa (SSA) [1]. The gap between the current state of HIV/AIDS and the UNAIDS goals of three zero (zero new HIV infections, zero discrimination, and zero AIDS related deaths) remains important. With barely two years remaining to the end-date of the Millennium Development Goals (MDG) target, HIV/AIDS remains a long-term global challenge [1].

Given the high cost of HIV/AIDs treatment estimated in 2010 to be globally between US$22 and US$24 billion annually by 2015 and individual cost of US $4,707 over lifetime to reach global targets [2,3], targeted interventions and evidence based prevention programmes have been advocated as cost-effective strategy to combat HIV/AIDS. Such a strategy reduces levels of vulnerability and risk as well as allowing HIV interventions to optimize coverage, reducing costs and lowering the number of new infections [4]. In the United State Virgin Islands, the recommended strategy of universal screening by 14 weeks gestation and screening the infant after birth has a cost savings of $1,122,787 and health benefits of 310 life year gained [5]. A prevention of Mother-to-Child Transmission intervention in Capet town, South Africa, revealed that a program at a scale sufficient to prevent 37 percent of pediatric HIV infections would cost about US$0.34 per person in South Africa and would be affordable to the health care system [6].

In Indian high HIV prevalence southern states, targeted interventions result in significant decline in HIV prevalence among female commercial sex workers (CSWs) and young pregnant women [7]. Evaluation of the cost-effectiveness of the female condom (FC) in preventing HIV infection and other sexually transmitted Diseases (STDs) among CSWs and their clients in the Mpumulanga Province of South Africa, showed that a well-designed FC program oriented to CSWs and other women with casual partners is likely to be highly cost-effective and can save public sector health payer US $12,090 in averted HIV/AIDS treatment costs in rural South Africa [8]

Likewise, analysis of targeting Voluntary HIV Counseling and Testing in Kenya and in Tanzania showed that increasing the proportion of couples to 70 percent reduces the cost per disability-adjusted life year (DALY) saved to $10.71 in Kenya and $13.39 in Tanzania, and that targeting a population with HIV-1 prevalence of 45 percent decreased the cost per DALY saved to $8.36 in Kenya and $11.74 in Tanzania [9].

However despite growing literature in health and social sciences on factors associated with HIV/AIDS during the last three decades, less is known about the most-at-risk populations regarding HIV prevalence [10, 11, 12, 13,14, 15]. Indeed, whereas in countries with concentrated HIV/AIDS epidemics (Latina America, East Asia and Eastern Europe), the most-at-risk populations including CSWs, long distance truck drivers, men who have sex with men, and unmarried youth [16,11,17] account for a large proportion of new infections, in countries with high prevalence, they account only for a smaller share of new infections [17].

Against this background, this study aims to identify HIV socioeconomic predictors as well as identify the most-at-risk groups among women in Malawi. With HIV prevalence of about 13.6 percent among women of reproductive age [18], HIV/AIDS constitutes a drain on the labor force and government expenditures in Malawi.

Data and Methods

Study setting

The Republic of Malawi is a landlocked country in southeast Africa. Malawi is over 118,000 km^2 with an estimated population of about 16 millions [18]. Its capital is Lilongwe, which is also Malawi's largest city; the second largest is Blantyre and the third is Mzuzu.

Malawi is among the world's least-developed countries. The economy is heavily based on agriculture, with a largely rural population. The country Gross National Income (GNI) per capita at purchasing power parity (PPP) is estimated at $860 while the world average is estimated at $10,780 [18, 19]. Ninety-one percent of Malawians live below 2 dollars (US) per day. The country's Human Development Index is estimated at 0.400, which gives the country a rank of 171 out of 187 countries with comparable data [19].

Malawi has a low life expectancy (53 years) and high infant mortality (66 deaths per 1,000 live births) compared to the world' average (70 years and 41 deaths per 1,000 live births). Averages for sub-Saharan Africa are estimated respectively at 55 years and 80 deaths for 1,000 live births. There is a high prevalence of HIV/AIDS, especially among women with about 13.6 percent HIV positive [18].

Malawi has actively responded to HIV since 1985 after the first AIDS case was

reported. In 1988, the government created the National AIDS Control Program to coordinate the country's HIV/AIDS education and prevention efforts. The Public Sector continues to set aside a minimum of two percent of their recurrent budget to support HIV and AIDS programme [20]. The HIV national commission budget has increased from US $98.1 million in 2010 to US $113.51 million in 2011 [20]. According to the Malawi 2012 Global AIDS Response progress report:

- Distribution of leaflets and HIV radio and TV programs. During the 2010-2011 financial year, 1,477 radio and 429 television (TV) programs were produced.
- In 2010 and 2011, around 3.8 million young people (50 percent males and 50 percent females) have been trained on life skills education (LSE) each year.
- Since 2003, the number of condoms distributed per capita has been increasing. Cumulatively, 21,049,592 condoms were distributed in the 2009-2010 fiscal year. During the fiscal year 2010-2011 the annual cumulative total of 26,461,079 condoms were distributed.
- The number of sites providing Prevention of Mother to Child Transmission (PMTCT) services has also been increased from152 facilities in 2006 to 544 sites 2011.
- Antiretroviral Therapy has been provided free of charge in the public sector since 2004. Number of patients alive and on treatment has increased from 10,761 in 2004 to 322,209 in 2011.

Data sources

This study uses data from the 2010 Malawi Health and Demographic Surveys (MDHS). The inclusion of HIV testing in the 2010 MDHS offers the opportunity to identify socioeconomic profile of women age 15-49 living with HIV. Participation in HIV testing was voluntary. To ensure confidentiality, case numbers (and not names) were used in linking the HIV test results to individual and household characteristics.

A subsample of one-third of the households was selected to conduct HIV testing for eligible women age 15-49 years. Ninety percent of all 2010 MDHS women who were eligible (8,174) for testing were interviewed and consented to HIV tests. The principal mode of HIV transmission in Malawi is heterosexual contact; therefore our analyses

focus on 6,395 women who ever had sexual intercourse. Details on the sample design are provided elsewhere [21,22].

Variables

The dependent variable for this analysis is HIV status, characterized as positive or negative blood test. The independent variables include 12 main variables grouped into two major types including, demographic and reproductive behavior variables (age, age at first sex, marital status, age at first birth, number of children ever born, Experience in premarital childbearing, and relationship to the head of household), and socioeconomic and contextual variables (religion, region of residence, place of residence, education, and household wealth index).

The choice of these variables is guided by the literature on factors associated with HIV in sub-Saharan Africa [10-15]. Most-at-risk populations refer to a combination of several factors because socioeconomics factors associated with HIV are not mutually exclusive.

Statistical analyses

Statistical analyses used Pearson Chi-square and Chi-square Automatic Interaction Detector (CHAID) using SPSS version 21. We used weighted data to take into account the complexity of the DHS design. We performed Pearson chi-square to identify associations between the HIV status (positive, negative) and demographic and reproductive behavior variables as well as socioeconomic and contextual variables.

We used CHAID to identify HIV predictors and the most-at-risk groups among women living with HIV [23]. CHAID is a nonparametric technique that makes no distributional assumptions on outliers, collinearities, heteroskedasticity, or distributional error structures. The dependent variable and predictor variables can be nominal (categorical), ordinal (ordered categories ranked from small to large), or interval (a "scale").

CHAID uses regression and classification algorithms and offers a non-algebraic method for partitioning data that lends itself to graphical displays. The method is a sequential fitting algorithm and its statistical tests are sequential with later effects being dependent upon earlier ones, and not simultaneous as would be the case in a regression model or analysis of variance where all effects are fit simultaneously.

CHAID solves the problem of simultaneous inference using Bonferroni multiplier. It automatically tests for and merges pairs of homogenous categories in independent variables.

At each step, CHAID chooses the independent (predictor) variable that has the strongest interaction with the HIV status (dependent variable). The variable having the strongest association with HIV status becomes the first branch in a tree with a leaf for each category that is significantly different relative to be HIV positive. It then assesses the category groupings, or interval breaks to pick the most significant combination of variables. The process is repeated to find the predictor variable on each leaf most significantly related to HIV status, until no significant predictors remain.

The developed model is a classification tree (or data partitioning tree) that shows how major "types" formed from the independent (predictor or splitter) variables differentially predict a criterion or dependent variable. The method permits also identification of women who are likely to be members of a particular group (Segmentation), and assign cases into one of several categories, such as high-, medium-, and low-risk groups (stratification). Selecting a useful subset of predictors from a large set of variables for use in building a formal parametric model (Data reduction and variable screening); Identify relationships that pertain only to specific subgroups and specify these in a formal parametric model (Interaction identification); and recoding group predictor categories and continuous variables with minimal loss of information. Categories of each predictor are merged if they are not significantly different with respect to the dependent variable (Category merging and discretizing continuous variables).

The output of CHAID prediction model is displayed in hierarchical tree-structured form, and consists of several levels of branches: root node, parent nodes, child nodes and terminal nodes. The root node, “Node 0” or “initial node” is the dependent variable or the target variable, HIV prevalence in our case. Parent node is the upper node compared with nodes on the subsequent (lower) level, whereas the lower level nodes are called child nodes. The terminal node or external node is any node that does not have child nodes. It is the last categories of the CHAID analysis tree.

For each terminal node CHAID provides in a table the following indicators:

1. Node: provides the number and percentage of people belonging to a selected category j (demographic weight in the sample);
2. Gain for each terminal node is the number of women who are living with HIV in absolute. In percentage, gain is calculated as number of women living with HIV in a selected node divided to the total of women living with HIV. Part of the population with the observed characteristic (living with HIV) in a selected category compared to total of women living with HIV.
3. Response can be defined as HIV prevalence among women belonging to each terminal node. Number of women living with HIV in a selected node divided by total of women of the node.
4. Gain index percentage reporting how much greater the proportion of a given target category at each node differs from the overall proportion. It is obtained by dividing the proportion of records that present category j in each terminal node into the proportion of records presenting category j in the total sample. Thus, it represents the increased probability of belonging to the selected category j that contains the records presenting the characteristics defined for each terminal node.

The method allows: (1) identifying complex interactions between variables across the measurement space; (2) Identifying the most significant explanatory variable; (3) Merge categories of nominal variable and categorize continuous variables without loss of information. Furthermore, CHAID as other decision trees can be applied to any data structure.

However, CHAID have two major shortcomings. First the method needs large sample sizes to work effectively because it uses multi-way splits. Indeed, with small sample sizes the respondent groups can quickly become too small for reliable analysis. Last, CHAID does not take into account the hierarchical structure of this data.

Results

Sampling description

Table 1 presents the characteristics of the study population. Since the principal mode of HIV transmission in Malawi is heterosexual contact, our analyses focus on women who ever had sexual intercourse. The distribution of the sample by age shows that

more than half (56 percent) of the populations are age less than 30 years old. The average age of the sample is estimated at 29 years old. Women who are in union (i.e., currently married or living with a man) constitute about 77 percent. The proportion of women who have never been married is estimated at 8 percent. Regarding the relationship to the head of household, the majority of women are spouse (63 percent).

Table 1 also shows that the majority of women (more than 80 percent) live in rural areas. By region, the majority of women live in Southern Region and the Central Region. Furthermore, 17 percent of women never attended school, while more than 60 percent have attended only primary school. Regarding the reproductive behavior, a large majority of women had their first sexual intercourse before 20 years (average 16.6 years old).

Table 1 – Description of the sample

Socioeconomic and demographic Characteristics	Percentage		Number[2]
	Weighted[1]	Unweight	
Age			
15-19	11.5	11.6	744
20-24	21.5	20.8	1,327
25-29	22.7	21.9	1,402
30-34	15.2	15.7	1,001
35-39	13.0	12.7	814
40-44	8.6	9.1	579
45-49	7.5	8.3	528
Average	29.6	29.8	-
Age at first sex			
<15	19.1	19.2	1,230
15-19	68.5	68.4	4,376
20-24	11.2	11.1	708
25+	1.3	1.3	81
Average	16.6	16.6	16.6
Marital status			
Single	7.5	7.6	484
In union	77.4	77.1	4,929
Ever married	15.1	15.4	982
Number of ever born children			
0	9.9	9.6	617
1+	90.1	90.4	5,778
Age at first birth			
Never give birth	10.3	10.6	660
< 20 years old	64.8	64.2	4,144
20 +	24.9	25.2	1,591
Ever had premarital child			
No	88.7	88.4	5,652
Yes	11.3	11.6	743
Relationship to the head of household			
Head of household	19.4	19.0	1,213
Spouse	62.6	62.4	3,992
Daughter & Grand daughter	11.0	11.7	748
Others	7.1	6.9	442
Province of residence			
Northern	11.1	17.5	1,122
Central	42.2	34.1	2,181
Southern	46.7	48.4	3,092
Place of residence			
Urban	19.2	13.1	837
Rural	80.8	86.9	5,558
Religion			
Catholic	21.2	20.6	1,316
Protestant	24.3	25.2	1,610
Other Christians	39.7	42.3	2,708
Muslim	13.5	10.9	695
Others	1.3	1.0	66
Education			
None	17.5	16.6	1,060
Primary	63.8	66.4	4,246
Secondary +	18.7	17.0	1,089
Household wealth quintiles			
Poorest	17.6	19.0	1,215
Poorer	20.1	20.6	1,319
Middle	19.7	20.9	1,334
Richer	19.3	20.7	1,323
Richest	23.3	18.8	1,204
Total			6,395

[1] Interpretations are based on weighted percentage.
[2] Unweight crude numbers

HIV prevalence by selected background characteristics

Table 2 describes HIV prevalence in Malawi by women's selected background characteristics. Overall, 14 percent of the women are HIV positive. All independent variables are statistically associated with HIV infection status except for religion.

HIV infection prevalence was high (20 percent) among women aged 30-39 years. Women who are no longer in union (widowed, divorced and separated) had significantly higher prevalence (4 percent) compared to those who have never been in a marital union (11 percent). HIV prevalence was high among heads of household (25.0 percent). Furthermore, while 25 percent of women in urban area were HIV positive, the prevalence was less than half (12 percent) compared to their counterparts from the rural areas. The HIV epidemic shows regional heterogeneity with a higher prevalence (20 percent) observed in the Southern region. Women with secondary education had higher HIV prevalence compared to those who never attended school (17 percent versus 14 percent). Regarding the household wealth quintiles the prevalence of HIV infection is higher among the women from the highest quintiles. With reference to sexual and reproductive behavior, HIV prevalence was higher among women who had their first sexual intercourse before the 15th birthday or from their 25th birthday, and /or who have experienced premarital childbearing.

Table 2 – HIV prevalence by selected socioeconomic and demographic characteristics

Socioeconomic and demographic Characteristics	HIV+ Prevalence (%)	Total (N)	Chi-Square	p-value
Age				
15-19	5.0	744		
20-24	6.9	1,327		
25-29	12.4	1,402		
30-34	19.8	1,001	190.35	<0.001
35-39	21.4	814		
40-44	18.7	579		
45-49	16.9	528		
Age at first sex				
<15	15.9	1,230		
15-19	13.1	4,376	9.13	0.028
20-24	12.4	708		
25&+	18.5	81		
Marital status				
Single	7.9	484		
In union	10.7	4,929	316.15	<0.001
Ever married	31.5	982		
Number of ever born children				
0	7.9	617	18.80	<0.001
1&+	14.2	5,778		
Age at first birth				
Never give birth	9.1	660		
< 20 years old	14.1	4,144	12.96	0.002
20 +	14.3	1,591		
Ever experience premarital childbearing				
No	12.8	5,652	26.99	<0.001
Yes	19.8	743		
Relationship to the head of household				
Head of household	25.0	1,213		
Spouse	10.2	3,992	179.93	<0.001
Daughter & Grand daughter	11.9	748		
Others	17.0	442		
Region of residence				
Northern	10.0	1,494		
Central	9.5	3,062	184.90	<0.001
Southern	20.0	4,444		
Place of residence				
Urban	24.7	1,156	157.00	<0.001
Rural	12.3	7,844		
Religion				
Catholic	12.6	1,316		
Protestant	14.3	1,610		
Other Christians	13.4	2,708	2.66	0.616
Muslim	14.8	695		
Others	13.6	66		
Education				
None	13.9	1,060		
Primary	12.8	4,246	10.73	0.005
Secondary +	16.6	1,089		
Household wealth quintiles				
Poorest	10.3	1,215		
Poorer	10.9	1,319		
Middle	11.6	1,334	88.34	<0.001
Richer	14.3	1,323		
Richest	21.5	1,204		
Total	13.6	6,395		

HIV predictors in Malawi: results from CHAID analysis

Table 3 shows summary information on the specifications used to build the CHAID model and the resulting model. Twelve independent variables were specified, but only six were included in the final model. The variables such as age at first sex, age at first birth and female education did not make a significant contribution to the model, so they were automatically dropped from the final model. Overall, there are 21 nodes among which 13 terminal nodes. Parent nodes include at least 100 cases whereas child nodes account for 50 cases in minimum.

Model components	Model specification	Results
Dependent variable	HIV status	HIV+=13.6%
Independent Variables	Age, Age at first sex, Marital status, Ever had a child, age at first birth, Experience premarital childbearing, Relationship to the head of household, Region of residence, Place of residence, Education, Wealth Index, Religion	**Marital status**, Age, Wealth Index, Relationship to the head of household, Region of residence, Place of residence
Maximum Tree Depth	3	3
Minimum Cases in Parent Node	100	100
Minimum Cases in Child Node	50	50
Number of Nodes	-	21
Number of Terminal Nodes	-	13
Overall predicted correct percentage		86.8

The tree diagram depicted in Figure 1 shows that "Marital status" is the best predictor of HIV status among women in Malawi (Chi-square = 313.22, p-value<0.0001).

The tree is split into two main notes. Node 1 includes women formerly in union; and Node 2 is composed of women in union and never married women.

Node 1- Women formerly in union. For this group, including divorced, widowed, and not living together, age is the best predictor of HIV prevalence (Chi-square=56.30, p-value <0.001). The group is further splits into four sub age groups: 15-24, 25-29, 30-34 and 40-49, and 35-39.

Among women aged 30-34, 40-44 and 40-49 (Node 3) with HIV prevalence of 37 percent, household wealth quintiles are the best predictor of HIV infection (Chi-square=29.81, p-value<0.001). Indeed in this group women in the highest wealth

quintile (Node 9) have HIV prevalence about three times higher than their counterpartners from the lowest wealth quintile-Node 11 (60 percent versus 22 percent). For women in the age groups of 14-19 and 20-24 years (Node 4) with a HIV prevalence of 13 percent, the relationship to the head of household is the best predictor of HIV infection (Chi-square=11.1, p-value <0.003) whereby women head of household (Node 12) have higher HIV prevalence compared to other women with different relationship to the head of household – Node 13 (24 percent versus 7 percent). The region of residence is the best predictor of HIV infection among women aged 35-39 (Chi-square=11.5, p-value<0.002) with women living in the Southern region (Node 14) having HIV prevalence about twice the one of the women from the Central and Northern regions-Node 15 (59 percent versus 31 percent). Among women aged 25-29 years (Node 6) accounting for 3 percent of the study population with HIV prevalence of 27 percent, age remains the only significant and final predictor of HIV prevalence.

Node 2- This group includes women in union (married or living together) and those have **never been in union**, representing 85 percent of the study population and have HIV prevalence of 10 percent. Place of residence (rural or urban) is the best predictor of HIV infection with a higher prevalence in urban area (Node 8) compared to rural area – Node 7 (21 percent vs 9 percent, Chi-square = 89.8, p-value<0.001).

For the women living in the rural area (Node 7) and representing 74 percent of the population, the best predictor for HIV infection is age (Chi-square=86.0, p-value <0.001) with the highest prevalence among women aged 30-44 years (13 percent) followed by the age group 25-49 (Node 18: 9 percent) and the age group 15-24 (Node 18: 4 percent). Similarly, age is strong predictor of HIV infection (Chi-square=86.0, p-value <0.001) among women living in urban area (Node 8) whereby the age group 30-49 (Node 19) has a prevalence about twice the one among the age group 15-29-Node 20 (29 percent versus 15 percent).

HIV risk groups in Malawi

There are in total 13 homogenous subgroups or terminal nodes. Table 4 describes these 13 subgroups (terminal nodes) in terms of their composition, demographic weight in the sample (columns 1 and 2), their share in HIV burden (columns 3 and 4)

and their corresponding HIV prevalence (column 5). The 13 homogenous sub-groups could be grouped into 5 major groups.

Group 1 represents 3 percent of the sample with an overall HIV prevalence of 59 percent. This group accounts for 11 percent of all the women HIV positive. Group 1 includes two subgroups: a) women in union disruption living in richest household and age 30-34 or 40-49 years old; and b) women in union disruption living in the Southern region and age 35-39 years old.

Group 2 represents 5 percent of the sample with an overall HIV prevalence of 35 percent and accounts for 12 percent of all HIV positive women. This group is composed of two subgroups including women in union disruption living in intermediate wealth households (non-poorest and non-richest households) age 30-34 or 40-49; and women in union disruption age 35-39 and living in the Northern or Central region.

Group 3 represents about 10 percent of study population with an overall HIV prevalence of 27 percent and accounts for 20 percent of all HIV positive women. This group is divided into four subgroups: a) Never married and women in Union, living in urban area and age 30-49; b) Formerly in union (widowed or divorced) women age 25-29; c) Young women (15-24) formerly in union who are head of household; d) Formerly in union women living in poorest household and age 15-24, 30-34 or 40.

Group 4 represents about 33 percent of the study population with an overall HIV prevalence of 14 percent and account for 33 percent of all the HIV positive women. This group includes adolescent (15-19), never married women or in union living in urban area; and never married or women in union living in rural area age 30-44.

Group 5 represents 50 percent of the study population and has the lowest HIV prevalence of 7 percent, but account for 23 percent of all the HIV positive women. This group includes three subgroups: a) Never married or women in union living in rural area and age 25-29 or 45-49; b) Young women age 15-24 who are non longer in union and are not head of household; and c) Young women (15-24) who are never married or in Union and are living in rural areas.

Table 4 reports also the gain index percentage (column 6) expressing how much greater the proportion of a given target group at each node differs from the overall proportion. The index percentage is very high among women belonging to group with

high HIV prevalence but with small demographic weight in the population (categories 1 to 3). Opposite values are observed among groups accounting for the major part of the sample among which HIV prevalence is low (Group 5). The Index is equal to 100 in category 4.

Table 4 – CHAID Gains for Nodes

Group/Node No	Node description	Node		Gain		% HIV[5]	Index[6]
		N[1]	%[2]	N[3]	%[4]		
Group 1		165	2.6	98	11.2	59.4	435.6
9	Formerly in union, 30-34/ 40-49 and richest	82	1.3	49	5.6	59.8	438.2
14	Formerly in union, 35-39 and Southern region	83	1.3	49	5.6	59.0	433.0
Group 2		308	4.8	108	12.4	35.1	257.2
10	Formerly in union, 30-34/40-49 and Rich/Middle/Poor	246	3.8	89	10.2	36.2	265.3
15	Formerly in union, 35-39 and North/Central regions	62	1.0	19	2.2	30.6	224.7
Group 3		640	10	171	19.5	26.7	195.9
19	Never married/in Union, living in urban area, 30-49	268	4.2	78	8.9	29.1	213.4
6	Formerly in union and 25-29	180	2.8	49	5.6	27.2	199.6
12	Formerly in union, 15-24 and Head of household	67	1.0	16	1.8	23.9	175.1
11	Formerly in union, 15-24, 30-34/40-49 and Poorest	125	2.0	28	3.2	22.4	164.3
Group 4		2,117	33.1	291	33.3	13.8	100.8
20	Never married/in Union, living in urban area, 15-19	417	6.5	64	7.3	15.3	112.6
16	Never married/in Union, living in rural area, 30-44	1,700	26.6	227	26.0	13.4	97.9
Group 5		3,165	49.5	204	23.3	6.5	42.3
18	Never married/in Union, rural area, 25-29 and 45-49	1,404	22.0	126	14.4	9.0	65.8
13	Formerly in union, 15-24 and Not Head of household	137	2.1	10	1.1	7.3	53.5
17	Never married/in Union, living in rural area, 15-24	1,624	25.4	68	7.8	4.2	30.7
Total		6,395	100	872	100	13.6	-

Notes:;[1] Number of cases per node (demographic size in the sample);
[2] Demographic size in percentage = (.1/Σ.1)*100;[3] Number of HIV women;[4] Demographic size among HIV positive women in percentage = (.3/Σ.3)*100; [5] HIV prevalence in percentage = (.3/Σ.1)*100; [6] Node Index = ((.3/Σ3)/(.1/Σ.1))*100

Discussions

This paper aimed to describe and profile HIV prevalence among women in Malawi. The study used Chi-square and CHAID techniques to analyze data from the Malawi 2010 DHS.

Analyses suggested three keys findings. First, consistent with previous studies [13, 24], findings from bivariate analysis and chi-square test showed high HIV prevalence among women in union dissolution, among the most educated women, women living

in wealthy households and/or among women living in urban areas. The finding confirmed also region heterogeneity in HIV prevalence. The Southern region being the most affected. In general the most educated women are more likely to marry husbands with high education level, and belonging to high socio-economic class of the society [25,26]. In parallel, relatively rich and better-educated men have higher rates of partner change because they have greater personal autonomy and spatial mobility [27, 28, 29, 30]. Women's economic dependence on their partners may also make it difficult for them to insist on safer sex (e.g. condom use). Concentration (about 50 percent) of the most educated, richest and urban women in the Southern region may explain high HIV prevalence in that region.

Second, results from CHAID models reported that marital status is the best predictor of HIV status among women in Malawi. Non-poorest women who are no longer in union (widowed and divorced or separated) age 30-34 or 40-49 have significantly higher HIV prevalence. This may be because: (1) husband from the highest quintile or a male partner may have more access to transactional sex and other risk behaviors such as polygamy which may increase women's vulnerability to HIV; (2) wealthier HIV positive widowed may have better quality of life as well as better access to treatment and survive longer [31]. Furthermore, divorced and separated are more frequent among the most educated women with economic autonomy [32]. Their causes (polygyny and/or infidelity) as well as consequences (multiple sexual partnerships) are also factors associated with HIV prevalence [33,34].

Third and last, CHAID models depicted also different interactions between risk factors and profiled HIV risk groups in Malawi. For instance, whilst overall HIV prevalence among women living in urban areas (25 percent) is twice the prevalence observed among women living in rural areas (12 percent), HIV prevalence is estimated at 15 percent among never married or women in union living in urban areas age 15-29, and at 13 percent among never married or women in union living in the rural areas age 30-44. Likewise, whereas in general HIV prevalence is low among never married and women in union (10 percent), CHAID results revealed a higher HIV prevalence (29 percent) among never married and women in union age 30-49 who live in urban areas compared to: (1) women in union disruption age 15-24 (7 percent if they are not head of households and 23 percent for head of household); (2) women in union disruption age 25-29 (27 percent); and (3) women in union

dissolution age 30-34 and 40-49 who live in poorest households (22 percent).

These findings showed the complexity of different interactions that may present challenges to conventional regression models. Indeed, CHAID is a sequential fitting algorithm and its statistical tests are sequential with later effects being dependent upon earlier ones and not simultaneous as would be the case in a regression model or analysis of variance where all effects are fit simultaneously. Furthermore, CHAID allows automatic detection of interaction between variables.

In the light of these findings, it is noteworthy that to reduce number of new infection, interventions should be targeted and prioritized according to the prevalence and demographic size of different risk groups. Furthermore, policy makers' prioritization of interventions may depend also on preference for preventive interventions compared to treatment of and care for HIV infected people and/or to treatment of and care for AIDS-patients. In Thailand, for instance policy makers expressed a preference for target preventive interventions that are highly effective compared to care and treatment [35].

Regarding preventive interventions, the findings suggested that:

1. Couples (males and women in union) and never married people age 25-49 (nodes 16 and 18) living in rural areas should be the first targets using universal HIV testing, "Abstinence", "Be faithful" and "use condom" campaign. Indeed, this group includes 49 percent of the study population, among whom the HIV prevalence is estimated at 11 percent on average. About 40 percent of women living with HIV in Malawi belong to this category.
2. Young age 15-24 living in rural areas (node 17) and urban adolescent 15-19 (node 20) is the second most important target. This group account for 32 percent of the studied population and 15 percent of women living with HIV. Besides, majority of adults living with HIV may be infected during adolescence. Unfortunately, the available dataset could not provide information on time of infection.
3. The country develops and implements a social policy to protect single mothers. Indeed, though overall HIV prevalence is estimated at 6 percent on average among young women age 15-24 (Table 2) that prevalence is estimated above 20 percent among young women formerly in union and among young

women who ever experience premarital childbearing living in urban areas. Likewise, HIV prevalence is very high among women in union disruption (32 percent on average) compared to other groups (10 percent). Though this high prevalence may be due to male mortality, some women in union disruption may be vulnerable because of poverty.

With reference to treatment and care, higher priority must be given to promoting HIV test, monitoring and evaluation of equity in access to treatment among women in union disruption and never married or women in union age 30-49 living in urban areas. Indeed, formerly in union women represent only about 13 percent of women of reproductive age in Malawi, they have the higher HIV prevalence ranges between 22 percent observed among poorest and 60 percent among richest.

Nevertheless, to achieve zero new infection as part of MDG 6, there is need of more comprehensive policy to combat HIV because of the complexity of HIV socioeconomic profile in Malawi. There are several groups built from several socioeconomic categories depending on individual marital status, wealth index, age, place of residence, and relationship to the head of household. In South Africa, Bendavid et al. [36] revealed that scaling up all aspects of HIV care including universal testing and treatment was associated with a life expectancy gain of 22.2 months, and new infections were 73 percent lower.

From the methodological point of view, this study has some limitations, which do not detract from its scientific importance and contribution. First, this study used cross-sectional data from the Demographic and Health Surveys, which does not permit one to draw causal association between HIV status and the associated factors. For instance, whether HIV infection has occurred before, during or after the union. Last, CHAID model ignores the hierarchical structure of the Demographic and Health Survey data and need large sample size.

In conclusion, this study recommends: (1) design and implementation of targeted interventions taking into account HIV prevalence and the demographic size of different groups at risk; (2) reinforcement of integration of family planning and HIV/AIDS services because the population understudied includes women of reproductive health. Integrating the two services (HIV and FP) could be cost-effective; (3) Community health workers, households based campaign, reproductive

health services and reproductive health courses at school could be used as canon to achieve universal prevention strategy, testing, counseling and treatment.

Acknowledgements: *This research was supported by the Luxembourg National Research Fund (FNR). The authors thank Macro international for providing free the Malawi 2010 DHS data-sets including HIV data. We thank Prof. Guillaume Wunsch for his comments. Furthermore, we* have appropriate approval and authorization from the DHS datasets research ethics board for conducting HIV secondary data analysis.

Contributors: JBOE participated in conception and design, literature review, data analysis and interpretation, drafting the article, critical revisions for important intellectual content and approval of final article for submission. MK, NJM, EMZ and YY participated in conception and design, interpretation of results, critical revisions for important intellectual content and approval of final article for submission.

Data sharing statement: This study is based on the Malawi 2004 and 2010 Demographic and Health Surveys. These data are available on www.measuredhs.com. Access to individual HIV status as well as individual background information required authorization from MACRO ORC (www.measuredhs.com). The Website depicts the process.

References

[1] United Nations.*The Millennium Development Goals. 2012 Report* . New York: United Nations, 2012.

[2] UNAIDS *UNAIDS Report on the Global AIDS Epidemic 2012. Global report.* Geneva: UNAIDS, 2012. http://www.unaids.org/en/resources/campaigns/20121120_globalreport2012/globalreport/

[3] International HIV / AIDS Alliance. (2010). *The cost efficiency of HIV prevention for vulnerable and most-at-risk populations and the reality of funding. What's Preventing Prevention Campaign Briefing 2.* Hove: International HIV / AIDS Alliance.

[4] UAIDS. *World AIDS day report 2011. How to get to zero: Faster, smarter, Better.* Geneva: UNAIDS, 2011.

[5] Udeh, B.; Udeh, C. & Graves, N., Perinatal HIV transmission and the cost-effectiveness of screening at 14 weeks gestation, at the onset of labour and the rapid testing of infants, *BMC Infectious Diseases* 2008, 8:174 doi:10.1186/1471-2334-8-174. http://www.biomedcentral.com/1471-2334/8/174

[6] Roux, P.; Henley, L.; Cotton, M.; Eley, B., Burden and cost of inpatient care for HIV-positive paediatric patients--status in the Cape Town metropole during the second week of March 1999. Paediatric HIV Census Group. *South African Medical Journal*, 2000, 90(10):1008-1011

[7] Kumar, R.; Mehendale, S.M.; Panda, S.; Venkatesh, S.; Lakshmi, PVM.; Kaur, M.; Prinja, S.; Singh, T.; Virdi, N.K.; Bahuguna, P.; Sharma, A.K.; Singh, S.; Godbole, S.V.; Risbud, A.; Manna, B.; Thirumugal, V.; Roy, T.; Sogarwal, R.; Pawar, N.D. Impact of targeted interventions on heterosexual transmission of HIV in India, *BMC Public Health*, 2011, 11:549, http://www.biomedcentral.com/1471-2458/11/549

[8] Marseille, E.; Kahn, J.G.; Billinghurst, K.; Saba, J. Cost-effectiveness of the female condom in preventing HIV and STDs in commercial sex workers in rural South Africa. *Social Sciences & Medicine*. 2001;52(1):135-148.

[9] Sweat, M; Gregorich, S; Sangiwa, G.; Furlonge, C.; Balmer, D.; Kamenga, C.; Grinstead, O. & Coates, T., Cost-effectiveness of voluntary HIV-1 counselling and

testing in reducing sexual transmission of HIV-1 in Kenya and Tanzania. *Lancet*. 2000 Jul;356 (9224):113-121

[10] Durevall, D, & Lindskog, A. *HIV/AIDS, Adult Mortality and Fertility: Evidence from Malawi. Working Papers in Ecomics, No 284*. Göteborg: School of Business, Economics and Law, University ofGothenburg, 2007.

[11] Rombo, D. *Marital risk factors and HIV infection among women: A comparison between Ghana and Kenya,* Ph.D. Dissertation. Minneapolis: University of Minnesota, 2009.

[12] Öjteg, K. *Socio-Economic determinants of HIV in Zambia. A district level Analysis*. Ph.Dissertation. Lund: Department of Economics at the University of Lund, 2009.

[13] Magadi, M., & Desta , M. A multilevel analysis of the determinants and cross-national variations of HIV seropositivity in sub-Saharan Africa: Evidence from the DHS. *Health & Place*, 2011; 17 (5): 1067–83.

[14] Magadi, M. Understanding the gender disparity in HIV infection across countries in sub-Saharan Africa: evidence from the Demographic and Health Surveys. *Sociology of Health & Illness,* 2011; 33 (4): 522-39.

[15] Asiedu, C., Asiedu, E., & Owusu, F. The Socio-Economic Determinants of HIV/AIDS Infection Rates in Lesotho, Malawi, Swaziland and Zimbabwe. *Development Policy Review ,* 2012; 30 (3): 305-326 .

[16] Green, E. *Rethinking AIDS Prevention. Learning from Successes in Developing Countries* . Westport, CT: Praeger publishers, 2004

[17] International HIV / AIDS Alliance. *The cost efficiency of HIV prevention for vulnerable and most-at-risk populations and the reality of funding. What's Preventing Prevention Campaign Briefing 2*. Hove: International HIV / AIDS Alliance, 2010.

[18] Population Reference Bureau, *2012 World Population Data sheet*, Washington DC, PRB, 2012.

[19] United Nations Development Programme/UNDP. Sustainability and Equity: A better future for all, Human Development Report 2011. New York, UNDP, 2011.

[20] Malawi, HIV national Commission. National 2012 Global AIDS Response Progress

Report: Malawi Country Report. Malawi, HIV National Commission, 2012, 67 p.

[21] Malawi National Statistical Office (NSO); ORC Macro. *Malawi Demographic and Health Survey 2004.* Calverton, Maryland: NSO and ORC Macro, 2005.

[22] Malawi National Statistical Office (NSO); ICF Macro. *Malawi Demographic and Health Survey 2010.* Zomba, Malawi, and Calverton, Maryland, USA: NSO and ICF Macro, 2011.

[23] Kass, G. V. An Exploratory Technique for Investigating Large Quantitaties of Categorical Data. *Applied Statistics,* 1980; 29 (2): 119-127.

[24] Adair, T. *HIV Status and Age at First Marriage among Women in Cameroon.* DHS Working Paper No 33, Calverton, Maryland, USA: Macro International Inc, 2007.

[25] Cleland, J., & Van Ginneken, J. Maternal education and child survival in developing countries: The search for pathways of influence. *Social Sciences and Medicine* , 1988 ; 27(12):1357-1368.

[26] Barrett, H., & Brown, A. Health, hygiene, and maternal education: Evidence from Gambia. *Social Sciences and Medicine*, 1996; 43(11): 1579-1590.

[27] Ainsworth, M. & Semali, I. Who is most likely to die of AIDS? Socioeconomic correlates of adult deaths in Kagera Region, Tanzania. In: *Confronting AIDS: evidence for developing world*. Edited by Ainsworth M, Fransen L, Over M. Washington, D.C.: The World Bank; 1998: 95-110.

[28] Gregson S, Garnett GP, Nyamukapa CA, Hallett TB, Lewis JJ, Mason PR, et al. HIV decline associated with behaviour change in eastern Zimbabwe. *Science,* 2006; 311:664–666.

[29] Hargreaves JR, Glynn JR. Educational attainment and HIV-1 infection in developing countries: a systematic review. *Tropical Medical and International Health*, 2002; 7:489–498.

[30] Gillespie S., Kadiyala, S. & Greener. Is poverty or wealth driving HIV transmission? *AIDS*, 2007; 21 (suppl 7):S5–S16.

[31] Rodrigo, C., & Rajapakse, S. (2010). HIV, poverty and women. *International Health*, 2010, 2(1): 9-16.

[32] Takyi, B., & Broughton, C. Marital Stability in Sub-Saharan Africa: Do Women's Autonomy and Socioeconomic Situation Matter? *Journal of Family and Economic Issues* , 2006; 27 (1): 113-132.

[33] Reniers, G. Divorce and remarriage in rural Malawi. *Demographic Research*, 2003; Special Collection 1, article 6: 175-206. www.demographic-research.org/special/1/6/S1-6.pdf.

[34] Reniers, G. Marital strategies for regulating exposure to HIV. *Demography*, 2008; 45 (2): 417-438.

[35] Youngkong, S.; Baltussen, R.; Tantivess, S.; Koolman, X.; & Teerawattananon, Y. Criteria for priority setting of HIV/AIDS interventions in Thailand: a discrete choice experiment. *BMC Health Services Research* 2010, 10:197. http://www.biomedcentral.com/1472-6963/10/197 retrieved on March 1, 2013

[36] Bendavid, E; Brandeau, M. L.; Wood, R.; Owens, D. K. Comparative Effectiveness of HIV Testing and Treatment in Highly Endemic Regions. *Archives of Internal Medicine, 2010;170 (15):1347-1354*.

Table 1 – Description of the sample

Socioeconomic and demographic Characteristics	Weight			Unweight		
	2004	2010	Total	2004	2010	Total
Age						
15-19	11.4	11.5	11.5	12.1	11.6	11.8
20-24	26.3	21.5	22.8	25.7	20.8	22.2
25-29	19.3	22.7	21.7	20.0	21.9	21.4
30-34	15.7	15.2	15.3	14.9	15.7	15.4
35-39	10.6	13.0	12.3	10.6	12.7	12.1
40-44	9.7	8.6	8.9	9.4	9.1	9.2
45-49	7.1	7.5	7.4	7.3	8.3	8.0
Average	29.2	29.6	29.5	29.1	29.8	29.6
Age at first sex						
<15	18.5	19.1	18.9	20.4	19.2	19.6
15-19	70.1	68.5	68.9	68.4	68.4	68.4
20-24	10.0	11.2	10.8	9.9	11.1	10.7
25&+	1.4	1.3	1.3	1.3	1.3	1.3
Average	16.6	16.6	16.6	16.5	16.6	16.6
Marital status						
Single	6.1	7.5	7.1	6.3	7.6	7.2
In union	81.6	77.4	78.6	80.3	77.1	78.0
Ever married	12.3	15.1	14.3	13.4	15.4	14.8
Number of ever born children						
0	9.7	9.9	9.9	10.3	9.6	9.8
1&+	90.3	90.1	90.1	89.7	90.4	90.2
Age at first birth						
Never give birth	11.0	10.3	10.5	10.2	10.6	10.5
< 20 years old	65.0	64.8	64.9	64.6	64.2	64.3
20 & +	24.0	24.9	24.6	25.2	25.2	25.2
Ever had premarital child						
No	87.2	88.7	88.3	87.9	88.4	88.2
Yes	12.8	11.3	11.7	12.1	11.6	11.8
Relationship to the head of household						
Head of household	16.9	19.4	18.7	17.9	19.0	18.6
Spouse	68.1	62.6	64.1	67.0	62.4	63.7
Daughter & Grand daughter	10.0	11.0	10.7	10.2	11.7	11.3
Others	5.1	7.1	6.5	5.0	6.9	6.3
Province of residence						
Northern	13.7	11.1	11.8	14.3	17.5	16.6
Central	38.0	42.2	41.0	33.8	34.1	34.0
Southern	48.3	46.7	47.1	51.9	48.4	49.4
Place of residence						
Urban	14.4	19.2	17.8	12.2	13.1	12.8
Rural	85.6	80.8	82.2	87.8	86.9	87.2
Religion						
Catholic	23.0	21.2	21.7	21.6	20.6	20.9
Protestant	25.9	24.3	24.7	24.9	25.2	25.1
Other Christians	38.6	39.7	39.4	37.1	42.3	40.8
Muslim	11.6	13.5	12.9	15.5	10.9	12.2
Others	0.9	1.3	1.2	0.9	1.0	1.0
Education						
None	25.7	17.5	19.8	25.4	16.6	19.1
Primary	61.4	63.8	63.1	61.8	66.4	65.1
Secondary & +	12.8	18.7	17.1	12.8	17.0	15.8
Household wealth Index						
Poorest	17.1	17.6	17.5	17.6	19.0	18.6
Poorer	21.2	20.1	20.4	20.8	20.6	20.7
Middle	21.9	19.7	20.3	22.9	20.9	21.4
Richer	22.4	19.3	20.2	22.3	20.7	21.2
Richest	17.4	23.3	21.6	16.4	18.8	18.1
Total	2605	6395	9000	2605	6395	9000

Table 2 – Factors associated with HIV prevalence: Descriptive analyses

Socioeconomic and demographic Characteristics	HIV+	N	Chi-Square	P-value
Age				
15-19	6.1	1,060		
20-24	8.9	1,996		
25-29	14.1	1,922		
30-34	20.0	1,389	205.10	0.000
35-39	22.1	1,091		
40-44	19.7	824		
45-49	15.3	718		
Age at first sex				
<15	18.1	1,762		
15-19	13.8	6,157	21.69	0.000
20-24	12.7	967		
25&+	14.9	114		
Marital status				
Single	9.0	649		
In union	11.9	7,021	331.20	0.000
Ever married	31.3	1,330		
Number of ever born children				
0	10.4	886	12.91	0.000
1&+	14.9	8,114		
Age at first birth				
Never give birth	11.5	946		
< 20 years old	14.8	5,837	7.080	0.029
20 & +	14.9	2,217		
Ever experience premarital childbearing				
No	13.7	7,941	29.54	0.000
Yes	20.2	1,059		
Relationship to the head of household				
Head of household	25.3	1,678		
Spouse	11.7	5,737	197.80	0.000
Daughter & Grand daughter	11.1	1,014		
Others	16.4	571		
Region of residence				
Northern	10.0	1,494		
Central	9.5	3,062	184.90	0.000
Southern	20.0	4,444		
Place of residence				
Urban	24.7	1,156	157.00	0.000
Rural	12.3	7,844		
Religion				
Catholic	13.1	1,879		
Protestant	15.1	2,258		
Other Christians	14.3	3,674	7.84	0.090
Muslim	16.4	1,100		
Others	10.6	89		
Education				
None	14.3	1,722		
Primary	13.4	5,855	27.92	0.000
Secondary & +	18.8	1,423		
Household wealth Index				
Poorest	10.3	1,673		
Poorer	10.5	1,862		
Middle	12.4	1,930	148.60	0.000
Richer	15.7	1,904		
Richest	22.5	1,631		
Year of survey				
2004	14.4	2,605	0.01	0.935
2007	14.5	6,395		
Total	14.5	9,000		

Table 3 - Model Summary

Model components	Model specification	Results
Growing Method:	CHAID	-
Dependent variable	HIV Marital status	HIV+=14.7%
Independent Variables	Age, Age at first sex, Marital status, Age at first union, Ever had a child, Experience premarital childbearing, age at first birth, Region of residence, Place of residence, Education, Wealth Index, Religion, Relationship to the head of household.	**Marital status**, Region of residence, Age, Place of residence, Wealth Index, Age at first sex, Ever had a child
Maximum Tree Depth	3	3
Minimum Cases in Parent Node	100	100
Minimum Cases in Child Node	50	50
Number of Nodes	-	27
Number of Terminal Nodes	-	16

Table 4 – Prevalence of HIV by groups

Node	Group description	Population %	Population N	HIV Prevalence
	Group 1			
22	Formerly in union-richest-had first sex from 25 years old	0.7	62	72.7
19	Formerly in union-richer or middle households- Southern region	2.7	242	45.7
21	Formerly in union-richest-had first sex between 15 and 24 years old	2.3	206	45.3
Total	*Group 1*	*5.7*	*510*	*54.6*
	Group 2			
7	Formerly in union-poorer households	2.8	251	27.5
25	Never married, living in urban area-Southern or Northern region	1.6	144	23.5
12	In union – living in Southern region – age 30-44	13.2	1,185	22.9
20	Formerly in union-richer or middle households-Central or Northern region	1.6	144	21.9
17	Formerly in union-poorest households-Southern region	2.3	206	20.9
Total	*Group 2*	*21.5*	*1,930*	*23.3*
	Group 3			
16	In union – living in Central or Northern region – urban area	6.1	549	18.0
14	In union – living in Southern region – age 25-29	9.0	810	18.0
23	Never married, living in rural area and ever gave birth	1.4	126	11.4
13	In union – living in Southern region – age 15-24/ 45-49	14.3	1,286	11.2
18	Formerly in union-poorest households-Central or Northern region	2.1	189	10.3
Total	*Group 3*	*32.9*	2,960	*13.8*
	Group 4			
15	In union, living in Northern or Central province – rural areas	35.9	3,231	6.1
26	Never married, living in urban area in Central region	0.9	81	2.7
24	Never married, living in rural area and never gave birth	3.2	288	2.3
Total	*Group 4*	*40.0*	3,600	*3.7*
	Total (Overall)	100	9,000	14.7

Figure 1 - HIV prevalence in Malawi: tree diagram

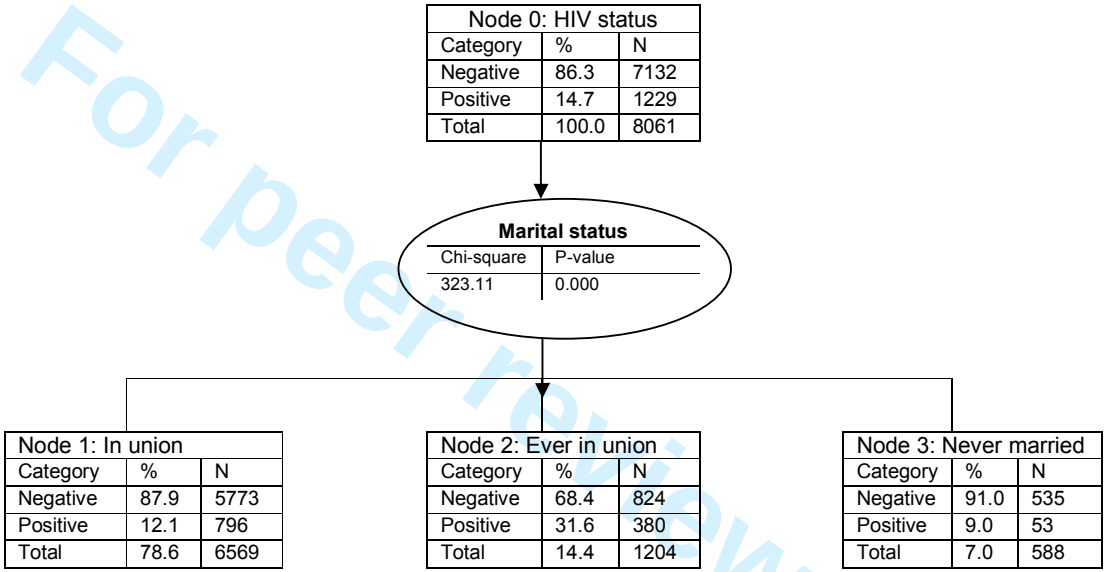


Figure 1a- HIV prevalence in Malawi: tree diagram for women in union

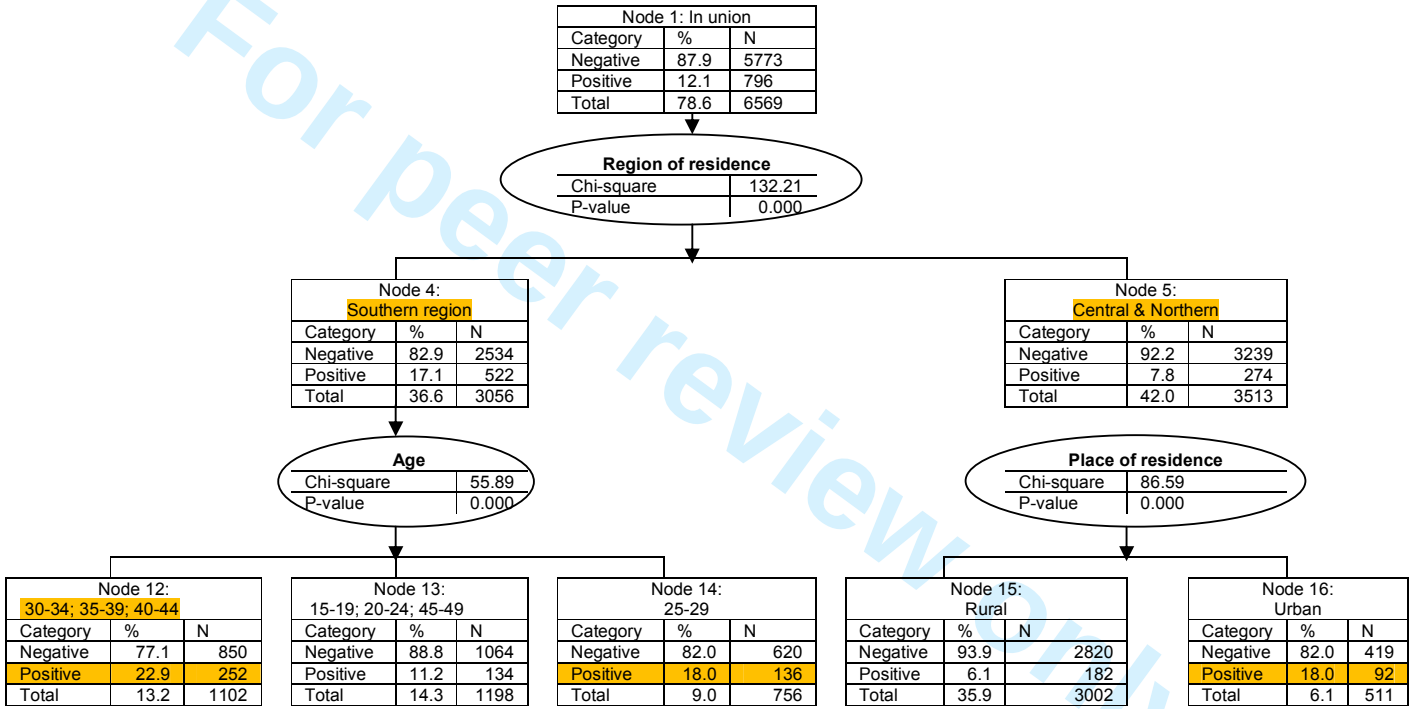


Figure 1b - HIV prevalence in Malawi: tree diagram for women in union disruption

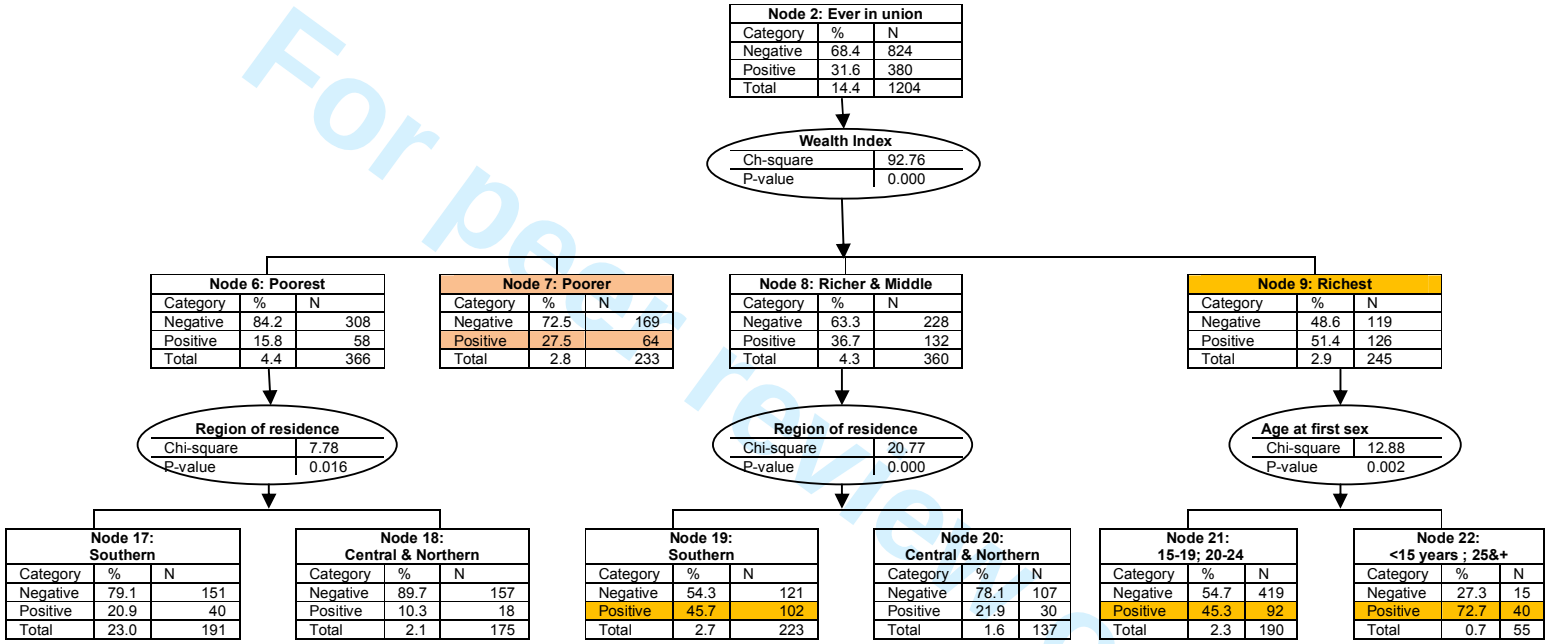
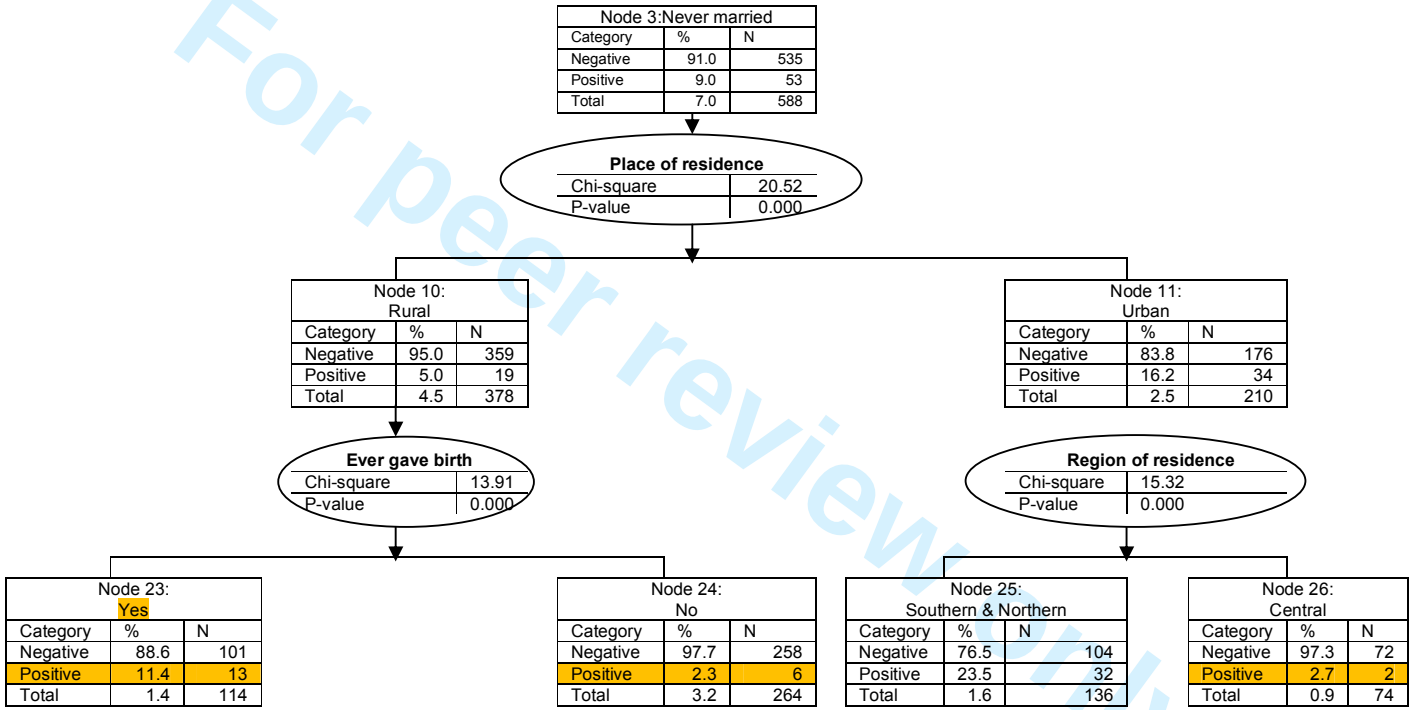


Figure 1c - HIV prevalence in Malawi: tree diagram for never married women



Article summary

1. Article focus

- Targeted interventions and evidence based prevention programmes have been advocated as cost-effective strategy to combat HIV/AIDS.
- Whereas in countries with concentrated AIDS epidemics (Latina America, East Asia and Eastern Europe), the most-at-risk populations including commercial sex workers (CSWs), long distance truck drivers, men who have sex with men, and unmarried youth account for a large proportion of new infections, in countries with high prevalence, they account only for a smaller share of new infections.
- Who are the most-at-risk populations regarding HIV prevalence in Malawi? With HIV prevalence of about 13 percent among women of reproductive age, HIV/AIDS constitutes a drain on the labor force and government expenditures in Malawi.

2. Key messages

- We use data from the Malawi 2004 and 2010 Demographic and Health Surveys to profile HIV most-at-risk groups of women in Malawi where about 14 percent of women are HIV positive.
- Our findings revealed that richest and formerly in union women are the most-at-risk population.
- We suggested targeted interventions considering the groups HIV prevalence and size. With 45 percent of the study population, among whom the HIV prevalence is estimated at 17 percent on average, couples (males and women in union) living in the Southern region and those living in the urban areas of the Central and the Northern should be the first targets of HIV interventions in Malawi.

Strengths and limitations of this study

- Our study seems the first in Malawi that has attempted to profile HIV most-at-risk groups of women in Malawi. The most-at-risk population refers to a combination of several factors because socioeconomics factors associated with HIV are not mutually exclusive.
- The major strength is the use of the Chi-square Automatic Interaction Detector (CHAID) to identify HIV predictors and the most at risk groups of women for intervention. CHAID uses regression and classification algorithms and offers a non-algebraic method for partitioning data that lends itself to graphical displays.
- The major limitation of this study is the cross- sectional nature of the data, which does not permit one to draw causal association between HIV status and the associated factors. For instance, whether HIV infection has occurred before, during or after the union.

STROBE 2007 (v4) Statement—Checklist of items that should be included in reports of *cross-sectional studies*

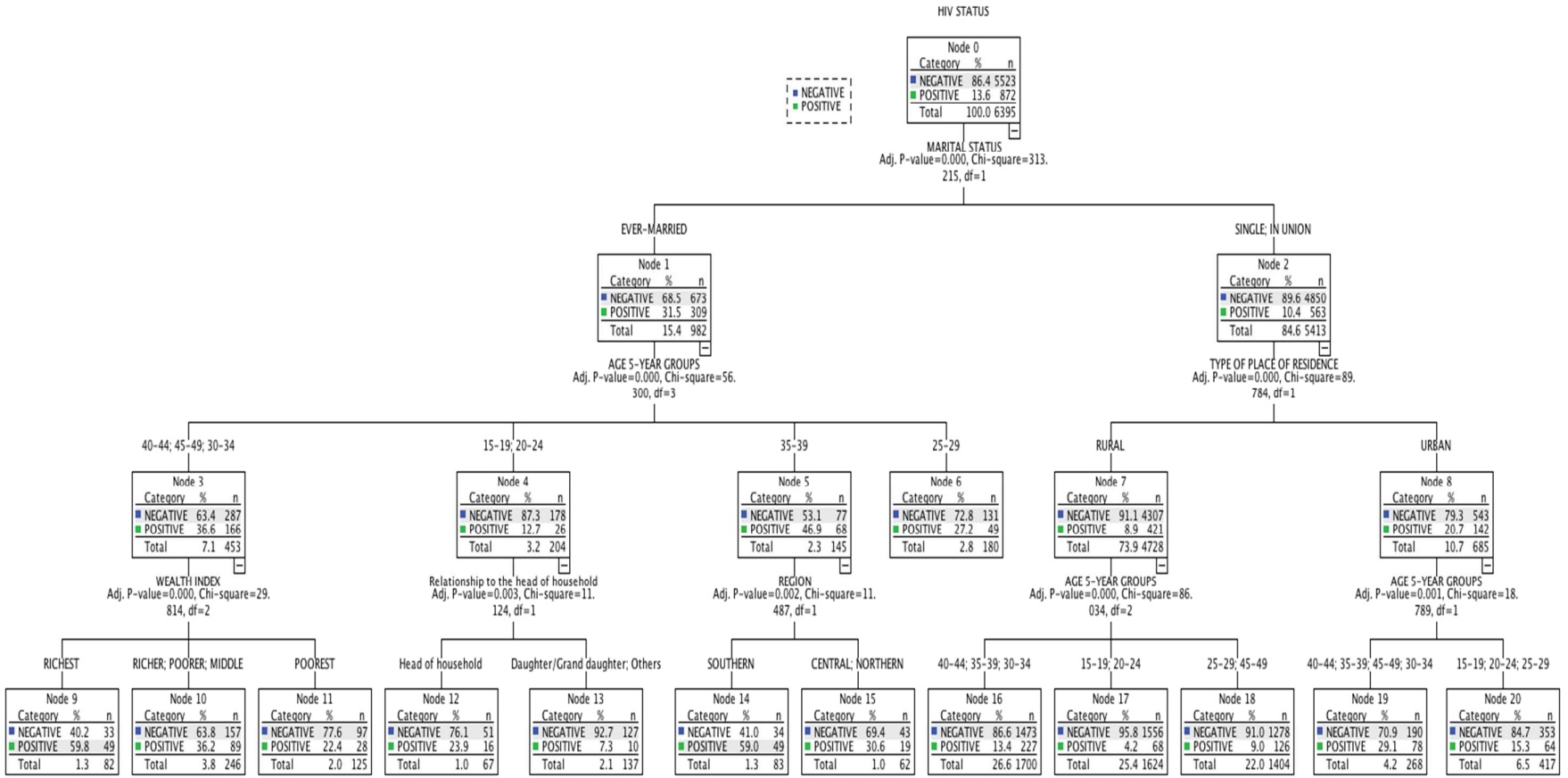
Section/Topic	Item #	Recommendation	Reported on page #
Title and abstract	1	(*a*) Indicate the study's design with a commonly used term in the title or the abstract	1
		(*b*) Provide in the abstract an informative and balanced summary of what was done and what was found	2
Introduction			
Background/rationale	2	Explain the scientific background and rationale for the investigation being reported	4
Objectives	3	State specific objectives, including any prespecified hypotheses	4
Methods			
Study design	4	Present key elements of study design early in the paper	5
Setting	5	Describe the setting, locations, and relevant dates, including periods of recruitment, exposure, follow-up, and data collection	5
Participants	6	(*a*) Give the eligibility criteria, and the sources and methods of selection of participants	5
Variables	7	Clearly define all outcomes, exposures, predictors, potential confounders, and effect modifiers. Give diagnostic criteria, if applicable	6
Data sources/ measurement	8*	For each variable of interest, give sources of data and details of methods of assessment (measurement). Describe comparability of assessment methods if there is more than one group	5-6
Bias	9	Describe any efforts to address potential sources of bias	*5-7*
Study size	10	Explain how the study size was arrived at	5
Quantitative variables	11	Explain how quantitative variables were handled in the analyses. If applicable, describe which groupings were chosen and why	6
Statistical methods	12	(*a*) Describe all statistical methods, including those used to control for confounding	6-7
		(*b*) Describe any methods used to examine subgroups and interactions	6-7
		(*c*) Explain how missing data were addressed	6-7
		(*d*) If applicable, describe analytical methods taking account of sampling strategy	6-7
		(*e*) Describe any sensitivity analyses	6-7
Results			

Participants	13*	(a) Report numbers of individuals at each stage of study—eg numbers potentially eligible, examined for eligibility, confirmed eligible, included in the study, completing follow-up, and analysed	5
		(b) Give reasons for non-participation at each stage	5
		(c) Consider use of a flow diagram	5
Descriptive data	14*	(a) Give characteristics of study participants (eg demographic, clinical, social) and information on exposures and potential confounders	7-8
		(b) Indicate number of participants with missing data for each variable of interest	7-8
Outcome data	15*	Report numbers of outcome events or summary measures	7-8
Main results	16	(*a*) Give unadjusted estimates and, if applicable, confounder-adjusted estimates and their precision (eg, 95% confidence interval). Make clear which confounders were adjusted for and why they were included	8-12
		(*b*) Report category boundaries when continuous variables were categorized	8-12
		(*c*) If relevant, consider translating estimates of relative risk into absolute risk for a meaningful time period	8-12
Other analyses	17	Report other analyses done—eg analyses of subgroups and interactions, and sensitivity analyses	8-12
Discussion			
Key results	18	Summarise key results with reference to study objectives	12-14
Limitations	19	Discuss limitations of the study, taking into account sources of potential bias or imprecision. Discuss both direction and magnitude of any potential bias	12-14
Interpretation	20	Give a cautious overall interpretation of results considering objectives, limitations, multiplicity of analyses, results from similar studies, and other relevant evidence	12-14
Generalisability	21	Discuss the generalisability (external validity) of the study results	12-14
Other information			
Funding	22	Give the source of funding and the role of the funders for the present study and, if applicable, for the original study on which the present article is based	15

*Give information separately for cases and controls in case-control studies and, if applicable, for exposed and unexposed groups in cohort and cross-sectional studies.

Note: An Explanation and Elaboration article discusses each checklist item and gives methodological background and published examples of transparent reporting. The STROBE checklist is best used in conjunction with this article (freely available on the Web sites of PLoS Medicine at http://www.plosmedicine.org/, Annals of Internal Medicine at http://www.annals.org/, and Epidemiology at http://www.epidem.com/). Information on the STROBE Initiative is available at www.strobe-statement.org.

Figure 1 – Tree diagram of HIV prevalence in Malawi by selected background characteristics



Article summary

1. Article focus

- Targeted interventions and evidence based prevention programmes have been advocated as cost-effective strategy to combat HIV/AIDS.
- Who are the most-at-risk populations regarding HIV prevalence in Malawi? With HIV prevalence of about 14 percent among women of reproductive age, HIV/AIDS constitutes a drain on the labor force and government expenditures in Malawi.

2. Key messages

- We use data from the Malawi 2010 Demographic and Health Surveys to profile HIV most-at-risk groups of women in Malawi where about 14 percent of women are HIV positive.
- Our findings revealed that richest and formerly in union women is the most-at-risk group.
- To achieve zero new infection as part of MDG 6, there is need of more comprehensive policy to combat HIV because of the complexity of HIV socioeconomic profile in Malawi. There are several groups built from several socioeconomic categories depending on individual marital status, wealth index, age, place of residence, and relationship to the head of household.

Strengths and limitations of this study

- From our knowledge this study may be the first in Malawi to attempt to profile HIV most-at-risk groups of women in Malawi. The most-at-risk population refers to a combination of several factors because factors associated with HIV are not mutually exclusive.
- The major strength is the use of the Chi-square Automatic Interaction Detector (CHAID) to identify HIV predictors and the most-at-risk groups among women for intervention. CHAID uses regression and classification algorithms and offers a non-algebraic method for partitioning data that lends itself to graphical displays.
- This study has two major limitations. First, this study used cross-sectional data from the Demographic and Health Surveys, which does not permit one to draw causal association between HIV status and the associated factors. For instance, whether HIV infection has occurred before, during or after the union. Last, CHAID model ignores the hierarchical structure of the Demographic and Health Survey data and need large sample size.

February 22, 2013

Mr. Richard Sands
Managing Editor, BMJ Open

RE: Manuscript ID bmjopen-2012-002459 entitled "Prevalence of HIV among women in Malawi: Identifying the most-at-risk groups for targeted and cost-effective interventions"

Dear Mr Richard Sands,

Please find enclosed our revised manuscript, which addresses the reviewers' concerns and suggestions. What follows is a point-by-point response to the comments provided as part of the review process. Each group of responses has been numbered to correspond with those on the comments. Moreover, in the revised manuscript we have highlighted in red colour the areas that have been modified compared to the original submission.

Reviewer(s)' Comments to Author:

Reviewer: Kandala Ngianga-Bakwin
University of Warwick, Warwick Medical School
No competing interest

This paper examines the relationship between the prevalence of HIV infection and HIV socioeconomic predictors as well as to identify the most at-risk groups of women in Malawi using consecutive cross-sectional surveys (the 2004 and 2010 Malawi Health and Demographic Surveys (MDHS)). Data were from 8,596 men and women aged 15-49 who participated in the MDHS to profile participants with high risk of HIV using Chi-square Automatic Interaction Detector (CHAID).

The CHAID procedure partitions data that lends itself to graphical displays and the method is a sequential fitting algorithm and its statistical tests are sequential with later effects being dependent upon earlier ones and not simultaneous as would be the case in a regression model or analysis of variance where all effects are fit simultaneously.

The authors do a good job profiling at risk groups of HIV infection women in Malawi and examining the prevalence of HIV infection and HIV socioeconomic predictors by applying CHAID. The statistical methods employed to describe the data (CHAID) is appropriate for this kind of data. The context is relevant; the findings are very informative and likely to contribute to policy in Malawi. Identifying at high risk of women with HIV infection reinforces the need for targeted programming based on the specific strengths and challenges of each community, even within an area as small as village. Nevertheless, the authors need to address some disadvantages and limitation of the methodology employed, the (CHAID) more appropriately.

Reply: We would like to thank the reviewer for his comments on our manuscript.

1. The methodology employed (CHAID) has some shortcomings. CHAID doesn't take into account the hierarchical structure of this data and the method (CHAID) is often used as an exploratory technique and needs rather large sample sizes to work effectively, since with small sample sizes the respondent groups can quickly become too small for reliable analysis.

Furthermore, CHAID methods are not able to assess and deal with the auto-correlation (dependence of nearby groups) in the data, non-linear (age) and time varying effects of covariates. Specifically, the study relied on the assumption of independence of the profiled high risk groups but in practice groups close to each other are related. CHAID models cannot resolve this above issue. Could the authors please comment on these issues?

Reply: The reviewer is right, like all statistical methods CHAID has some limitations including ignoring the hierarchical structure of the Demographic and Health Survey data, and need of large sample size. We have recognized these shortcomings in the document. CHAID is not only an exploratory technique though it cannot be use for testing causality because the technique has good prediction capacities. Using CHAID, one can uncover relationships between a dependent variable, HIV status in our case, and a host of predictor variables. Therefore, we used CHAID for identifying homogenous groups of women considering HIV prevalence risk. All our variables are categorical to take into account non-linearity of some variable such as age though CHAID identifies homogenous segments.

2. Another challenge of profiling high-risk groups of HIV prevalence rates using CHAID with the MDHS is that often the method cannot deal with the issue of interaction between variables such as rural/urban, education. Please comment!

Reply: Thanks for the comment. CHAID deals with issue of interaction between variable. The method assesses the category groupings, or interval breaks to pick the most significant combination of variables.

3. I have also some specific comments:

a) Article focus : The second bullet of this section is not a focus of this paper, please delete or amend;

Reply: Thanks for the comment. The bullet has been deleted as suggested.

b) Strengths and limitations of the study: in the second bullet, authors should rather detail how the use CHAID methodology has been able to improve the analysis of these data compared to other methods rather than mentioning what CHAID is able to perform;

Reply: We have mentioned one the advantage of this method. "The model depicted also different interactions between risk factors and profiled HIV risk groups in Malawi using several variables including marital status, wealth index, age, place of residence, and relationship to the head of household".

c) it is also worth mentioning that by not accounting for men risk factors in profiling women, the results might be limited because in the context of HIV in these settings, men have power over women sexuality, which might influence women risk factors.

Reply: Reviewer is right. However our recommendations take into account men in combating HIV (See last paragraph of page 15). "Couples (males and women in union) and never married people age 25-49 (nodes 16 and 18) living in rural

areas should be the first targets using universal HIV testing, “Abstinence”, “Be faithful” and “use condom” campaign. Indeed, this group includes 49 percent of the study population, among whom the HIV prevalence is estimated at 11 percent on average. About 40 percent of women living with HIV in Malawi belong to this category”.

Minor comments

4. Page 3 line 17, the HIV prevalence is said to be 13 %, which is different from line 26 p3, Line 33 page 3, Line 31 p4 and Line 38, p8. Please confirm the correct one.

Reply: The HIV prevalence among women of reproductive age is estimated at 13.6 percent. The number has been consistently reported in the document.

5. P8 Line 21. Since the majority of women (82%) live in rural areas. Can the author comments of the impact of this imbalance in the results and whether the methods used were able to account for it.

Reply: We have worked with weighted file to take into account the sampling design effect including population distribution by place of residence.

6. The notion of nodes is introduce in results section page 9 without explaining it in statistical the methods section. Please explain it for your readers in the statistical the methods section.

Reply: The notion of nodes has been introduced and explained in the statistical methods section as suggested.

7. Line 48 page 12 should read 2004 and 2010. Not 2004 and 2004 DHS. Please also explain clarify how the CHAID in this study offers a useful alternative to traditional regression models.

Reply: Thanks for the comments. The suggestion has been implemented taking into account that analyses are based on the 2010 DHS. We have inserted a paragraph to show advantages of using CHAID in this study.

As a data mining technique with a hierarchical nature, it allows identifying population subgroups that share similar characteristics. It provides a graphical display of the results, which facilitate their interpretation and can assist health decision makers in visualizing how predictors interact to define high-risk subgroups and understand the importance of each variable examined in predicting the modeled response. Furthermore, studies have found the two methodologies to be comparable in predictive performance.

8. P13 Lines 52-57 Family planning was not formerly examined in this study and I think this conclusion is not part of this study. Please revise.

Reply: Yes, the reviewer is right the study does not examine Family Planning. However, the population understudied includes women of reproductive health. Integrating the two services (HIV and FP) could be cost-effective.

Reviewer: Name: Kavita Singh

Position: PhD Scholar and Research Associate

Institution: All India Institute of Medical Sciences, New Delhi

Country: India

I have no competing / conflict of interest with the research study / manuscript I reviewed.

9. Research question needs to be clearly specified, whether the authors are interested in estimating prevalence of HIV in Malawi women or just aims to identify the predictors of HIV prevalence in Malawi women across different socioeconomic groups. Accordingly, the manuscript title can be clearly formed.

Reply: We have clarified the research question and objectives.

10. Study design: Pooling data from two cross-sectional survey conducted in 2004 and 2010, seems not to be methodologically sound from purist perspective. I think, Outcomes to two samples can be compared but can it be summed and then used for estimation of prevalence and its association with risk factors can be analyzed is a point of concern from my viewpoint.

Reply: Current analyses exclude the 2004 DHS data.

11. Total number of participants used from different surveys is not clear and again different total sample size has been written invariable through out the manuscript.

Reply: Number of participants has been checked, corrected and reported consistently.

12. Methods section, needs to described clearly.

Reply: We have added paragraphs and sentences to improve the method section.

13. Abstract section has some major limitations, which are compiled in my comments pasted below.

Reply: The abstract section has been improved to take into account the comments.

14. Manuscript can be improved further with a couple of thoughtful iterations.

Reply: We thank the reviewer for the suggestion.

15. Background section is not appropriately references at many places. Complete web URL's for UN MDG and UNAIDS reports should be provided.

Reply: References have been completed

16. STROBE checklist included at the end of the manuscript, though is indicated and marked to have covered all aspects, however I don't think the checklist is rigorously followed. For example a flow chart of the number of eligible participants included is marked as YES referring to Page 5 under Participants section, 28 of the manuscript, however no such flow diagram has actually been included. Authors may wish to review the checklist again and improve on the quality of reporting the study findings further.

Reply: Thank you for the comment. We have revised the checklist.

17. I am sorry to mention this, however interpretation of results and result tables are not clear and very ambiguous at many places. Needs to be closely reviewed and presented in concrete fashion.

Reply: Thanks for the comment.

18. A statement on appropriate approval from research ethics board / DHS dataset authorised personnel, for conducting secondary data analysis has not been included in the manuscript.

Reply: We have included the statement acknowledging special authorization from the DHS research ethics board for conducting secondary analysis of HIV datasets.

19. I think, the authors have identified an important area to be addressed in Malawi community and that is of identifying the most-at risk groups for HIV infection. There are few major observations, which I have made after reading through this manuscript many times. I have summarized below my comments and also pasted some of these on the manuscript which is enclosed along with these comments for author's reference:

Reply: We thank the reviewer for his kind comments. Comments have been addressed and suggestions have been implemented to make the document more insightful and informative in the Malawi's context for public health policy making.

20. Manuscript title doesn't reflect or is consistent with the actual study objectives, because the main objective of this study is not estimate the prevalence of HIV among women in Malawi but rather just to identify the women groups who are high risk of HIV and accordingly to inform the cost-effective interventions. Authors may consider to concise the manuscript title and make it more suggestive of the actual planned study.

Reply: The title has been changed to "Identifying HIV most-at-risk groups in Malawi for targeted interventions"

21. There are a number of data discrepancies observed at number of places, it might be because the authors did not discuss clearly column heading or the way they have based their calculations, but still there are discrepancies. For example, look into the attached manuscript, where the discrepancies in total number of participants included in the study have been flagged.

Reply: Column headings have been clearly labeled and numbers have been checked.

22. Page 2: Abstract; conclusion – line21; primary and secondary outcome measure can be clarified; line 37, needs to be re-worded to make it clear and consistent with actual findings.

Reply: We have clarified "HIV status (positive or negative)" is the outcome. Furthermore we have implemented the suggestion to clarify the sentence.

> *"Women in union and their partners as well as never married people age 25-49 (nodes 16 and 18) living in rural areas should be the first prevention targets. This population accounts for 49 percent of the sample among and 40 percent of women living with HIV in Malawi"*

23. Page 4; Line 21 – current costs of HIV treatment is presented as (US $ 4,707 over lifetime), is not clear for which setting, age group and time period. Also, whether it's the out of pocket expenditure or cost borne by government? I think, clarifying on this aspect will help readers better understand the context of cost of HIV treatment and its implications. Appropriate reference can be cited for this.

Reply: We have provided additional: "*Based on the global costs of HIV treatment in 2010 estimated between US$22-24 billion annually by 2015 and individual cost of US $4,707 over lifetime in order to reach global targets [i,ii], targeted interventions and evidence based prevention programmes have been advocated as cost-effective strategy to combat HIV/AIDS*".

24. Page 4; Line 21 – examples of few cost-effective interventions/strategies should be included to provide a comprehensive detail in the context.

Reply: We have cited some targeted cost-effective interventions research.

25. Page 5; line 34: Data sources: The approach of combining two cross sectional surveys conducted at two different time points (2004 and 2010) with a gap of 6 years, doesn't make the study methodologically sound. Authors can provide justification and specific reasoning behind combining cross-sectional data sets from two different period for estimating prevalence of HIV in the community, which can be flawed by many factors, such as HIV being a chronic type of condition, there might be higher chances of calculating the same person twice in prevalence estimation, when you combine two dataset for prevalence calculation. Additionally, there might be lots of demographic and socio-economic changes that can occur in the society over 6 years period and I would be concerned in the way then, how productive and reliable this exercise may result in by combining the cross-sectional survey datasets from 2004 and 2010. Just a suggestion, authors may consider restricting the statistical analysis to the participants surveyed in 2010, which has pretty good sample size of ~8174 participants. I don't think adding meager 2000 additional participants from 2004 survey would add any incremental benefits to the statistical analysis part. Interpretation of these numbers also seems to be faulty. I have added specific comments to the tables in enclosed manuscript.

Reply: We thank reviewer for the comment and suggestion. Current analyses exclude the 2004 DHS data.

26. Page 14; line 27-31 is repetition of same paragraph on the above page 13; line 51-56

Reply: Thanks for the comment. We have implemented the comment.

27. Page 8; line 7-13; Invariably in the results section, the average figure reported for many variables like average age, or percent of women never married doesn't clearly quote the figures represented in the Table 1. Also, it's unclear whether the authors have used weighted or un-weighted average.

Reply: We have interpreted weighted figures. We have updated the result section to take into account the comment.

28. Page 19; Table 2: At the bottom of the table year of surveys mentioned are 2004 and 2007. I found most part of statistical analysis and results section not very clear. Authors may be advised to make be concrete, consistent and reader friendly.

Reply: Thank for the comments. Variable year of surveys has been removed from the analysis and the table has been updated accordingly.

29. I would recommend, if authors can provide methodological sound justification of combining two survey datasets conduced at a gap of 6 years for estimating the prevalence and then calculating the socio-economic predictors, would be much appreciated. Also, all tables needs to be organized, clearly labeled and cross-verified for the total numbers.

Reply: Thank for the comments. Current analyses exclude the 2004 DHS data. All tables have been clearly labelled and numbers have been cross-verified as recommended.

We would like to thank the reviewers for thoughtful comments and suggestions. We truly appreciate your interest in our work. We believe that as a result of the review process our paper has greatly improved and hope that it is now acceptable for publication in BMJ Open.

Yours sincerely,

Jacques B.O. Emina, PhD

[i] UNAIDS *UNAIDS Report on the Global AIDS Epidemic 2012. Global report.* Geneva: UNAIDS, 2012. http://www.unaids.org/en/resources/campaigns/20121120_globalreport2012/globalreport/

[ii] International HIV / AIDS Alliance. (2010). *The cost efficiency of HIV prevention for vulnerable and most-at-risk populations and the reality of funding. What's Preventing Prevention Campaign Briefing 2.* Hove: International HIV / AIDS Alliance.

Identifying HIV most-at-risk groups in Malawi for targeted interventions

Journal:	*BMJ Open*
Manuscript ID:	bmjopen-2012-002459.R2
Article Type:	Research
Date Submitted by the Author:	08-Apr-2013
Complete List of Authors:	Emina, Jacques; University of Kinshasa, ; CEPS/ INSTEAD, Madise, Nyovani; University of Southampton, Faculty of Social and Human Sciences Kuepie, Mathias; CEPS/INSTEAD, AFRILUX Zulu, Eliya; African Institute for Development Policy, Yazoume , Ye ; ICF International,
Primary Subject Heading:	HIV/AIDS
Secondary Subject Heading:	Epidemiology, Health policy, Sociology, Public health, Infectious diseases
Keywords:	HIV & AIDS < INFECTIOUS DISEASES, Epidemiology < INFECTIOUS DISEASES, Public health < INFECTIOUS DISEASES

Identifying HIV most-at-risk groups in Malawi for targeted interventions

Jacques B. O. Emina [1,2], Nyovani Madise [3], Mathias Kuepie [2], Eliya M. Zulu [4], Yazoume Ye [5]

1 Department of Population and Development Studies, University of Kinshasa, P.O.BOX 176 Kinshasa XI, Democratic Republic of Congo

2 CEPS/INSTEAD 3, avenue de la Fonte, L-4364 Esch-sur-Alzette Luxembourg

3 Faculty of Social and Human Sciences, Room 4103 Building 58, University of Southampton, Southampton, SO17 1BJ.

4 African Institute for Development Policy (AFIDEP), P. O. Box 14688-00800, Westlands, Nairobi, Kenya.

5 ICF International 11785 Beltsville Drive, Suite 300 Calverton, MD 20705, USA.

Correspondence to: Jacques B.O. Emina; jacques.emina@gmail.com

Abstract

Objectives: To identify HIV socioeconomic predictors as well as identify the most-at-risk groups of women in Malawi.

Design: cross-sectional survey

Setting: Malawi

Participants: The study used a sample of 6,395 women age 15-49 years from the 2010 Malawi Health and Demographic Surveys (MDHS).

Interventions: N/A

Primary and secondary outcome measures: Individual HIV status: positive or not.

Results: Findings from Pearson Chi-square and Chi-square Automatic Interaction Detector (CHAID) analyses revealed that marital status is the most significant predictor of HIV. Women who are no longer in union and living households in the highest wealth quintiles constitute the most-at-risk group; whereas the less-at-risk group includes young women (15-24) never married or in union and living in rural area.

Conclusion: In the light of these findings, this study recommends: (1) design and implementation of targeted interventions taking into account HIV prevalence and the demographic size of different groups at risk groups. Preventive interventions should prioritize couples and never married people age 25-49 living in rural areas because this group accounts for 49 percent of the study population and 40 percent of women living with HIV in Malawi; (2) With reference to treatment and care, higher priority must be given to promoting HIV test, monitoring and evaluation of equity in access to treatment among women in union disruption and never married or women in union age 30-49 living in urban areas; (3) Community health workers, households based campaign, reproductive health services and reproductive health courses at school could be used as canon to achieve universal prevention strategy, testing, counseling and treatment.

Keywords: HIV/AIDS, Malawi, CHAID methods, decision analysis, most-at-risk groups, targeted interventions.

Article summary

1. Article focus

- Targeted interventions and evidence based prevention programmes have been advocated as cost-effective strategy to combat HIV/AIDS.
- Who are the most-at-risk populations regarding HIV prevalence in Malawi? With HIV prevalence of about 14 percent among women of reproductive age, HIV/AIDS constitutes a drain on the labor force and government expenditures in Malawi.

2. Key messages

- We use data from the Malawi 2010 Demographic and Health Surveys to profile HIV most-at-risk groups of women in Malawi where about 14 percent of women are HIV positive.
- Our findings revealed that richest and formerly in union women is the most-at-risk group.
- To achieve zero new infection as part of MDG 6, there is need of more comprehensive policy to combat HIV because of the complexity of HIV socioeconomic profile in Malawi. There are several groups built from several socioeconomic categories depending on individual marital status, wealth index, age, place of residence, and relationship to the head of household.

Strengths and limitations of this study

- From our knowledge this study may be the first in Malawi to attempt to profile HIV most-at-risk groups of women in Malawi. The most-at-risk population refers to a combination of several factors because factors associated with HIV are not mutually exclusive.
- The major strength is the use of the Chi-square Automatic Interaction Detector (CHAID) to identify HIV predictors and the most-at-risk groups among women for intervention. CHAID uses regression and classification algorithms and offers a non-algebraic method for partitioning data that lends itself to graphical displays.
- This study has two major limitations. First, this study used cross-sectional data from the Demographic and Health Surveys, which does not permit one to draw causal association between HIV status and the associated factors. For instance, whether HIV infection has occurred before, during or after the union. Last, CHAID model ignores the hierarchical structure of the Demographic and Health Survey data and need large sample size.

Introduction

In 2000, the United Nations' Millennium summit identified the reduction of HIV prevalence as one of the eight fundamental goals for furthering human development. Though global HIV/ AIDS incidence is declining, HIV/AIDS has remained the

leading cause of death in women of reproductive age in low-and middle-income countries, particularly in sub-Sahara Africa (SSA) [1]. The gap between the current state of HIV/AIDS and the UNAIDS goals of three zero (zero new HIV infections, zero discrimination, and zero AIDS related deaths) remains important. With barely two years remaining to the end-date of the Millennium Development Goals (MDG) target, HIV/AIDS remains a long-term global challenge [1].

Given the high cost of HIV/AIDs treatment estimated in 2010 to be globally between US$22 and US$24 billion annually by 2015 and individual cost of US $4,707 over lifetime to reach global targets [2,3], targeted interventions and evidence based prevention programmes have been advocated as cost-effective strategy to combat HIV/AIDS. Such a strategy reduces levels of vulnerability and risk as well as allowing HIV interventions to optimize coverage, reducing costs and lowering the number of new infections [4]. In the United State Virgin Islands, the recommended strategy of universal screening by 14 weeks gestation and screening the infant after birth has a cost savings of $1,122,787 and health benefits of 310 life year gained [5]. A prevention of Mother-to-Child Transmission intervention in Capet town, South Africa, revealed that a program at a scale sufficient to prevent 37 percent of pediatric HIV infections would cost about US$0.34 per person in South Africa and would be affordable to the health care system [6].

In Indian high HIV prevalence southern states, targeted interventions result in significant decline in HIV prevalence among female commercial sex workers (CSWs) and young pregnant women [7]. Evaluation of the cost-effectiveness of the female condom (FC) in preventing HIV infection and other sexually transmitted Diseases (STDs) among CSWs and their clients in the Mpumulanga Province of South Africa, showed that a well-designed FC program oriented to CSWs and other women with casual partners is likely to be highly cost-effective and can save public sector health payer US $12,090 in averted HIV/AIDS treatment costs in rural South Africa [8]

Likewise, analysis of targeting Voluntary HIV Counseling and Testing in Kenya and in Tanzania showed that increasing the proportion of couples to 70 percent reduces the cost per disability-adjusted life year (DALY) saved to $10.71 in Kenya and $13.39 in Tanzania, and that targeting a population with HIV-1 prevalence of 45 percent decreased the cost per DALY saved to $8.36 in Kenya and $11.74 in Tanzania [9].

However despite growing literature in health and social sciences on factors associated with HIV/AIDS during the last three decades, less is known about the most-at-risk populations regarding HIV prevalence [10, 11, 12, 13,14, 15]. Indeed, whereas in countries with concentrated HIV/AIDS epidemics (Latina America, East Asia and Eastern Europe), the most-at-risk populations including CSWs, long distance truck drivers, men who have sex with men, and unmarried youth [16,11,17] account for a large proportion of new infections, in countries with high prevalence, they account only for a smaller share of new infections [17].

Against this background, this study aims to identify HIV socioeconomic predictors as well as identify the most-at-risk groups among women in Malawi. With HIV prevalence of about 13.6 percent among women of reproductive age [18], HIV/AIDS constitutes a drain on the labor force and government expenditures in Malawi.

Data and Methods

Study setting

The Republic of Malawi is a landlocked country in southeast Africa. Malawi is over 118,000 km^2 with an estimated population of about 16 millions [18]. Its capital is Lilongwe, which is also Malawi's largest city; the second largest is Blantyre and the third is Mzuzu.

Malawi is among the world's least-developed countries. The economy is heavily based on agriculture, with a largely rural population. The country Gross National Income (GNI) per capita at purchasing power parity (PPP) is estimated at $860 while the world average is estimated at $10,780 [18, 19]. Ninety-one percent of Malawians live below 2 dollars (US) per day. The country's Human Development Index is estimated at 0.400, which gives the country a rank of 171 out of 187 countries with comparable data [19].

Malawi has a low life expectancy (53 years) and high infant mortality (66 deaths per 1,000 live births) compared to the world' average (70 years and 41 deaths per 1,000 live births). Averages for sub-Saharan Africa are estimated respectively at 55 years and 80 deaths for 1,000 live births. There is a high prevalence of HIV/AIDS, especially among women with about 13.6 percent HIV positive [18].

Malawi has actively responded to HIV since 1985 after the first AIDS case was

reported. In 1988, the government created the National AIDS Control Program to coordinate the country's HIV/AIDS education and prevention efforts. The Public Sector continues to set aside a minimum of two percent of their recurrent budget to support HIV and AIDS programme [20]. The HIV national commission budget has increased from US $98.1 million in 2010 to US $113.51 million in 2011 [20]. According to the Malawi 2012 Global AIDS Response progress report:

- Distribution of leaflets and HIV radio and TV programs. During the 2010-2011 financial year, 1,477 radio and 429 television (TV) programs were produced.
- In 2010 and 2011, around 3.8 million young people (50 percent males and 50 percent females) have been trained on life skills education (LSE) each year.
- Since 2003, the number of condoms distributed per capita has been increasing. Cumulatively, 21,049,592 condoms were distributed in the 2009-2010 fiscal year. During the fiscal year 2010-2011 the annual cumulative total of 26,461,079 condoms were distributed.
- The number of sites providing Prevention of Mother to Child Transmission (PMTCT) services has also been increased from152 facilities in 2006 to 544 sites 2011.
- Antiretroviral Therapy has been provided free of charge in the public sector since 2004. Number of patients alive and on treatment has increased from 10,761 in 2004 to 322,209 in 2011.

Data sources

This study uses data from the 2010 Malawi Health and Demographic Surveys (MDHS). The inclusion of HIV testing in the 2010 MDHS offers the opportunity to identify socioeconomic profile of women age 15-49 living with HIV. Participation in HIV testing was voluntary. To ensure confidentiality, case numbers (and not names) were used in linking the HIV test results to individual and household characteristics.

A subsample of one-third of the households was selected to conduct HIV testing for eligible women age 15-49 years. Ninety percent of all 2010 MDHS women who were eligible (8,174) for testing were interviewed and consented to HIV tests. The principal mode of HIV transmission in Malawi is heterosexual contact; therefore our analyses

focus on 6,395 women who ever had sexual intercourse. Details on the sample design are provided elsewhere [21,22].

Variables

The dependent variable for this analysis is HIV status, characterized as positive or negative blood test. The independent variables include 12 main variables grouped into two major types including, demographic and reproductive behavior variables (age, age at first sex, marital status, age at first birth, number of children ever born, Experience in premarital childbearing, and relationship to the head of household), and socioeconomic and contextual variables (religion, region of residence, place of residence, education, and household wealth index).

The choice of these variables is guided by the literature on factors associated with HIV in sub-Saharan Africa [10-15]. Most-at-risk populations refer to a combination of several factors because socioeconomics factors associated with HIV are not mutually exclusive.

Statistical analyses

Statistical analyses used Pearson Chi-square and Chi-square Automatic Interaction Detector (CHAID) using SPSS version 21. We used weighted data to take into account the complexity of the DHS design. We performed Pearson chi-square to identify associations between the HIV status (positive, negative) and demographic and reproductive behavior variables as well as socioeconomic and contextual variables.

We used CHAID to identify HIV predictors and the most-at-risk groups among women living with HIV [23]. CHAID is a nonparametric technique that makes no distributional assumptions on outliers, collinearities, heteroskedasticity, or distributional error structures. The dependent variable and predictor variables can be nominal (categorical), ordinal (ordered categories ranked from small to large), or interval (a "scale").

CHAID uses regression and classification algorithms and offers a non-algebraic method for partitioning data that lends itself to graphical displays. The method is a sequential fitting algorithm and its statistical tests are sequential with later effects being dependent upon earlier ones, and not simultaneous as would be the case in a regression model or analysis of variance where all effects are fit simultaneously.

CHAID solves the problem of simultaneous inference using Bonferroni multiplier. It automatically tests for and merges pairs of homogenous categories in independent variables.

At each step, CHAID chooses the independent (predictor) variable that has the strongest interaction with the HIV status (dependent variable). The variable having the strongest association with HIV status becomes the first branch in a tree with a leaf for each category that is significantly different relative to be HIV positive. It then assesses the category groupings, or interval breaks to pick the most significant combination of variables. The process is repeated to find the predictor variable on each leaf most significantly related to HIV status, until no significant predictors remain.

The developed model is a classification tree (or data partitioning tree) that shows how major "types" formed from the independent (predictor or splitter) variables differentially predict a criterion or dependent variable. The method permits also identification of women who are likely to be members of a particular group (Segmentation), and assign cases into one of several categories, such as high-, medium-, and low-risk groups (stratification). Selecting a useful subset of predictors from a large set of variables for use in building a formal parametric model (Data reduction and variable screening); Identify relationships that pertain only to specific subgroups and specify these in a formal parametric model (Interaction identification); and recoding group predictor categories and continuous variables with minimal loss of information. Categories of each predictor are merged if they are not significantly different with respect to the dependent variable (Category merging and discretizing continuous variables).

The output of CHAID prediction model is displayed in hierarchical tree-structured form, and consists of several levels of branches: root node, parent nodes, child nodes and terminal nodes. The root node, “Node 0” or “initial node” is the dependent variable or the target variable, HIV prevalence in our case. Parent node is the upper node compared with nodes on the subsequent (lower) level, whereas the lower level nodes are called child nodes. The terminal node or external node is any node that does not have child nodes. It is the last categories of the CHAID analysis tree.

For each terminal node CHAID provides in a table the following indicators:

1. Node: provides the number and percentage of people belonging to a selected category j (demographic weight in the sample);
2. Gain for each terminal node is the number of women who are living with HIV in absolute. In percentage, gain is calculated as number of women living with HIV in a selected node divided to the total of women living with HIV. Part of the population with the observed characteristic (living with HIV) in a selected category compared to total of women living with HIV.
3. Response can be defined as HIV prevalence among women belonging to each terminal node. Number of women living with HIV in a selected node divided by total of women of the node.
4. Gain index percentage reporting how much greater the proportion of a given target category at each node differs from the overall proportion. It is obtained by dividing the proportion of records that present category j in each terminal node into the proportion of records presenting category j in the total sample. Thus, it represents the increased probability of belonging to the selected category j that contains the records presenting the characteristics defined for each terminal node.

The method allows: (1) identifying complex interactions between variables across the measurement space; (2) Identifying the most significant explanatory variable; (3) Merge categories of nominal variable and categorize continuous variables without loss of information. Furthermore, CHAID as other decision trees can be applied to any data structure.

However, CHAID have two major shortcomings. First the method needs large sample sizes to work effectively because it uses multi-way splits. Indeed, with small sample sizes the respondent groups can quickly become too small for reliable analysis. Last, CHAID does not take into account the hierarchical structure of this data.

Results

Sampling description

Table 1 presents the characteristics of the study population. Since the principal mode of HIV transmission in Malawi is heterosexual contact, our analyses focus on women who ever had sexual intercourse. The distribution of the sample by age shows that

more than half (56 percent) of the populations are age less than 30 years old. The average age of the sample is estimated at 29 years old. Women who are in union (i.e., currently married or living with a man) constitute about 77 percent. The proportion of women who have never been married is estimated at 8 percent. Regarding the relationship to the head of household, the majority of women are spouse (63 percent).

Table 1 also shows that the majority of women (more than 80 percent) live in rural areas. By region, the majority of women live in Southern Region and the Central Region. Furthermore, 17 percent of women never attended school, while more than 60 percent have attended only primary school. Regarding the reproductive behavior, a large majority of women had their first sexual intercourse before 20 years (average 16.6 years old).

Table 1 – Description of the sample

Socioeconomic and demographic Characteristics	Percentage		Number[2]
	Weighted[1]	Unweight	
Age			
15-19	11.5	11.6	744
20-24	21.5	20.8	1,327
25-29	22.7	21.9	1,402
30-34	15.2	15.7	1,001
35-39	13.0	12.7	814
40-44	8.6	9.1	579
45-49	7.5	8.3	528
Average	29.6	29.8	-
Age at first sex			
<15	19.1	19.2	1,230
15-19	68.5	68.4	4,376
20-24	11.2	11.1	708
25+	1.3	1.3	81
Average	16.6	16.6	16.6
Marital status			
Single	7.5	7.6	484
In union	77.4	77.1	4,929
Ever married	15.1	15.4	982
Number of ever born children			
0	9.9	9.6	617
1+	90.1	90.4	5,778
Age at first birth			
Never give birth	10.3	10.6	660
< 20 years old	64.8	64.2	4,144
20 +	24.9	25.2	1,591
Ever had premarital child			
No	88.7	88.4	5,652
Yes	11.3	11.6	743
Relationship to the head of household			
Head of household	19.4	19.0	1,213
Spouse	62.6	62.4	3,992
Daughter & Grand daughter	11.0	11.7	748
Others	7.1	6.9	442
Province of residence			
Northern	11.1	17.5	1,122
Central	42.2	34.1	2,181
Southern	46.7	48.4	3,092
Place of residence			
Urban	19.2	13.1	837
Rural	80.8	86.9	5,558
Religion			
Catholic	21.2	20.6	1,316
Protestant	24.3	25.2	1,610
Other Christians	39.7	42.3	2,708
Muslim	13.5	10.9	695
Others	1.3	1.0	66
Education			
None	17.5	16.6	1,060
Primary	63.8	66.4	4,246
Secondary +	18.7	17.0	1,089
Household wealth quintiles			
Poorest	17.6	19.0	1,215
Poorer	20.1	20.6	1,319
Middle	19.7	20.9	1,334
Richer	19.3	20.7	1,323
Richest	23.3	18.8	1,204
Total			6,395

[1] Interpretations are based on weighted percentage.
[2] Unweight crude numbers

HIV prevalence by selected background characteristics

Table 2 describes HIV prevalence in Malawi by women's selected background characteristics. Overall, 14 percent of the women are HIV positive. All independent variables are statistically associated with HIV infection status except for religion.

HIV infection prevalence was high (20 percent) among women aged 30-39 years. Women who are no longer in union (widowed, divorced and separated) had significantly higher prevalence (4 percent) compared to those who have never been in a marital union (11 percent). HIV prevalence was high among heads of household (25.0 percent). Furthermore, while 25 percent of women in urban area were HIV positive, the prevalence was less than half (12 percent) compared to their counterparts from the rural areas. The HIV epidemic shows regional heterogeneity with a higher prevalence (20 percent) observed in the Southern region. Women with secondary education had higher HIV prevalence compared to those who never attended school (17 percent versus 14 percent). Regarding the household wealth quintiles the prevalence of HIV infection is higher among the women from the highest quintiles. With reference to sexual and reproductive behavior, HIV prevalence was higher among women who had their first sexual intercourse before the 15th birthday or from their 25th birthday, and /or who have experienced premarital childbearing.

Table 2 – HIV prevalence by selected socioeconomic and demographic characteristics

Socioeconomic and demographic Characteristics	HIV+ Prevalence (%)	Total (N)	Chi-Square	p-value
Age				
15-19	5.0	744		
20-24	6.9	1,327		
25-29	12.4	1,402		
30-34	19.8	1,001	190.35	<0.001
35-39	21.4	814		
40-44	18.7	579		
45-49	16.9	528		
Age at first sex				
<15	15.9	1,230		
15-19	13.1	4,376	9.13	0.028
20-24	12.4	708		
25&+	18.5	81		
Marital status				
Single	7.9	484		
In union	10.7	4,929	316.15	<0.001
Ever married	31.5	982		
Number of ever born children				
0	7.9	617	18.80	<0.001
1&+	14.2	5,778		
Age at first birth				
Never give birth	9.1	660		
< 20 years old	14.1	4,144	12.96	0.002
20 +	14.3	1,591		
Ever experience premarital childbearing				
No	12.8	5,652	26.99	<0.001
Yes	19.8	743		
Relationship to the head of household				
Head of household	25.0	1,213		
Spouse	10.2	3,992	179.93	<0.001
Daughter & Grand daughter	11.9	748		
Others	17.0	442		
Region of residence				
Northern	10.0	1,494		
Central	9.5	3,062	184.90	<0.001
Southern	20.0	4,444		
Place of residence				
Urban	24.7	1,156	157.00	<0.001
Rural	12.3	7,844		
Religion				
Catholic	12.6	1,316		
Protestant	14.3	1,610		
Other Christians	13.4	2,708	2.66	0.616
Muslim	14.8	695		
Others	13.6	66		
Education				
None	13.9	1,060		
Primary	12.8	4,246	10.73	0.005
Secondary +	16.6	1,089		
Household wealth quintiles				
Poorest	10.3	1,215		
Poorer	10.9	1,319		
Middle	11.6	1,334	88.34	<0.001
Richer	14.3	1,323		
Richest	21.5	1,204		
Total	13.6	6,395		

HIV predictors in Malawi: results from CHAID analysis

Table 3 shows summary information on the specifications used to build the CHAID model and the resulting model. Twelve independent variables were specified, but only six were included in the final model. The variables such as age at first sex, age at first birth and female education did not make a significant contribution to the model, so they were automatically dropped from the final model. Overall, there are 21 nodes among which 13 terminal nodes. Parent nodes include at least 100 cases whereas child nodes account for 50 cases in minimum.

Model components	Model specification	Results
Dependent variable	HIV status	HIV+=13.6%
Independent Variables	Age, Age at first sex, Marital status, Ever had a child, age at first birth, Experience premarital childbearing, Relationship to the head of household, Region of residence, Place of residence, Education, Wealth Index, Religion	**Marital status**, Age, Wealth Index, Relationship to the head of household, Region of residence, Place of residence
Maximum Tree Depth	3	3
Minimum Cases in Parent Node	100	100
Minimum Cases in Child Node	50	50
Number of Nodes	-	21
Number of Terminal Nodes	-	13
Overall predicted correct percentage		86.8

The tree diagram depicted in Figure 1 shows that "Marital status" is the best predictor of HIV status among women in Malawi (Chi-square = 313.22, p-value<0.0001).

The tree is split into two main notes. Node 1 includes women formerly in union; and Node 2 is composed of women in union and never married women.

Node 1- Women formerly in union. For this group, including divorced, widowed, and not living together, age is the best predictor of HIV prevalence (Chi-square=56.30, p-value <0.001). The group is further splits into four sub age groups: 15-24, 25-29, 30-34 and 40-49, and 35-39.

Among women aged 30-34, 40-44 and 40-49 (Node 3) with HIV prevalence of 37 percent, household wealth quintiles are the best predictor of HIV infection (Chi-square=29.81, p-value<0.001). Indeed in this group women in the highest wealth

quintile (Node 9) have HIV prevalence about three times higher than their counterpartners from the lowest wealth quintile-Node 11 (60 percent versus 22 percent). For women in the age groups of 14-19 and 20-24 years (Node 4) with a HIV prevalence of 13 percent, the relationship to the head of household is the best predictor of HIV infection (Chi-square=11.1, p-value <0.003) whereby women head of household (Node 12) have higher HIV prevalence compared to other women with different relationship to the head of household – Node 13 (24 percent versus 7 percent). The region of residence is the best predictor of HIV infection among women aged 35-39 (Chi-square=11.5, p-value<0.002) with women living in the Southern region (Node 14) having HIV prevalence about twice the one of the women from the Central and Northern regions-Node 15 (59 percent versus 31 percent). Among women aged 25-29 years (Node 6) accounting for 3 percent of the study population with HIV prevalence of 27 percent, age remains the only significant and final predictor of HIV prevalence.

Node 2- This group includes women in union (married or living together) and those have **never been in union**, representing 85 percent of the study population and have HIV prevalence of 10 percent. Place of residence (rural or urban) is the best predictor of HIV infection with a higher prevalence in urban area (Node 8) compared to rural area – Node 7 (21 percent vs 9 percent, Chi-square = 89.8, p-value<0.001).

For the women living in the rural area (Node 7) and representing 74 percent of the population, the best predictor for HIV infection is age (Chi-square=86.0, p-value <0.001) with the highest prevalence among women aged 30-44 years (13 percent) followed by the age group 25-49 (Node 18: 9 percent) and the age group 15-24 (Node 18: 4 percent). Similarly, age is strong predictor of HIV infection (Chi-square=86.0, p-value <0.001) among women living in urban area (Node 8) whereby the age group 30-49 (Node 19) has a prevalence about twice the one among the age group 15-29-Node 20 (29 percent versus 15 percent).

HIV risk groups in Malawi

There are in total 13 homogenous subgroups or terminal nodes. Table 4 describes these 13 subgroups (terminal nodes) in terms of their composition, demographic weight in the sample (columns 1 and 2), their share in HIV burden (columns 3 and 4)

and their corresponding HIV prevalence (column 5). The 13 homogenous sub-groups could be grouped into 5 major groups.

Group 1 represents 3 percent of the sample with an overall HIV prevalence of 59 percent. This group accounts for 11 percent of all the women HIV positive. Group 1 includes two subgroups: a) women in union disruption living in richest household and age 30-34 or 40-49 years old; and b) women in union disruption living in the Southern region and age 35-39 years old.

Group 2 represents 5 percent of the sample with an overall HIV prevalence of 35 percent and accounts for 12 percent of all HIV positive women. This group is composed of two subgroups including women in union disruption living in intermediate wealth households (non-poorest and non-richest households) age 30-34 or 40-49; and women in union disruption age 35-39 and living in the Northern or Central region.

Group 3 represents about 10 percent of study population with an overall HIV prevalence of 27 percent and accounts for 20 percent of all HIV positive women. This group is divided into four subgroups: a) Never married and women in Union, living in urban area and age 30-49; b) Formerly in union (widowed or divorced) women age 25-29; c) Young women (15-24) formerly in union who are head of household; d) Formerly in union women living in poorest household and age 15-24, 30-34 or 40.

Group 4 represents about 33 percent of the study population with an overall HIV prevalence of 14 percent and account for 33 percent of all the HIV positive women. This group includes adolescent (15-19), never married women or in union living in urban area; and never married or women in union living in rural area age 30-44.

Group 5 represents 50 percent of the study population and has the lowest HIV prevalence of 7 percent, but account for 23 percent of all the HIV positive women. This group includes three subgroups: a) Never married or women in union living in rural area and age 25-29 or 45-49; b) Young women age 15-24 who are non longer in union and are not head of household; and c) Young women (15-24) who are never married or in Union and are living in rural areas.

Table 4 reports also the gain index percentage (column 6) expressing how much greater the proportion of a given target group at each node differs from the overall proportion. The index percentage is very high among women belonging to group with

high HIV prevalence but with small demographic weight in the population (categories 1 to 3). Opposite values are observed among groups accounting for the major part of the sample among which HIV prevalence is low (Group 5). The Index is equal to 100 in category 4.

Table 4 – CHAID Gains for Nodes

Group/Node No	Node description	Node		Gain		% HIV[5]	Index[6]
		N[1]	%[2]	N[3]	%[4]		
Group 1		165	2.6	98	11.2	59.4	435.6
9	Formerly in union, 30-34/ 40-49 and richest	82	1.3	49	5.6	59.8	438.2
14	Formerly in union, 35-39 and Southern region	83	1.3	49	5.6	59.0	433.0
Group 2		308	4.8	108	12.4	35.1	257.2
10	Formerly in union, 30-34/40-49 and Rich/Middle/Poor	246	3.8	89	10.2	36.2	265.3
15	Formerly in union, 35-39 and North/Central regions	62	1.0	19	2.2	30.6	224.7
Group 3		640	10	171	19.5	26.7	195.9
19	Never married/in Union, living in urban area, 30-49	268	4.2	78	8.9	29.1	213.4
6	Formerly in union and 25-29	180	2.8	49	5.6	27.2	199.6
12	Formerly in union, 15-24 and Head of household	67	1.0	16	1.8	23.9	175.1
11	Formerly in union, 15-24, 30-34/40-49 and Poorest	125	2.0	28	3.2	22.4	164.3
Group 4		2,117	33.1	291	33.3	13.8	100.8
20	Never married/in Union, living in urban area, 15-19	417	6.5	64	7.3	15.3	112.6
16	Never married/in Union, living in rural area, 30-44	1,700	26.6	227	26.0	13.4	97.9
Group 5		3,165	49.5	204	23.3	6.5	42.3
18	Never married/in Union, rural area, 25-29 and 45-49	1,404	22.0	126	14.4	9.0	65.8
13	Formerly in union, 15-24 and Not Head of household	137	2.1	10	1.1	7.3	53.5
17	Never married/in Union, living in rural area, 15-24	1,624	25.4	68	7.8	4.2	30.7
Total		6,395	100	872	100	13.6	-

Notes:;[1] Number of cases per node (demographic size in the sample);
[2] Demographic size in percentage = (.1/Σ.1)*100;[3] Number of HIV women;[4] Demographic size among HIV positive women in percentage = (.3/Σ.3)*100; [5] HIV prevalence in percentage = (.3/Σ.1)*100; [6] Node Index = ((.3/Σ3)/(.1/Σ.1))*100

Discussions

This paper aimed to describe and profile HIV prevalence among women in Malawi. The study used Chi-square and CHAID techniques to analyze data from the Malawi 2010 DHS.

Analyses suggested three keys findings. First, consistent with previous studies [13, 24], findings from bivariate analysis and chi-square test showed high HIV prevalence among women in union dissolution, among the most educated women, women living

in wealthy households and/or among women living in urban areas. The finding confirmed also region heterogeneity in HIV prevalence. The Southern region being the most affected. In general the most educated women are more likely to marry husbands with high education level, and belonging to high socio-economic class of the society [25,26]. In parallel, relatively rich and better-educated men have higher rates of partner change because they have greater personal autonomy and spatial mobility [27, 28, 29, 30]. Women's economic dependence on their partners may also make it difficult for them to insist on safer sex (e.g. condom use). Concentration (about 50 percent) of the most educated, richest and urban women in the Southern region may explain high HIV prevalence in that region.

Second, results from CHAID models reported that marital status is the best predictor of HIV status among women in Malawi. Non-poorest women who are no longer in union (widowed and divorced or separated) age 30-34 or 40-49 have significantly higher HIV prevalence. This may be because: (1) husband from the highest quintile or a male partner may have more access to transactional sex and other risk behaviors such as polygamy which may increase women's vulnerability to HIV; (2) wealthier HIV positive widowed may have better quality of life as well as better access to treatment and survive longer [31]. Furthermore, divorced and separated are more frequent among the most educated women with economic autonomy [32]. Their causes (polygyny and/or infidelity) as well as consequences (multiple sexual partnerships) are also factors associated with HIV prevalence [33,34].

Third and last, CHAID models depicted also different interactions between risk factors and profiled HIV risk groups in Malawi. For instance, whilst overall HIV prevalence among women living in urban areas (25 percent) is twice the prevalence observed among women living in rural areas (12 percent), HIV prevalence is estimated at 15 percent among never married or women in union living in urban areas age 15-29, and at 13 percent among never married or women in union living in the rural areas age 30-44. Likewise, whereas in general HIV prevalence is low among never married and women in union (10 percent), CHAID results revealed a higher HIV prevalence (29 percent) among never married and women in union age 30-49 who live in urban areas compared to: (1) women in union disruption age 15-24 (7 percent if they are not head of households and 23 percent for head of household); (2) women in union disruption age 25-29 (27 percent); and (3) women in union

dissolution age 30-34 and 40-49 who live in poorest households (22 percent).

These findings showed the complexity of different interactions that may present challenges to conventional regression models. Indeed, CHAID is a sequential fitting algorithm and its statistical tests are sequential with later effects being dependent upon earlier ones and not simultaneous as would be the case in a regression model or analysis of variance where all effects are fit simultaneously. Furthermore, CHAID allows automatic detection of interaction between variables.

In the light of these findings, it is noteworthy that to reduce number of new infection, interventions should be targeted and prioritized according to the prevalence and demographic size of different risk groups. Furthermore, policy makers' prioritization of interventions may depend also on preference for preventive interventions compared to treatment of and care for HIV infected people and/or to treatment of and care for AIDS-patients. In Thailand, for instance policy makers expressed a preference for target preventive interventions that are highly effective compared to care and treatment [35].

Regarding preventive interventions, the findings suggested that:

1. Couples (males and women in union) and never married people age 25-49 (nodes 16 and 18) living in rural areas should be the first targets using universal HIV testing, "Abstinence", "Be faithful" and "use condom" campaign. Indeed, this group includes 49 percent of the study population, among whom the HIV prevalence is estimated at 11 percent on average. About 40 percent of women living with HIV in Malawi belong to this category.
2. Young age 15-24 living in rural areas (node 17) and urban adolescent 15-19 (node 20) is the second most important target. This group account for 32 percent of the studied population and 15 percent of women living with HIV. Besides, majority of adults living with HIV may be infected during adolescence. Unfortunately, the available dataset could not provide information on time of infection.
3. The country develops and implements a social policy to protect single mothers. Indeed, though overall HIV prevalence is estimated at 6 percent on average among young women age 15-24 (Table 2) that prevalence is estimated above 20 percent among young women formerly in union and among young

women who ever experience premarital childbearing living in urban areas. Likewise, HIV prevalence is very high among women in union disruption (32 percent on average) compared to other groups (10 percent). Though this high prevalence may be due to male mortality, some women in union disruption may be vulnerable because of poverty.

With reference to treatment and care, higher priority must be given to promoting HIV test, monitoring and evaluation of equity in access to treatment among women in union disruption and never married or women in union age 30-49 living in urban areas. Indeed, formerly in union women represent only about 13 percent of women of reproductive age in Malawi, they have the higher HIV prevalence ranges between 22 percent observed among poorest and 60 percent among richest.

Nevertheless, to achieve zero new infection as part of MDG 6, there is need of more comprehensive policy to combat HIV because of the complexity of HIV socioeconomic profile in Malawi. There are several groups built from several socioeconomic categories depending on individual marital status, wealth index, age, place of residence, and relationship to the head of household. In South Africa, Bendavid et al. [36] revealed that scaling up all aspects of HIV care including universal testing and treatment was associated with a life expectancy gain of 22.2 months, and new infections were 73 percent lower.

From the methodological point of view, this study has some limitations, which do not detract from its scientific importance and contribution. First, this study used cross-sectional data from the Demographic and Health Surveys, which does not permit one to draw causal association between HIV status and the associated factors. For instance, whether HIV infection has occurred before, during or after the union. Last, CHAID model ignores the hierarchical structure of the Demographic and Health Survey data and need large sample size.

In conclusion, this study recommends: (1) design and implementation of targeted interventions taking into account HIV prevalence and the demographic size of different groups at risk; (2) reinforcement of integration of family planning and HIV/AIDS services because the population understudied includes women of reproductive health. Integrating the two services (HIV and FP) could be cost-effective; (3) Community health workers, households based campaign, reproductive

health services and reproductive health courses at school could be used as canon to achieve universal prevention strategy, testing, counseling and treatment.

Acknowledgements*: This research was supported by the Luxembourg National Research Fund (FNR). The authors thank Macro international for providing free the Malawi 2010 DHS data-sets including HIV data. We thank Prof. Guillaume Wunsch for his comments. Furthermore, we* have appropriate approval and authorization from the DHS datasets research ethics board for conducting HIV secondary data analysis.

Contributors: JBOE participated in conception and design, literature review, data analysis and interpretation, drafting the article, critical revisions for important intellectual content and approval of final article for submission. MK, NJM, EMZ and YY participated in conception and design, interpretation of results, critical revisions for important intellectual content and approval of final article for submission.

Data sharing statement: This study is based on the Malawi 2004 and 2010 Demographic and Health Surveys. These data are available on www.measuredhs.com. Access to individual HIV status as well as individual background information required authorization from MACRO ORC (www.measuredhs.com). The Website depicts the process.

References

[1] United Nations.*The Millennium Development Goals. 2012 Report* . New York: United Nations, 2012.

[2] UNAIDS *UNAIDS Report on the Global AIDS Epidemic 2012. Global report.* Geneva: UNAIDS, 2012. http://www.unaids.org/en/resources/campaigns/20121120_globalreport2012/globalreport/

[3] International HIV / AIDS Alliance. (2010). *The cost efficiency of HIV prevention for vulnerable and most-at-risk populations and the reality of funding. What's Preventing Prevention Campaign Briefing 2.* Hove: International HIV / AIDS Alliance.

[4] UAIDS. *World AIDS day report 2011. How to get to zero: Faster, smarter, Better.* Geneva: UNAIDS, 2011.

[5] Udeh, B.; Udeh, C. & Graves, N., Perinatal HIV transmission and the cost-effectiveness of screening at 14 weeks gestation, at the onset of labour and the rapid testing of infants, *BMC Infectious Diseases* 2008, 8:174 doi:10.1186/1471-2334-8-174. http://www.biomedcentral.com/1471-2334/8/174

[6] Roux, P.; Henley, L.; Cotton, M.; Eley, B., Burden and cost of inpatient care for HIV-positive paediatric patients--status in the Cape Town metropole during the second week of March 1999. Paediatric HIV Census Group. *South African Medical Journal*, 2000, 90(10):1008-1011

[7] Kumar, R.; Mehendale, S.M.; Panda, S.; Venkatesh, S.; Lakshmi, PVM.; Kaur, M.; Prinja, S.; Singh, T.; Virdi, N.K.; Bahuguna, P.; Sharma, A.K.; Singh, S.; Godbole, S.V.; Risbud, A.; Manna, B.; Thirumugal, V.; Roy, T.; Sogarwal, R.; Pawar, N.D. Impact of targeted interventions on heterosexual transmission of HIV in India, *BMC Public Health*, 2011, 11:549, http://www.biomedcentral.com/1471-2458/11/549

[8] Marseille, E.; Kahn, J.G.; Billinghurst, K.; Saba, J. Cost-effectiveness of the female condom in preventing HIV and STDs in commercial sex workers in rural South Africa. *Social Sciences & Medicine*. 2001;52(1):135-148.

[9] Sweat, M; Gregorich, S; Sangiwa, G.; Furlonge, C.; Balmer, D.; Kamenga, C.; Grinstead, O. & Coates, T., Cost-effectiveness of voluntary HIV-1 counselling and

testing in reducing sexual transmission of HIV-1 in Kenya and Tanzania. *Lancet*. 2000 Jul;356 (9224):113-121

[10] Durevall, D, & Lindskog, A. *HIV/AIDS, Adult Mortality and Fertility: Evidence from Malawi. Working Papers in Ecomics, No 284*. Göteborg: School of Business, Economics and Law, University ofGothenburg, 2007.

[11] Rombo, D. *Marital risk factors and HIV infection among women: A comparison between Ghana and Kenya,* Ph.D. Dissertation. Minneapolis: University of Minnesota, 2009.

[12] Öjteg, K. *Socio-Economic determinants of HIV in Zambia. A district level Analysis*. Ph.Dissertation. Lund: Department of Economics at the University of Lund, 2009.

[13] Magadi, M., & Desta , M. A multilevel analysis of the determinants and cross-national variations of HIV seropositivity in sub-Saharan Africa: Evidence from the DHS. *Health & Place*, 2011; 17 (5): 1067–83.

[14] Magadi, M. Understanding the gender disparity in HIV infection across countries in sub-Saharan Africa: evidence from the Demographic and Health Surveys. *Sociology of Health & Illness,* 2011; 33 (4): 522-39.

[15] Asiedu, C., Asiedu, E., & Owusu, F. The Socio-Economic Determinants of HIV/AIDS Infection Rates in Lesotho, Malawi, Swaziland and Zimbabwe. *Development Policy Review ,* 2012; 30 (3): 305-326 .

[16] Green, E. *Rethinking AIDS Prevention. Learning from Successes in Developing Countries* . Westport, CT: Praeger publishers, 2004

[17] International HIV / AIDS Alliance. *The cost efficiency of HIV prevention for vulnerable and most-at-risk populations and the reality of funding. What's Preventing Prevention Campaign Briefing 2*. Hove: International HIV / AIDS Alliance, 2010.

[18] Population Reference Bureau, *2012 World Population Data sheet*, Washington DC, PRB, 2012.

[19] United Nations Development Programme/UNDP. Sustainability and Equity: A better future for all, Human Development Report 2011. New York, UNDP, 2011.

[20] Malawi, HIV national Commission. National 2012 Global AIDS Response Progress

Report: Malawi Country Report. Malawi, HIV National Commission, 2012, 67 p.

[21] Malawi National Statistical Office (NSO); ORC Macro. *Malawi Demographic and Health Survey 2004.* Calverton, Maryland: NSO and ORC Macro, 2005.

[22] Malawi National Statistical Office (NSO); ICF Macro. *Malawi Demographic and Health Survey 2010.* Zomba, Malawi, and Calverton, Maryland, USA: NSO and ICF Macro, 2011.

[23] Kass, G. V. An Exploratory Technique for Investigating Large Quantitaties of Categorical Data. *Applied Statistics,* 1980; 29 (2): 119-127.

[24] Adair, T. *HIV Status and Age at First Marriage among Women in Cameroon.* DHS Working Paper No 33, Calverton, Maryland, USA: Macro International Inc, 2007.

[25] Cleland, J., & Van Ginneken, J. Maternal education and child survival in developing countries: The search for pathways of influence. *Social Sciences and Medicine* , 1988 ; 27(12):1357-1368.

[26] Barrett, H., & Brown, A. Health, hygiene, and maternal education: Evidence from Gambia. *Social Sciences and Medicine*, 1996; 43(11): 1579-1590.

[27] Ainsworth, M. & Semali, I. Who is most likely to die of AIDS? Socioeconomic correlates of adult deaths in Kagera Region, Tanzania. In: *Confronting AIDS: evidence for developing world*. Edited by Ainsworth M, Fransen L, Over M. Washington, D.C.: The World Bank; 1998: 95-110.

[28] Gregson S, Garnett GP, Nyamukapa CA, Hallett TB, Lewis JJ, Mason PR, et al. HIV decline associated with behaviour change in eastern Zimbabwe. *Science,* 2006; 311:664–666.

[29] Hargreaves JR, Glynn JR. Educational attainment and HIV-1 infection in developing countries: a systematic review. *Tropical Medical and International Health*, 2002; 7:489–498.

[30] Gillespie S., Kadiyala, S. & Greener. Is poverty or wealth driving HIV transmission? *AIDS*, 2007; 21 (suppl 7):S5–S16.

[31] Rodrigo, C., & Rajapakse, S. (2010). HIV, poverty and women. *International Health*, 2010, 2(1): 9-16.

[32] Takyi, B., & Broughton, C. Marital Stability in Sub-Saharan Africa: Do Women's Autonomy and Socioeconomic Situation Matter? *Journal of Family and Economic Issues* , 2006; 27 (1): 113-132.

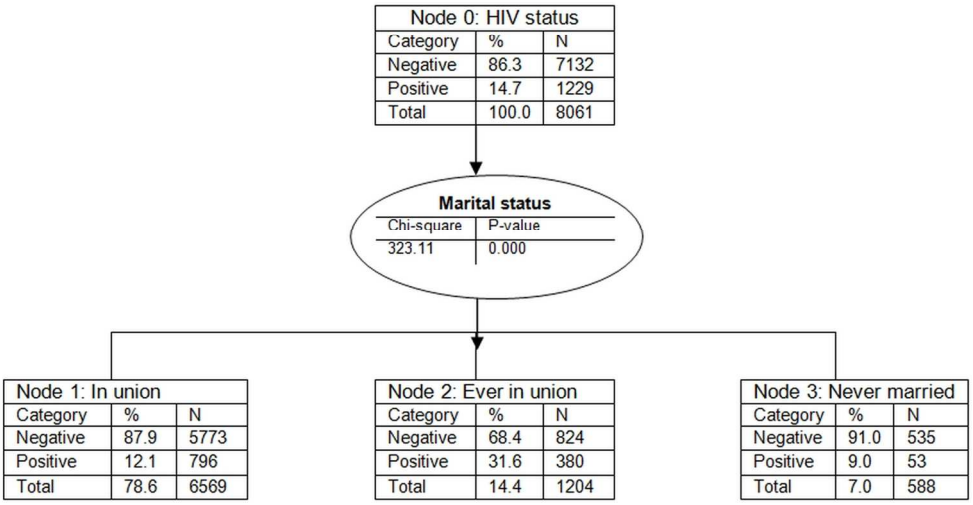
[33] Reniers, G. Divorce and remarriage in rural Malawi. *Demographic Research*, 2003; Special Collection 1, article 6: 175-206. www.demographic-research.org/special/1/6/S1-6.pdf.

[34] Reniers, G. Marital strategies for regulating exposure to HIV. *Demography*, 2008; 45 (2): 417-438.

[35] Youngkong, S.; Baltussen, R.; Tantivess, S.; Koolman, X.; & Teerawattananon, Y. Criteria for priority setting of HIV/AIDS interventions in Thailand: a discrete choice experiment. *BMC Health Services Research* 2010, 10:197. http://www.biomedcentral.com/1472-6963/10/197 retrieved on March 1, 2013

[36] Bendavid, E; Brandeau, M. L.; Wood, R.; Owens, D. K. Comparative Effectiveness of HIV Testing and Treatment in Highly Endemic Regions. *Archives of Internal Medicine, 2010;170 (15):1347-1354*.

Figure 1 - HIV prevalence in Malawi: tree diagram



154x90mm (300 x 300 DPI)

STROBE 2007 (v4) Statement—Checklist of items that should be included in reports of *cross-sectional studies*

Section/Topic	Item #	Recommendation	Reported on page #
Title and abstract	1	(*a*) Indicate the study's design with a commonly used term in the title or the abstract	1
		(*b*) Provide in the abstract an informative and balanced summary of what was done and what was found	2
Introduction			
Background/rationale	2	Explain the scientific background and rationale for the investigation being reported	4
Objectives	3	State specific objectives, including any prespecified hypotheses	4
Methods			
Study design	4	Present key elements of study design early in the paper	5
Setting	5	Describe the setting, locations, and relevant dates, including periods of recruitment, exposure, follow-up, and data collection	5
Participants	6	(*a*) Give the eligibility criteria, and the sources and methods of selection of participants	5
Variables	7	Clearly define all outcomes, exposures, predictors, potential confounders, and effect modifiers. Give diagnostic criteria, if applicable	6
Data sources/ measurement	8*	For each variable of interest, give sources of data and details of methods of assessment (measurement). Describe comparability of assessment methods if there is more than one group	5-6
Bias	9	Describe any efforts to address potential sources of bias	*5-7*
Study size	10	Explain how the study size was arrived at	5
Quantitative variables	11	Explain how quantitative variables were handled in the analyses. If applicable, describe which groupings were chosen and why	6
Statistical methods	12	(*a*) Describe all statistical methods, including those used to control for confounding	6-7
		(*b*) Describe any methods used to examine subgroups and interactions	6-7
		(*c*) Explain how missing data were addressed	6-7
		(*d*) If applicable, describe analytical methods taking account of sampling strategy	6-7
		(*e*) Describe any sensitivity analyses	6-7
Results			

Participants	13*	(a) Report numbers of individuals at each stage of study—eg numbers potentially eligible, examined for eligibility, confirmed eligible, included in the study, completing follow-up, and analysed	5
		(b) Give reasons for non-participation at each stage	5
		(c) Consider use of a flow diagram	5
Descriptive data	14*	(a) Give characteristics of study participants (eg demographic, clinical, social) and information on exposures and potential confounders	7-8
		(b) Indicate number of participants with missing data for each variable of interest	7-8
Outcome data	15*	Report numbers of outcome events or summary measures	7-8
Main results	16	(*a*) Give unadjusted estimates and, if applicable, confounder-adjusted estimates and their precision (eg, 95% confidence interval). Make clear which confounders were adjusted for and why they were included	8-12
		(*b*) Report category boundaries when continuous variables were categorized	8-12
		(*c*) If relevant, consider translating estimates of relative risk into absolute risk for a meaningful time period	8-12
Other analyses	17	Report other analyses done—eg analyses of subgroups and interactions, and sensitivity analyses	8-12
Discussion			
Key results	18	Summarise key results with reference to study objectives	12-14
Limitations	19	Discuss limitations of the study, taking into account sources of potential bias or imprecision. Discuss both direction and magnitude of any potential bias	12-14
Interpretation	20	Give a cautious overall interpretation of results considering objectives, limitations, multiplicity of analyses, results from similar studies, and other relevant evidence	12-14
Generalisability	21	Discuss the generalisability (external validity) of the study results	12-14
Other information			
Funding	22	Give the source of funding and the role of the funders for the present study and, if applicable, for the original study on which the present article is based	15

*Give information separately for cases and controls in case-control studies and, if applicable, for exposed and unexposed groups in cohort and cross-sectional studies.

Note: An Explanation and Elaboration article discusses each checklist item and gives methodological background and published examples of transparent reporting. The STROBE checklist is best used in conjunction with this article (freely available on the Web sites of PLoS Medicine at http://www.plosmedicine.org/, Annals of Internal Medicine at http://www.annals.org/, and Epidemiology at http://www.epidem.com/). Information on the STROBE Initiative is available at www.strobe-statement.org.

February 22, 2013

Mr. Richard Sands
Managing Editor, BMJ Open

RE: Manuscript ID bmjopen-2012-002459 entitled "Prevalence of HIV among women in Malawi: Identifying the most-at-risk groups for targeted and cost-effective interventions"

Dear Mr Richard Sands,

Please find enclosed our revised manuscript, which addresses the reviewers' concerns and suggestions. What follows is a point-by-point response to the comments provided as part of the review process. Each group of responses has been numbered to correspond with those on the comments. Moreover, in the revised manuscript we have highlighted in red colour the areas that have been modified compared to the original submission.

Reviewer(s)' Comments to Author:

Reviewer: Kandala Ngianga-Bakwin
University of Warwick, Warwick Medical School
No competing interest

This paper examines the relationship between the prevalence of HIV infection and HIV socioeconomic predictors as well as to identify the most at-risk groups of women in Malawi using consecutive cross-sectional surveys (the 2004 and 2010 Malawi Health and Demographic Surveys (MDHS)). Data were from 8,596 men and women aged 15-49 who participated in the MDHS to profile participants with high risk of HIV using Chi-square Automatic Interaction Detector (CHAID).

The CHAID procedure partitions data that lends itself to graphical displays and the method is a sequential fitting algorithm and its statistical tests are sequential with later effects being dependent upon earlier ones and not simultaneous as would be the case in a regression model or analysis of variance where all effects are fit simultaneously.

The authors do a good job profiling at risk groups of HIV infection women in Malawi and examining the prevalence of HIV infection and HIV socioeconomic predictors by applying CHAID. The statistical methods employed to describe the data (CHAID) is appropriate for this kind of data. The context is relevant; the findings are very informative and likely to contribute to policy in Malawi. Identifying at high risk of women with HIV infection reinforces the need for targeted programming based on the specific strengths and challenges of each community, even within an area as small as village. Nevertheless, the authors need to address some disadvantages and limitation of the methodology employed, the (CHAID) more appropriately.

Reply: We would like to thank the reviewer for his comments on our manuscript.

1. The methodology employed (CHAID) has some shortcomings. CHAID doesn't take into account the hierarchical structure of this data and the method (CHAID) is often used as an exploratory technique and needs rather large sample sizes to work effectively, since with small sample sizes the respondent groups can quickly become too small for reliable analysis.

Furthermore, CHAID methods are not able to assess and deal with the auto-correlation (dependence of nearby groups) in the data, non-linear (age) and time varying effects of covariates. Specifically, the study relied on the assumption of independence of the profiled high risk groups but in practice groups close to each other are related. CHAID models cannot resolve this above issue. Could the authors please comment on these issues?

Reply: The reviewer is right, like all statistical methods CHAID has some limitations including ignoring the hierarchical structure of the Demographic and Health Survey data, and need of large sample size. We have recognized these shortcomings in the document. CHAID is not only an exploratory technique though it cannot be use for testing causality because the technique has good prediction capacities. Using CHAID, one can uncover relationships between a dependent variable, HIV status in our case, and a host of predictor variables. Therefore, we used CHAID for identifying homogenous groups of women considering HIV prevalence risk. All our variables are categorical to take into account non-linearity of some variable such as age though CHAID identifies homogenous segments.

2. Another challenge of profiling high-risk groups of HIV prevalence rates using CHAID with the MDHS is that often the method cannot deal with the issue of interaction between variables such as rural/urban, education. Please comment!

Reply: Thanks for the comment. CHAID deals with issue of interaction between variable. The method assesses the category groupings, or interval breaks to pick the most significant combination of variables.

3. I have also some specific comments:

a) Article focus : The second bullet of this section is not a focus of this paper, please delete or amend;

Reply: Thanks for the comment. The bullet has been deleted as suggested.

b) Strengths and limitations of the study: in the second bullet, authors should rather detail how the use CHAID methodology has been able to improve the analysis of these data compared to other methods rather than mentioning what CHAID is able to perform;

Reply: We have mentioned one the advantage of this method. "The model depicted also different interactions between risk factors and profiled HIV risk groups in Malawi using several variables including marital status, wealth index, age, place of residence, and relationship to the head of household".

c) it is also worth mentioning that by not accounting for men risk factors in profiling women, the results might be limited because in the context of HIV in these settings, men have power over women sexuality, which might influence women risk factors.

Reply: Reviewer is right. However our recommendations take into account men in combating HIV (See last paragraph of page 15). "Couples (males and women in union) and never married people age 25-49 (nodes 16 and 18) living in rural

areas should be the first targets using universal HIV testing, “Abstinence”, “Be faithful” and “use condom” campaign. Indeed, this group includes 49 percent of the study population, among whom the HIV prevalence is estimated at 11 percent on average. About 40 percent of women living with HIV in Malawi belong to this category”.

Minor comments

4. Page 3 line 17, the HIV prevalence is said to be 13 %, which is different from line 26 p3, Line 33 page 3, Line 31 p4 and Line 38, p8. Please confirm the correct one.

Reply: The HIV prevalence among women of reproductive age is estimated at 13.6 percent. The number has been consistently reported in the document.

5. P8 Line 21. Since the majority of women (82%) live in rural areas. Can the author comments of the impact of this imbalance in the results and whether the methods used were able to account for it.

Reply: We have worked with weighted file to take into account the sampling design effect including population distribution by place of residence.

6. The notion of nodes is introduce in results section page 9 without explaining it in statistical the methods section. Please explain it for your readers in the statistical the methods section.

Reply: The notion of nodes has been introduced and explained in the statistical methods section as suggested.

7. Line 48 page 12 should read 2004 and 2010. Not 2004 and 2004 DHS. Please also explain clarify how the CHAID in this study offers a useful alternative to traditional regression models.

Reply: Thanks for the comments. The suggestion has been implemented taking into account that analyses are based on the 2010 DHS. We have inserted a paragraph to show advantages of using CHAID in this study.

As a data mining technique with a hierarchical nature, it allows identifying population subgroups that share similar characteristics. It provides a graphical display of the results, which facilitate their interpretation and can assist health decision makers in visualizing how predictors interact to define high-risk subgroups and understand the importance of each variable examined in predicting the modeled response. Furthermore, studies have found the two methodologies to be comparable in predictive performance.

8. P13 Lines 52-57 Family planning was not formerly examined in this study and I think this conclusion is not part of this study. Please revise.

Reply: Yes, the reviewer is right the study does not examine Family Planning. However, the population understudied includes women of reproductive health. Integrating the two services (HIV and FP) could be cost-effective.

Reviewer: Name: Kavita Singh

Position: PhD Scholar and Research Associate

Institution: All India Institute of Medical Sciences, New Delhi

Country: India

I have no competing / conflict of interest with the research study / manuscript I reviewed.

9. Research question needs to be clearly specified, whether the authors are interested in estimating prevalence of HIV in Malawi women or just aims to identify the predictors of HIV prevalence in Malawi women across different socioeconomic groups. Accordingly, the manuscript title can be clearly formed.

Reply: We have clarified the research question and objectives.

10. Study design: Pooling data from two cross-sectional survey conducted in 2004 and 2010, seems not to be methodologically sound from purist perspective. I think, Outcomes to two samples can be compared but can it be summed and then used for estimation of prevalence and its association with risk factors can be analyzed is a point of concern from my viewpoint.

Reply: Current analyses exclude the 2004 DHS data.

11. Total number of participants used from different surveys is not clear and again different total sample size has been written invariable through out the manuscript.

Reply: Number of participants has been checked, corrected and reported consistently.

12. Methods section, needs to described clearly.

Reply: We have added paragraphs and sentences to improve the method section.

13. Abstract section has some major limitations, which are compiled in my comments pasted below.

Reply: The abstract section has been improved to take into account the comments.

14. Manuscript can be improved further with a couple of thoughtful iterations.

Reply: We thank the reviewer for the suggestion.

15. Background section is not appropriately references at many places. Complete web URL's for UN MDG and UNAIDS reports should be provided.

Reply: References have been completed

16. STROBE checklist included at the end of the manuscript, though is indicated and marked to have covered all aspects, however I don't think the checklist is rigorously followed. For example a flow chart of the number of eligible participants included is marked as YES referring to Page 5 under Participants section, 28 of the manuscript, however no such flow diagram has actually been included. Authors may wish to review the checklist again and improve on the quality of reporting the study findings further.

Reply: Thank you for the comment. We have revised the checklist.

17. I am sorry to mention this, however interpretation of results and result tables are not clear and very ambiguous at many places. Needs to be closely reviewed and presented in concrete fashion.

Reply: Thanks for the comment.

18. A statement on appropriate approval from research ethics board / DHS dataset authorised personnel, for conducting secondary data analysis has not been included in the manuscript.

Reply: We have included the statement acknowledging special authorization from the DHS research ethics board for conducting secondary analysis of HIV datasets.

19. I think, the authors have identified an important area to be addressed in Malawi community and that is of identifying the most-at risk groups for HIV infection. There are few major observations, which I have made after reading through this manuscript many times. I have summarized below my comments and also pasted some of these on the manuscript which is enclosed along with these comments for author's reference:

Reply: We thank the reviewer for his kind comments. Comments have been addressed and suggestions have been implemented to make the document more insightful and informative in the Malawi's context for public health policy making.

20. Manuscript title doesn't reflect or is consistent with the actual study objectives, because the main objective of this study is not estimate the prevalence of HIV among women in Malawi but rather just to identify the women groups who are high risk of HIV and accordingly to inform the cost-effective interventions. Authors may consider to concise the manuscript title and make it more suggestive of the actual planned study.

Reply: The title has been changed to "Identifying HIV most-at-risk groups in Malawi for targeted interventions"

21. There are a number of data discrepancies observed at number of places, it might be because the authors did not discuss clearly column heading or the way they have based their calculations, but still there are discrepancies. For example, look into the attached manuscript, where the discrepancies in total number of participants included in the study have been flagged.

Reply: Column headings have been clearly labeled and numbers have been checked.

22. Page 2: Abstract; conclusion – line21; primary and secondary outcome measure can be clarified; line 37, needs to be re-worded to make it clear and consistent with actual findings.

Reply: We have clarified "HIV status (positive or negative)" is the outcome. Furthermore we have implemented the suggestion to clarify the sentence.

> *"Women in union and their partners as well as never married people age 25-49 (nodes 16 and 18) living in rural areas should be the first prevention targets. This population accounts for 49 percent of the sample among and 40 percent of women living with HIV in Malawi"*

23. Page 4; Line 21 – current costs of HIV treatment is presented as (US $ 4,707 over lifetime), is not clear for which setting, age group and time period. Also, whether it's the out of pocket expenditure or cost borne by government? I think, clarifying on this aspect will help readers better understand the context of cost of HIV treatment and its implications. Appropriate reference can be cited for this.

Reply: We have provided additional: "*Based on the global costs of HIV treatment in 2010 estimated between US$22-24 billion annually by 2015 and individual cost of US $4,707 over lifetime in order to reach global targets [i,ii], targeted interventions and evidence based prevention programmes have been advocated as cost-effective strategy to combat HIV/AIDS*".

24. Page 4; Line 21 – examples of few cost-effective interventions/strategies should be included to provide a comprehensive detail in the context.

Reply: We have cited some targeted cost-effective interventions research.

25. Page 5; line 34: Data sources: The approach of combining two cross sectional surveys conducted at two different time points (2004 and 2010) with a gap of 6 years, doesn't make the study methodologically sound. Authors can provide justification and specific reasoning behind combining cross-sectional data sets from two different period for estimating prevalence of HIV in the community, which can be flawed by many factors, such as HIV being a chronic type of condition, there might be higher chances of calculating the same person twice in prevalence estimation, when you combine two dataset for prevalence calculation. Additionally, there might be lots of demographic and socio-economic changes that can occur in the society over 6 years period and I would be concerned in the way then, how productive and reliable this exercise may result in by combining the cross-sectional survey datasets from 2004 and 2010. Just a suggestion, authors may consider restricting the statistical analysis to the participants surveyed in 2010, which has pretty good sample size of ~8174 participants. I don't think adding meager 2000 additional participants from 2004 survey would add any incremental benefits to the statistical analysis part. Interpretation of these numbers also seems to be faulty. I have added specific comments to the tables in enclosed manuscript.

Reply: We thank reviewer for the comment and suggestion. Current analyses exclude the 2004 DHS data.

26. Page 14; line 27-31 is repetition of same paragraph on the above page 13; line 51-56

Reply: Thanks for the comment. We have implemented the comment.

27. Page 8; line 7-13; Invariably in the results section, the average figure reported for many variables like average age, or percent of women never married doesn't clearly quote the figures represented in the Table 1. Also, it's unclear whether the authors have used weighted or un-weighted average.

Reply: We have interpreted weighted figures. We have updated the result section to take into account the comment.

28. Page 19; Table 2: At the bottom of the table year of surveys mentioned are 2004 and 2007. I found most part of statistical analysis and results section not very clear. Authors may be advised to make be concrete, consistent and reader friendly.

Reply: Thank for the comments. Variable year of surveys has been removed from the analysis and the table has been updated accordingly.

29. I would recommend, if authors can provide methodological sound justification of combining two survey datasets conduced at a gap of 6 years for estimating the prevalence and then calculating the socio-economic predictors, would be much appreciated. Also, all tables needs to be organized, clearly labeled and cross-verified for the total numbers.

Reply: Thank for the comments. Current analyses exclude the 2004 DHS data. All tables have been clearly labelled and numbers have been cross-verified as recommended.

We would like to thank the reviewers for thoughtful comments and suggestions. We truly appreciate your interest in our work. We believe that as a result of the review process our paper has greatly improved and hope that it is now acceptable for publication in BMJ Open.

Yours sincerely,

Jacques B.O. Emina, PhD

[i] UNAIDS *UNAIDS Report on the Global AIDS Epidemic 2012. Global report.* Geneva: UNAIDS, 2012. http://www.unaids.org/en/resources/campaigns/20121120_globalreport2012/globalreport/

[ii] International HIV / AIDS Alliance. (2010). *The cost efficiency of HIV prevention for vulnerable and most-at-risk populations and the reality of funding. What's Preventing Prevention Campaign Briefing 2.* Hove: International HIV / AIDS Alliance.

Identifying HIV most-at-risk groups in Malawi for targeted interventions

Jacques B. O. Emina [1,2,3], Nyovani Madise [4], Mathias Kuepie [2], Eliya M. Zulu [5], Yazoume Ye[6]

1 Department of Population and Development Studies, University of Kinshasa, P.O.BOX 176 Kinshasa XI, Democratic Republic of Congo

2 CEPS/INSTEAD 3, avenue de la Fonte, L-4364 Esch-sur-Alzette Luxembourg

3 INDEPTH Network, Accra, Ghana

4 Faculty of Social and Human Sciences, Room 4103 Building 58, University of Southampton, Southampton, SO17 1BJ.

5African Institute for Development Policy (AFIDEP), P. O. Box 14688-00800, Westlands, Nairobi, Kenya.

6 ICF International 11785 Beltsville Drive, Suite 300 Calverton, MD 20705, USA.

Correspondence to: Jacques B.O. Emina; jacques.emina@gmail.com

Abstract

Objectives: To identify HIV socioeconomic predictors as well as identify the most-at-risk groups of women in Malawi.

Design: cross-sectional survey

Setting: Malawi

Participants: The study used a sample of 6,395 women age 15-49 years from the 2010 Malawi Health and Demographic Surveys (MDHS).

Interventions: N/A

Primary and secondary outcome measures: Individual HIV status: positive or not.

Results: Findings from Pearson Chi-square and Chi-square Automatic Interaction Detector (CHAID) analyses revealed that marital status is the most significant predictor of HIV. Women who are no longer in union and living households in the highest wealth quintiles constitute the most-at-risk group; whereas the less-at-risk group includes young women (15-24) never married or in union and living in rural area.

Conclusion: In the light of these findings, this study recommends: (1) design and implementation of targeted interventions should consider the magnitude of HIV prevalence and demographic size of most at risk groups. Preventive interventions should prioritize couples and never married people age 25-49 living in rural areas because this group accounts for 49 percent of the study population and 40 percent of women living with HIV in Malawi; (2) With reference to treatment and care, higher priority must be given to promoting HIV test, monitoring and evaluation of equity in access to treatment among women in union disruption and never married or women in union age 30-49 living in urban areas; (3) Community health workers, households based campaign, reproductive health services and reproductive health courses at school could be used as canon to achieve universal prevention strategy, testing, counseling and treatment.

Keywords: HIV/AIDS, Malawi, CHAID methods, decision analysis, most-at-risk groups, targeted interventions.

Article summary

1. Article focus

- Targeted interventions and evidence based prevention programmes have been advocated as cost-effective strategy to combat HIV/AIDS.
- Who are the most-at-risk populations regarding HIV prevalence in Malawi? With HIV prevalence of about 14 percent among women of reproductive age, HIV/AIDS constitutes a drain on the labor force and government expenditures in Malawi.

2. Key messages

- We use data from the Malawi 2010 Demographic and Health Surveys to profile HIV most-at-risk groups of women in Malawi where about 14 percent of women are HIV positive.
- Our findings revealed that richest and formerly in union women constitute the most-at-risk group.
- To achieve zero new infection as part of MDG 6, there is need of more comprehensive policy to combat HIV because of the complexity of HIV socioeconomic profile in Malawi. There are several groups built from several socioeconomic categories depending on individual marital status, wealth index, age, place of residence, and relationship to the head of household.

Strengths and limitations of this study

- From our knowledge this study may be the first in Malawi to attempt to profile HIV most-at-risk groups of women in Malawi. The most-at-risk population refers to a combination of several factors because factors associated with HIV are not mutually exclusive.
- The major strength is the use of the Chi-square Automatic Interaction Detector (CHAID) to identify HIV predictors and the most-at-risk groups among women for intervention. CHAID uses regression and classification algorithms and offers a non-algebraic method for partitioning data that lends itself to graphical displays.
- This study has two major limitations. First, this study used cross-sectional data from the Demographic and Health Surveys, which does not permit one to draw causal association between HIV status and the associated factors. For instance, whether HIV infection has occurred before, during or after the union. Last, CHAID model ignores the hierarchical structure of the Demographic and Health Survey data and need large sample size.

Introduction

In 2000, the United Nations' Millennium summit identified the reduction of HIV prevalence as one of the eight fundamental goals for furthering human development. Though global HIV/ AIDS incidence is declining, HIV/AIDS has remained the leading cause of death in women of reproductive age in low-and middle-income countries, particularly in sub-Sahara Africa (SSA) [1]. The gap between the current state of HIV/AIDS and the UNAIDS goals of three zero (zero new HIV infections, zero discrimination, and zero AIDS related deaths) remains important. With barely two years remaining to the end-date of the Millennium Development Goals (MDG) target, HIV/AIDS remains a long-term global challenge [1].

Given the high cost of HIV/AIDs treatment estimated in 2010 to be globally between US$22 and US$24 billion annually by 2015 and individual cost of US $4,707 over lifetime to reach global targets [2,3], targeted interventions and evidence based prevention programmes have been advocated as cost-effective strategy to combat HIV/AIDS. Such a strategy reduces levels of vulnerability and risk as well as allowing HIV interventions to optimize coverage, reducing costs and lowering the number of new infections [4]. In the United State Virgin Islands, the recommended strategy of universal screening by 14 weeks gestation and screening the infant after birth has a cost savings of $1,122,787 and health benefits of 310 life year gained [5]. A prevention of Mother-to-Child Transmission intervention in Cape town, South Africa, revealed that a program at a scale sufficient to prevent 37 percent of pediatric HIV infections would cost about US$0.34 per person in South Africa and would be affordable to the health care system [6].

In Indian high HIV prevalence southern states, targeted interventions result in significant decline in HIV prevalence among female commercial sex workers (CSWs) and young pregnant women [7]. Evaluation of the cost-effectiveness of the female condom (FC) in preventing HIV infection and other sexually transmitted Diseases (STDs) among CSWs and their clients in the Mpumulanga Province of South Africa, showed that a well-designed FC program oriented to CSWs and other women with casual partners is likely to be highly cost-effective and can save public sector health payer US $12,090 in averted HIV/AIDS treatment costs in rural South Africa [8]

Likewise, analysis of targeting Voluntary HIV Counseling and Testing in Kenya and

in Tanzania showed that increasing the proportion of couples to 70 percent reduces the cost per disability-adjusted life year (DALY) saved to $10.71 in Kenya and $13.39 in Tanzania, and that targeting a population with HIV-1 prevalence of 45 percent decreased the cost per DALY saved to $8.36 in Kenya and $11.74 in Tanzania [9].

However despite growing literature in health and social sciences on factors associated with HIV/AIDS during the last three decades, less is known about the most-at-risk populations regarding HIV prevalence [10, 11, 12, 13,14, 15]. Indeed, whereas in countries with concentrated HIV/AIDS epidemics (Latina America, East Asia and Eastern Europe), the most-at-risk populations including CSWs, long distance truck drivers, men who have sex with men, and unmarried youth [16,11,17] account for a large proportion of new infections, in countries with high prevalence, they account only for a smaller share of new infections [17].

Against this background, this study aims to identify HIV socioeconomic predictors as well as identify the most-at-risk groups among women in Malawi. With HIV prevalence of about 13.6 percent among women of reproductive age [18], HIV/AIDS constitutes a drain on the labor force and government expenditures in Malawi.

Data and Methods

Study setting

The Republic of Malawi is a landlocked country in southeast Africa. Malawi is over 118,000 km^2 with an estimated population of about 16 millions [18]. Its capital is Lilongwe, which is also Malawi's largest city; the second largest is Blantyre and the third is Mzuzu.

Malawi is among the world's least-developed countries. The economy is heavily based on agriculture, with a largely rural population. The country Gross National Income (GNI) per capita at purchasing power parity (PPP) is estimated at $860 while the world average is estimated at $10,780 [18, 19]. Ninety-one percent of Malawians live below 2 dollars (US) per day. The country's Human Development Index is estimated at 0.400, which gives the country a rank of 171 out of 187 countries with comparable data [19].

Malawi has a low life expectancy (53 years) and high infant mortality (66 deaths per

1,000 live births) compared to the world' average (70 years and 41 deaths per 1,000 live births). Averages for sub-Saharan Africa are estimated respectively at 55 years and 80 deaths for 1,000 live births. There is a high prevalence of HIV/AIDS, especially among women with about 13.6 percent HIV positive [18].

Malawi has actively responded to HIV since 1985 after the first AIDS case was reported. In 1988, the government created the National AIDS Control Program to coordinate the country's HIV/AIDS education and prevention efforts. The Public Sector continues to set aside a minimum of two percent of their recurrent budget to support HIV and AIDS programme [20]. The HIV national commission budget has increased from US $98.1 million in 2010 to US $113.51 million in 2011 [20]. According to the Malawi 2012 Global AIDS Response progress report:

- Distribution of leaflets and HIV radio and TV programs. During the 2010-2011 financial year, 1,477 radio and 429 television (TV) programs were produced.
- In 2010 and 2011, around 3.8 million young people (50 percent males and 50 percent females) have been trained on life skills education (LSE) each year.
- Since 2003, the number of condoms distributed per capita has been increasing. Cumulatively, 21,049,592 condoms were distributed in the 2009-2010 fiscal year. During the fiscal year 2010-2011 the annual cumulative total of 26,461,079 condoms were distributed.
- The number of sites providing Prevention of Mother to Child Transmission (PMTCT) services has also been increased from152 facilities in 2006 to 544 sites 2011.
- Antiretroviral Therapy has been provided free of charge in the public sector since 2004. Number of patients alive and on treatment has increased from 10,761 in 2004 to 322,209 in 2011.

Data sources

This study uses data from the 2010 Malawi Health and Demographic Surveys (MDHS). The inclusion of HIV testing in the 2010 MDHS offers the opportunity to identify socioeconomic profile of women age 15-49 living with HIV. Participation in

HIV testing was voluntary. To ensure confidentiality, case numbers (and not names) were used in linking the HIV test results to individual and household characteristics.

A subsample of one-third of the households was selected to conduct HIV testing for eligible women age 15-49 years. Ninety percent of all 2010 MDHS women who were eligible (8,174) for testing were interviewed and consented to HIV tests. The principal mode of HIV transmission in Malawi is heterosexual contact; therefore our analyses focus on 6,395 women who ever had sexual intercourse. Details on the sample design are provided elsewhere [21,22].

Variables

The dependent variable for this analysis is HIV status, characterized as positive or negative blood test. The independent variables include 12 main variables grouped into two major types including, demographic and reproductive behavior variables (age, age at first sex, marital status, age at first birth, number of children ever born, Experience in premarital childbearing, and relationship to the head of household), and socioeconomic and contextual variables (religion, region of residence, place of residence, education, and household wealth index).

The choice of these variables is guided by the literature on factors associated with HIV in sub-Saharan Africa [10-15]. Most-at-risk populations refer to a combination of several factors because socioeconomics factors associated with HIV are not mutually exclusive.

Statistical analyses

Statistical analyses used Pearson Chi-square and Chi-square Automatic Interaction Detector (CHAID) using SPSS version 21. We used weighted data to take into account the complexity of the DHS design. We performed Pearson chi-square to identify associations between the HIV status (positive, negative) and demographic and reproductive behavior variables as well as socioeconomic and contextual variables.

We used CHAID to identify HIV predictors and the most-at-risk groups among women living with HIV [23]. CHAID is a nonparametric technique that makes no distributional assumptions on outliers, collinearities, heteroskedasticity, or distributional error structures. The dependent variable and predictor variables can be nominal (categorical), ordinal (ordered categories ranked from small to large), or

interval (a "scale").

CHAID uses regression and classification algorithms and offers a non-algebraic method for partitioning data that lends itself to graphical displays. The method is a sequential fitting algorithm and its statistical tests are sequential with later effects being dependent upon earlier ones, and not simultaneous as would be the case in a regression model or analysis of variance where all effects are fit simultaneously. CHAID solves the problem of simultaneous inference using Bonferroni multiplier. It automatically tests for and merges pairs of homogenous categories in independent variables.

At each step, CHAID chooses the independent (predictor) variable that has the strongest interaction with the HIV status (dependent variable). The variable having the strongest association with HIV status becomes the first branch in a tree with a leaf for each category that is significantly different relative to be HIV positive. It then assesses the category groupings, or interval breaks to pick the most significant combination of variables. The process is repeated to find the predictor variable on each leaf most significantly related to HIV status, until no significant predictors remain.

The developed model is a classification tree (or data partitioning tree) that shows how major "types" formed from the independent (predictor or splitter) variables differentially predict a criterion or dependent variable. The method permits also identification of women who are likely to be members of a particular group (Segmentation), and assign cases into one of several categories, such as high-, medium-, and low-risk groups (stratification). Selecting a useful subset of predictors from a large set of variables for use in building a formal parametric model (Data reduction and variable screening); Identify relationships that pertain only to specific subgroups and specify these in a formal parametric model (Interaction identification); and recoding group predictor categories and continuous variables with minimal loss of information. Categories of each predictor are merged if they are not significantly different with respect to the dependent variable (Category merging and discretizing continuous variables).

The output of CHAID prediction model is displayed in hierarchical tree-structured form, and consists of several levels of branches: root node, parent nodes, child nodes

and terminal nodes. The root node, "Node 0" or "initial node" is the dependent variable or the target variable, HIV prevalence in our case. Parent node is the upper node compared with nodes on the subsequent (lower) level, whereas the lower level nodes are called child nodes. The terminal node or external node is any node that does not have child nodes. It is the last categories of the CHAID analysis tree.

For each terminal node CHAID provides in a table the following indicators:

1. Node: provides the number and percentage of people belonging to a selected category j (demographic weight in the sample);
2. Gain for each terminal node is the number of women who are living with HIV in absolute. In percentage, gain is calculated as number of women living with HIV in a selected node divided to the total of women living with HIV. Part of the population with the observed characteristic (living with HIV) in a selected category compared to total of women living with HIV.
3. Response can be defined as HIV prevalence among women belonging to each terminal node. Number of women living with HIV in a selected node divided by total of women of the node.
4. Gain index percentage reporting how much greater the proportion of a given target category at each node differs from the overall proportion. It is obtained by dividing the proportion of records that present category j in each terminal node into the proportion of records presenting category j in the total sample. Thus, it represents the increased probability of belonging to the selected category j that contains the records presenting the characteristics defined for each terminal node.

The method allows: (1) identifying complex interactions between variables across the measurement space; (2) Identifying the most significant explanatory variable; (3) Merge categories of nominal variable and categorize continuous variables without loss of information. Furthermore, CHAID as other decision trees can be applied to any data structure.

However, CHAID have two major shortcomings. First the method needs large sample sizes to work effectively because it uses multi-way splits. Indeed, with small sample sizes the respondent groups can quickly become too small for reliable analysis. Last, CHAID does not take into account the hierarchical structure of this data.

Results

Sampling description

Table 1 presents the characteristics of the study population. Since the principal mode of HIV transmission in Malawi is heterosexual contact, our analyses focus on women who ever had sexual intercourse. The distribution of the sample by age shows that more than half (56 percent) of the populations are age less than 30 years old. The average age of the sample is estimated at 29 years old. Women who are in union (i.e., currently married or living with a man) constitute about 77 percent. The proportion of women who have never been married is estimated at 8 percent. Regarding the relationship to the head of household, the majority of women are spouse (63 percent).

Table 1 also shows that the majority of women (more than 80 percent) live in rural areas. By region, the majority of women live in Southern Region and the Central Region. Furthermore, 17 percent of women never attended school, while more than 60 percent have attended only primary school. Regarding the reproductive behavior, a large majority of women had their first sexual intercourse before 20 years (average 16.6 years old).

Table 1 – Description of the sample

Socioeconomic and demographic Characteristics	Percentage		Number[2]
	Weighted[1]	Unweight	
Age			
15-19	11.5	11.6	744
20-24	21.5	20.8	1,327
25-29	22.7	21.9	1,402
30-34	15.2	15.7	1,001
35-39	13.0	12.7	814
40-44	8.6	9.1	579
45-49	7.5	8.3	528
Average	29.6	29.8	-
Age at first sex			
<15	19.1	19.2	1,230
15-19	68.5	68.4	4,376
20-24	11.2	11.1	708
25+	1.3	1.3	81
Average	16.6	16.6	16.6
Marital status			
Single	7.5	7.6	484
In union	77.4	77.1	4,929
No longer in union/Ever married	15.1	15.4	982
Number of ever born children			
0	9.9	9.6	617
1+	90.1	90.4	5,778
Age at first birth			
Never give birth	10.3	10.6	660
< 20 years old	64.8	64.2	4,144
20 +	24.9	25.2	1,591
Ever had premarital child			
No	88.7	88.4	5,652
Yes	11.3	11.6	743
Relationship to the head of household			
Head of household	19.4	19.0	1,213
Spouse	62.6	62.4	3,992
Daughter & Grand daughter	11.0	11.7	748
Others	7.1	6.9	442
Province of residence			
Northern	11.1	17.5	1,122
Central	42.2	34.1	2,181
Southern	46.7	48.4	3,092
Place of residence			
Urban	19.2	13.1	837
Rural	80.8	86.9	5,558
Religion			
Catholic	21.2	20.6	1,316
Protestant	24.3	25.2	1,610
Other Christians	39.7	42.3	2,708
Muslim	13.5	10.9	695
Others	1.3	1.0	66
Education			
None	17.5	16.6	1,060
Primary	63.8	66.4	4,246
Secondary +	18.7	17.0	1,089
Household wealth quintiles			
Poorest	17.6	19.0	1,215
Poorer	20.1	20.6	1,319
Middle	19.7	20.9	1,334
Richer	19.3	20.7	1,323
Richest	23.3	18.8	1,204
Total			6,395

[1] Interpretations are based on weighted percentage.
[2] Unweight crude numbers

HIV prevalence by selected background characteristics

Table 2 describes HIV prevalence in Malawi by women's selected background characteristics. Overall, 14 percent of the women are HIV positive. All independent variables are statistically associated with HIV infection status except for religion.

HIV infection prevalence was high (20 percent) among women aged 30-39 years. Women who are no longer in union (widowed, divorced and separated) had significantly higher prevalence (32 percent) compared to those who have never been in a marital union (single) (8 percent) or those living in union (11 percent). HIV prevalence was high among heads of household (25.0 percent). Furthermore, while 25 percent of women in urban area were HIV positive, the prevalence was less than half (12 percent) compared to their counterparts from the rural areas. The HIV epidemic shows regional heterogeneity with a higher prevalence (20 percent) observed in the Southern region. Women with secondary education had higher HIV prevalence compared to those who never attended school (17 percent versus 14 percent). Regarding the household wealth quintiles the prevalence of HIV infection is higher among the women from the highest quintiles. With reference to sexual and reproductive behavior, HIV prevalence was higher among women who had their first sexual intercourse before the 15th birthday or from their 25th birthday, and /or who have experienced premarital childbearing.

Table 2 – HIV prevalence by selected socioeconomic and demographic characteristics

Socioeconomic and demographic Characteristics	HIV+ Prevalence (%)	Total (N)	Chi-Square	p-value
Age				
15-19	5.0	744		
20-24	6.9	1,327		
25-29	12.4	1,402		
30-34	19.8	1,001	190.35	<0.001
35-39	21.4	814		
40-44	18.7	579		
45-49	16.9	528		
Age at first sex				
<15	15.9	1,230		
15-19	13.1	4,376	9.13	0.028
20-24	12.4	708		
25&+	18.5	81		
Marital status				
Single/ Never married	7.9	484		
In union	10.7	4,929	316.15	<0.001
Ever married/ No longer in union	31.5	982		
Number of ever born children				
0	7.9	617	18.80	<0.001
1&+	14.2	5,778		
Age at first birth				
Never give birth	9.1	660		
< 20 years old	14.1	4,144	12.96	0.002
20 +	14.3	1,591		
Ever experience premarital childbearing				
No	12.8	5,652	26.99	<0.001
Yes	19.8	743		
Relationship to the head of household				
Head of household	25.0	1,213		
Spouse	10.2	3,992	179.93	<0.001
Daughter & Grand daughter	11.9	748		
Others	17.0	442		
Region of residence				
Northern	10.0	1,494		
Central	9.5	3,062	184.90	<0.001
Southern	20.0	4,444		
Place of residence				
Urban	24.7	1,156	157.00	<0.001
Rural	12.3	7,844		
Religion				
Catholic	12.6	1,316		
Protestant	14.3	1,610		
Other Christians	13.4	2,708	2.66	0.616
Muslim	14.8	695		
Others	13.6	66		
Education				
None	13.9	1,060		
Primary	12.8	4,246	10.73	0.005
Secondary +	16.6	1,089		
Household wealth quintiles				
Poorest	10.3	1,215		
Poorer	10.9	1,319		
Middle	11.6	1,334	88.34	<0.001
Richer	14.3	1,323		
Richest	21.5	1,204		
Total	13.6	6,395		

HIV predictors in Malawi: results from CHAID analysis

Table 3 shows summary information on the specifications used to build the CHAID model and the resulting model. Twelve independent variables were specified, but only six were included in the final model. The variables such as age at first sex, age at first birth and female education did not make a significant contribution to the model, so they were automatically dropped from the final model. Overall, there are 21 nodes among which 13 terminal nodes. Parent nodes include at least 100 cases whereas child nodes account for 50 cases in minimum.

Model components	Model specification	Results
Dependent variable	HIV status	HIV+=13.6%
Independent Variables	Age, Age at first sex, Marital status, Ever had a child, age at first birth, Experience premarital childbearing, Relationship to the head of household, Region of residence, Place of residence, Education, Wealth Index, Religion	**Marital status**, Age, Wealth Index, Relationship to the head of household, Region of residence, Place of residence
Maximum Tree Depth	3	3
Minimum Cases in Parent Node	100	100
Minimum Cases in Child Node	50	50
Number of Nodes	-	21
Number of Terminal Nodes	-	13
Overall predicted correct percentage		86.8

The tree diagram depicted in Figure 1 shows that "Marital status" is the best predictor of HIV status among women in Malawi (Chi-square = 313.22, p-value<0.0001).

The tree is split into two main notes. Node 1 includes women formerly in union; and Node 2 is composed of women in union and never married women.

Node 1- Women formerly in union. For this group, including divorced, widowed, and not living together, age is the best predictor of HIV prevalence (Chi-square=56.30, p-value <0.001). The group is further splits into four sub age groups: 15-24, 25-29, 30-34 and 40-49, and 35-39.

Among women aged 30-34, 40-44 and 40-49 (Node 3) with HIV prevalence of 37 percent, household wealth quintiles are the best predictor of HIV infection (Chi-square=29.81, p-value<0.001). Indeed in this group women in the highest wealth

quintile (Node 9) have HIV prevalence about three times higher than their counterpartners from the lowest wealth quintile-Node 11 (60 percent versus 22 percent). For women in the age groups of 14-19 and 20-24 years (Node 4) with a HIV prevalence of 13 percent, the relationship to the head of household is the best predictor of HIV infection (Chi-square=11.1, p-value <0.003) whereby women head of household (Node 12) have higher HIV prevalence compared to other women with different relationship to the head of household – Node 13 (24 percent versus 7 percent). The region of residence is the best predictor of HIV infection among women aged 35-39 (Chi-square=11.5, p-value<0.002) with women living in the Southern region (Node 14) having HIV prevalence about twice the one of the women from the Central and Northern regions-Node 15 (59 percent versus 31 percent). Among women aged 25-29 years (Node 6) accounting for 3 percent of the study population with HIV prevalence of 27 percent, age remains the only significant and final predictor of HIV prevalence.

Node 2- This group includes women in union (married or living together) and those have **never been in union**, representing 85 percent of the study population and have HIV prevalence of 10 percent. Place of residence (rural or urban) is the best predictor of HIV infection with a higher prevalence in urban area (Node 8) compared to rural area – Node 7 (21 percent vs 9 percent, Chi-square = 89.8, p-value<0.001).

For the women living in the rural area (Node 7) and representing 74 percent of the population, the best predictor for HIV infection is age (Chi-square=86.0, p-value <0.001) with the highest prevalence among women aged 30-44 years (13 percent) followed by the age group 25-49 (Node 18: 9 percent) and the age group 15-24 (Node 18: 4 percent). Similarly, age is strong predictor of HIV infection (Chi-square=86.0, p-value <0.001) among women living in urban area (Node 8) whereby the age group 30-49 (Node 19) has a prevalence about twice the one among the age group 15-29-Node 20 (29 percent versus 15 percent).

HIV risk groups in Malawi

There are in total 13 homogenous subgroups or terminal nodes. Table 4 describes these 13 subgroups (terminal nodes) in terms of their composition, demographic weight in the sample (columns 1 and 2), their share in HIV burden (columns 3 and 4)

and their corresponding HIV prevalence (column 5). The 13 homogenous sub-groups could be grouped into 5 major groups.

Group 1 represents 3 percent of the sample with an overall HIV prevalence of 59 percent. This group accounts for 11 percent of all the women HIV positive. Group 1 includes two subgroups: a) women in union disruption living in richest household and age 30-34 or 40-49 years old; and b) women in union disruption living in the Southern region and age 35-39 years old.

Group 2 represents 5 percent of the sample with an overall HIV prevalence of 35 percent and accounts for 12 percent of all HIV positive women. This group is composed of two subgroups including women in union disruption living in intermediate wealth households (non-poorest and non-richest households) age 30-34 or 40-49; and women in union disruption age 35-39 and living in the Northern or Central region.

Group 3 represents about 10 percent of study population with an overall HIV prevalence of 27 percent and accounts for 20 percent of all HIV positive women. This group is divided into four subgroups: a) Never married and women in Union, living in urban area and age 30-49; b) Formerly in union (widowed or divorced) women age 25-29; c) Young women (15-24) formerly in union who are head of household; d) Formerly in union women living in poorest household and age 15-24, 30-34 or 40.

Group 4 represents about 33 percent of the study population with an overall HIV prevalence of 14 percent and account for 33 percent of all the HIV positive women. This group includes adolescent (15-19), never married women or in union living in urban area; and never married or women in union living in rural area age 30-44.

Group 5 represents 50 percent of the study population and has the lowest HIV prevalence of 7 percent, but account for 23 percent of all the HIV positive women. This group includes three subgroups: a) Never married or women in union living in rural area and age 25-29 or 45-49; b) Young women age 15-24 who are non longer in union and are not head of household; and c) Young women (15-24) who are never married or in Union and are living in rural areas.

Table 4 reports also the gain index percentage (column 6) expressing how much greater the proportion of a given target group at each node differs from the overall proportion. The index percentage is very high among women belonging to group with

high HIV prevalence but with small demographic weight in the population (categories 1 to 3). Opposite values are observed among groups accounting for the major part of the sample among which HIV prevalence is low (Group 5). The Index is equal to 100 in category 4.

Table 4 – CHAID Gains for Nodes

Group/Node No	Node description	Node		Gain		% HIV[5]	Index[6]
		N[1]	%[2]	N[3]	%[4]		
Group 1		165	2.6	98	11.2	59.4	435.6
9	Formerly in union, 30-34/ 40-49 and richest	82	1.3	49	5.6	59.8	438.2
14	Formerly in union, 35-39 and Southern region	83	1.3	49	5.6	59.0	433.0
Group 2		308	4.8	108	12.4	35.1	257.2
10	Formerly in union, 30-34/40-49 and Rich/Middle/Poor	246	3.8	89	10.2	36.2	265.3
15	Formerly in union, 35-39 and North/Central regions	62	1.0	19	2.2	30.6	224.7
Group 3		640	10	171	19.5	26.7	195.9
19	Never married/in Union, living in urban area, 30-49	268	4.2	78	8.9	29.1	213.4
6	Formerly in union and 25-29	180	2.8	49	5.6	27.2	199.6
12	Formerly in union, 15-24 and Head of household	67	1.0	16	1.8	23.9	175.1
11	Formerly in union, 15-24, 30-34/40-49 and Poorest	125	2.0	28	3.2	22.4	164.3
Group 4		2,117	33.1	291	33.3	13.8	100.8
20	Never married/in Union, living in urban area, 15-19	417	6.5	64	7.3	15.3	112.6
16	Never married/in Union, living in rural area, 30-44	1,700	26.6	227	26.0	13.4	97.9
Group 5		3,165	49.5	204	23.3	6.5	42.3
18	Never married/in Union, rural area, 25-29 and 45-49	1,404	22.0	126	14.4	9.0	65.8
13	Formerly in union, 15-24 and Not Head of household	137	2.1	10	1.1	7.3	53.5
17	Never married/in Union, living in rural area, 15-24	1,624	25.4	68	7.8	4.2	30.7
Total		6,395	100	872	100	13.6	-

Notes:;[1] Number of cases per node (demographic size in the sample);
[2] Demographic size in percentage = (.1/Σ.1)*100;[3] Number of HIV women;[4] Demographic size among HIV positive women in percentage = (.3/Σ.3)*100; [5] HIV prevalence in percentage = (.3/Σ.1)*100; [6] Node Index = ((.3/Σ3)/(.1/Σ.1))*100

Discussions

This paper aimed to describe and profile HIV prevalence among women in Malawi. The study used Chi-square and CHAID techniques to analyze data from the Malawi 2010 DHS.

Analyses suggested three keys findings. First, consistent with previous studies [13, 24], findings from bivariate analysis and chi-square test showed high HIV prevalence among women in union dissolution, among the most educated women, women living in wealthy households and/or among women living in urban areas. The finding confirmed also region heterogeneity in HIV prevalence. The Southern region being the most affected. In general the most educated women are more likely to marry husbands with high education level, and belonging to high socio-economic class of the society [25,26]. In parallel, relatively rich and better-educated men have higher rates of partner change because they have greater personal autonomy and spatial mobility [27, 28, 29, 30]. Women's economic dependence on their partners may also make it difficult for them to insist on safer sex (e.g. condom use). Concentration (about 50 percent) of the most educated, richest and urban women in the Southern region may explain high HIV prevalence in that region.

Second, results from CHAID models reported that marital status is the best predictor of HIV status among women in Malawi. Non-poorest women who are no longer in union (widowed and divorced or separated) age 30-34 or 40-49 have significantly higher HIV prevalence. This may be because: (1) husband from the highest quintile or a male partner may have more access to transactional sex and other risk behaviors such as polygamy which may increase women's vulnerability to HIV; (2) wealthier HIV positive widowed may have better quality of life as well as better access to treatment and survive longer [31]. Furthermore, divorced and separated are more frequent among the most educated women with economic autonomy [32]. Their causes (polygyny and/or infidelity) as well as consequences (multiple sexual partnerships) are also factors associated with HIV prevalence [33,34].

Third and last, CHAID models depicted also different interactions between risk factors and profiled HIV risk groups in Malawi. For instance, whilst overall HIV prevalence among women living in urban areas (25 percent) is twice the prevalence

observed among women living in rural areas (12 percent), HIV prevalence is estimated at 15 percent among never married or women in union living in urban areas age 15-29, and at 13 percent among never married or women in union living in the rural areas age 30-44. Likewise, whereas in general HIV prevalence is low among never married and women in union (10 percent), CHAID results revealed a higher HIV prevalence (29 percent) among never married and women in union age 30-49 who live in urban areas compared to: (1) women in union disruption age 15-24 (7 percent if they are not head of households and 23 percent for head of household); (2) women in union disruption age 25-29 (27 percent); and (3) women in union dissolution age 30-34 and 40-49 who live in poorest households (22 percent).

These findings showed the complexity of different interactions that may present challenges to conventional regression models. Indeed, CHAID is a sequential fitting algorithm and its statistical tests are sequential with later effects being dependent upon earlier ones and not simultaneous as would be the case in a regression model or analysis of variance where all effects are fit simultaneously. Furthermore, CHAID allows automatic detection of interaction between variables.

In the light of these findings, it is noteworthy that to reduce number of new infection, interventions should be targeted and prioritized according to the prevalence and demographic size of different risk groups. Furthermore, policy makers' prioritization of interventions may depend also on preference for preventive interventions compared to treatment of and care for HIV infected people and/or to treatment of and care for AIDS-patients. In Thailand, for instance policy makers expressed a preference for target preventive interventions that are highly effective compared to care and treatment [35].

Regarding preventive interventions, the findings suggested that:

1. Couples (males and women in union) and never married people age 25-49 (nodes 16 and 18) living in rural areas should be the first targets using universal HIV testing, "Abstinence", "Be faithful" and "use condom" campaign. Indeed, this group includes 49 percent of the study population, among whom the HIV prevalence is estimated at 11 percent on average. About 40 percent of women living with HIV in Malawi belong to this category.
2. Young age 15-24 living in rural areas (node 17) and urban adolescent 15-19

(node 20) is the second most important target. This group account for 32 percent of the studied population and 15 percent of women living with HIV. Besides, majority of adults living with HIV may be infected during adolescence. Unfortunately, the available dataset could not provide information on time of infection.

3. The country develops and implements a social policy to protect single mothers. Indeed, though overall HIV prevalence is estimated at 6 percent on average among young women age 15-24 (Table 2) that prevalence is estimated above 20 percent among young women formerly in union and among young women who ever experience premarital childbearing living in urban areas. Likewise, HIV prevalence is very high among women in union disruption (32 percent on average) compared to other groups (10 percent). Though this high prevalence may be due to male mortality, some women in union disruption may be vulnerable because of poverty.

With reference to treatment and care, higher priority must be given to promoting HIV test, monitoring and evaluation of equity in access to treatment among women in union disruption and never married or women in union age 30-49 living in urban areas. Indeed, formerly in union women represent only about 13 percent of women of reproductive age in Malawi, they have the higher HIV prevalence ranges between 22 percent observed among poorest and 60 percent among richest.

Nevertheless, to achieve zero new infection as part of MDG 6, there is need of more comprehensive policy to combat HIV because of the complexity of HIV socioeconomic profile in Malawi. There are several groups built from several socioeconomic categories depending on individual marital status, wealth index, age, place of residence, and relationship to the head of household. In South Africa, Bendavid et al. [36] revealed that scaling up all aspects of HIV care including universal testing and treatment was associated with a life expectancy gain of 22.2 months, and new infections were 73 percent lower.

From the methodological point of view, this study has some limitations, which do not detract from its scientific importance and contribution. First, this study used cross-sectional data from the Demographic and Health Surveys, which does not permit one to draw causal association between HIV status and the associated factors. For

instance, whether HIV infection has occurred before, during or after the union. Last, CHAID model ignores the hierarchical structure of the Demographic and Health Survey data and need large sample size.

In conclusion, this study recommends: (1) design and implementation of targeted interventions taking into account HIV prevalence and the demographic size of different groups at risk; (2) reinforcement of integration of family planning and HIV/AIDS services because the population understudied includes women of reproductive health. Integrating the two services (HIV and FP) could be cost-effective; (3) Community health workers, households based campaign, reproductive health services and reproductive health courses at school could be used as canon to achieve universal prevention strategy, testing, counseling and treatment.

Acknowledgements*: This research was supported by the Luxembourg National Research Fund (FNR). The authors thank Macro international for providing free the Malawi 2010 DHS data-sets including HIV data. We thank Prof. Guillaume Wunsch for his comments. Furthermore, we* have appropriate approval and authorization from the DHS datasets research ethics board for conducting HIV secondary data analysis.

Contributors: JBOE participated in conception and design, literature review, data analysis and interpretation, drafting the article, critical revisions for important intellectual content and approval of final article for submission. MK, NJM, EMZ and YY participated in conception and design, interpretation of results, critical revisions for important intellectual content and approval of final article for submission.

Data sharing statement: This study is based on the Malawi 2010 Demographic and Health Surveys. These data are available on www.measuredhs.com. Access to individual HIV status as well as individual background information required authorization from MACRO ORC (www.measuredhs.com). The Website depicts the process.

References

[1] United Nations.*The Millennium Development Goals. 2012 Report* . New York: United Nations, 2012.

[2] UNAIDS *UNAIDS Report on the Global AIDS Epidemic 2012. Global report.* Geneva: UNAIDS, 2012. http://www.unaids.org/en/resources/campaigns/20121120_globalreport2012/globalreport/

[3] International HIV / AIDS Alliance. (2010). *The cost efficiency of HIV prevention for vulnerable and most-at-risk populations and the reality of funding. What's Preventing Prevention Campaign Briefing 2.* Hove: International HIV / AIDS Alliance.

[4] UAIDS. *World AIDS day report 2011. How to get to zero: Faster, smarter, Better.* Geneva: UNAIDS, 2011.

[5] Udeh, B.; Udeh, C. & Graves, N., Perinatal HIV transmission and the cost-effectiveness of screening at 14 weeks gestation, at the onset of labour and the rapid testing of infants, *BMC Infectious Diseases* 2008, 8:174 doi:10.1186/1471-2334-8-174. http://www.biomedcentral.com/1471-2334/8/174

[6] Roux, P.; Henley, L.; Cotton, M.; Eley, B., Burden and cost of inpatient care for HIV-positive paediatric patients--status in the Cape Town metropole during the second week of March 1999. Paediatric HIV Census Group. *South African Medical Journal*, 2000, 90(10):1008-1011

[7] Kumar, R.; Mehendale, S.M.; Panda, S.; Venkatesh, S.; Lakshmi, PVM.; Kaur, M.; Prinja, S.; Singh, T.; Virdi, N.K.; Bahuguna, P.; Sharma, A.K.; Singh, S.; Godbole, S.V.; Risbud, A.; Manna, B.; Thirumugal, V.; Roy, T.; Sogarwal, R.; Pawar, N.D. Impact of targeted interventions on heterosexual transmission of HIV in India, *BMC Public Health*, 2011, 11:549, http://www.biomedcentral.com/1471-2458/11/549

[8] Marseille, E.; Kahn, J.G.; Billinghurst, K.; Saba, J. Cost-effectiveness of the female condom in preventing HIV and STDs in commercial sex workers in rural South Africa. *Social Sciences & Medicine*. 2001;52(1):135-148.

[9] Sweat, M; Gregorich, S; Sangiwa, G.; Furlonge, C.; Balmer, D.; Kamenga, C.; Grinstead, O. & Coates, T., Cost-effectiveness of voluntary HIV-1 counselling and

testing in reducing sexual transmission of HIV-1 in Kenya and Tanzania. *Lancet*. 2000 Jul;356 (9224):113-121

[10] Durevall, D, & Lindskog, A. *HIV/AIDS, Adult Mortality and Fertility: Evidence from Malawi. Working Papers in Ecomics, No 284*. Göteborg: School of Business, Economics and Law, University ofGothenburg, 2007.

[11] Rombo, D. *Marital risk factors and HIV infection among women: A comparison between Ghana and Kenya,* Ph.D. Dissertation. Minneapolis: University of Minnesota, 2009.

[12] Öjteg, K. *Socio-Economic determinants of HIV in Zambia. A district level Analysis*. Ph.Dissertation. Lund: Department of Economics at the University of Lund, 2009.

[13] Magadi, M., & Desta , M. A multilevel analysis of the determinants and cross-national variations of HIV seropositivity in sub-Saharan Africa: Evidence from the DHS. *Health & Place*, 2011; 17 (5): 1067–83.

[14] Magadi, M. Understanding the gender disparity in HIV infection across countries in sub-Saharan Africa: evidence from the Demographic and Health Surveys. *Sociology of Health & Illness,* 2011; 33 (4): 522-39.

[15] Asiedu, C., Asiedu, E., & Owusu, F. The Socio-Economic Determinants of HIV/AIDS Infection Rates in Lesotho, Malawi, Swaziland and Zimbabwe. *Development Policy Review ,* 2012; 30 (3): 305-326 .

[16] Green, E. *Rethinking AIDS Prevention. Learning from Successes in Developing Countries* . Westport, CT: Praeger publishers, 2004

[17] International HIV / AIDS Alliance. *The cost efficiency of HIV prevention for vulnerable and most-at-risk populations and the reality of funding. What's Preventing Prevention Campaign Briefing 2*. Hove: International HIV / AIDS Alliance, 2010.

[18] Population Reference Bureau, *2012 World Population Data sheet*, Washington DC, PRB, 2012.

[19] United Nations Development Programme/UNDP. Sustainability and Equity: A better future for all, Human Development Report 2011. New York, UNDP, 2011.

[20] Malawi, HIV national Commission. National 2012 Global AIDS Response Progress

Report: Malawi Country Report. Malawi, HIV National Commission, 2012, 67 p.

[21] Malawi National Statistical Office (NSO); ORC Macro. *Malawi Demographic and Health Survey 2004.* Calverton, Maryland: NSO and ORC Macro, 2005.

[22] Malawi National Statistical Office (NSO); ICF Macro. *Malawi Demographic and Health Survey 2010.* Zomba, Malawi, and Calverton, Maryland, USA: NSO and ICF Macro, 2011.

[23] Kass, G. V. An Exploratory Technique for Investigating Large Quantitaties of Categorical Data. *Applied Statistics,* 1980; 29 (2): 119-127.

[24] Adair, T. *HIV Status and Age at First Marriage among Women in Cameroon.* DHS Working Paper No 33, Calverton, Maryland, USA: Macro International Inc, 2007.

[25] Cleland, J., & Van Ginneken, J. Maternal education and child survival in developing countries: The search for pathways of influence. *Social Sciences and Medicine* , 1988 ; 27(12):1357-1368.

[26] Barrett, H., & Brown, A. Health, hygiene, and maternal education: Evidence from Gambia. *Social Sciences and Medicine*, 1996; 43(11): 1579-1590.

[27] Ainsworth, M. & Semali, I. Who is most likely to die of AIDS? Socioeconomic correlates of adult deaths in Kagera Region, Tanzania. In: *Confronting AIDS: evidence for developing world*. Edited by Ainsworth M, Fransen L, Over M. Washington, D.C.: The World Bank; 1998: 95-110.

[28] Gregson S, Garnett GP, Nyamukapa CA, Hallett TB, Lewis JJ, Mason PR, et al. HIV decline associated with behaviour change in eastern Zimbabwe. *Science,* 2006; 311:664–666.

[29] Hargreaves JR, Glynn JR. Educational attainment and HIV-1 infection in developing countries: a systematic review. *Tropical Medical and International Health*, 2002; 7:489–498.

[30] Gillespie S., Kadiyala, S. & Greener. Is poverty or wealth driving HIV transmission? *AIDS*, 2007; 21 (suppl 7):S5–S16.

[31] Rodrigo, C., & Rajapakse, S. (2010). HIV, poverty and women. *International Health*, 2010, 2(1): 9-16.

[32] Takyi, B., & Broughton, C. Marital Stability in Sub-Saharan Africa: Do Women's Autonomy and Socioeconomic Situation Matter? *Journal of Family and Economic Issues* , 2006; 27 (1): 113-132.

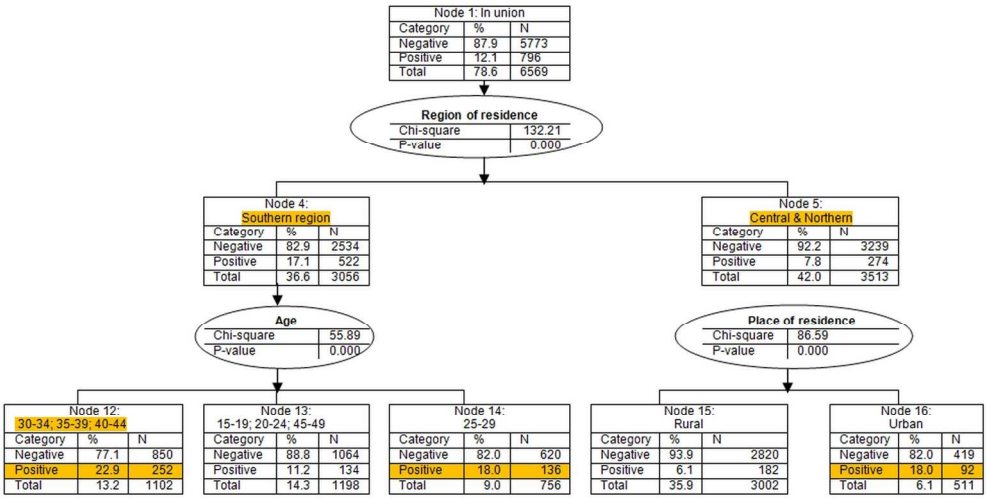
[33] Reniers, G. Divorce and remarriage in rural Malawi. *Demographic Research*, 2003; Special Collection 1, article 6: 175-206. www.demographic-research.org/special/1/6/S1-6.pdf.

[34] Reniers, G. Marital strategies for regulating exposure to HIV. *Demography*, 2008; 45 (2): 417-438.

[35] Youngkong, S.; Baltussen, R.; Tantivess, S.; Koolman, X.; & Teerawattananon, Y. Criteria for priority setting of HIV/AIDS interventions in Thailand: a discrete choice experiment. *BMC Health Services Research* 2010, 10:197. http://www.biomedcentral.com/1472-6963/10/197 retrieved on March 1, 2013

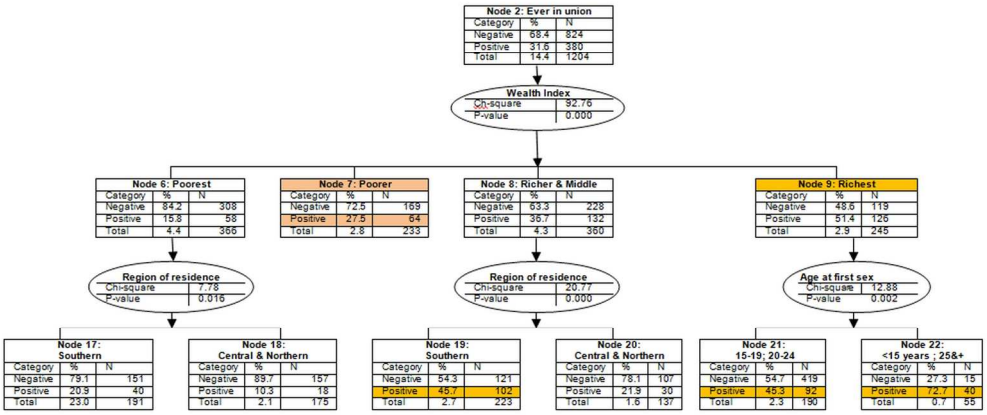
[36] Bendavid, E; Brandeau, M. L.; Wood, R.; Owens, D. K. Comparative Effectiveness of HIV Testing and Treatment in Highly Endemic Regions. *Archives of Internal Medicine, 2010;170 (15):1347-1354*.

Figure 1a- HIV prevalence in Malawi: tree diagram for women in union



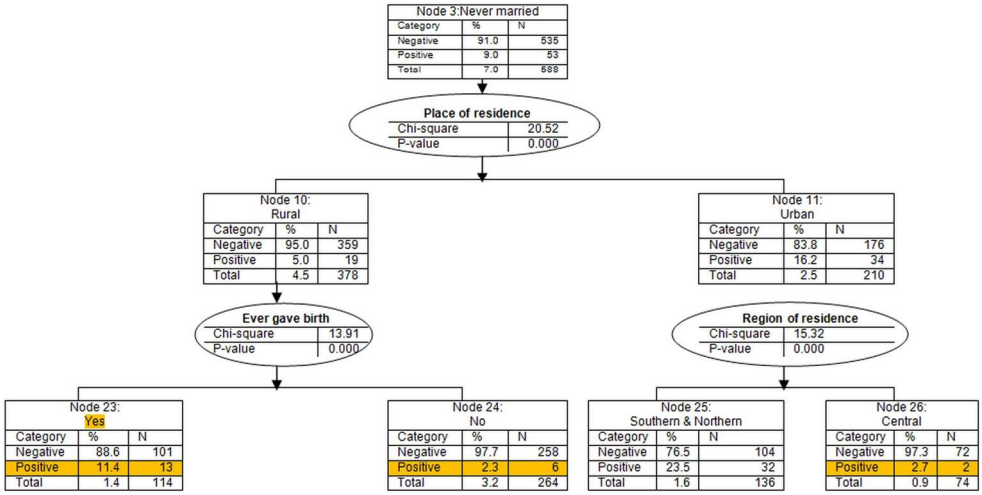
165x90mm (300 x 300 DPI)

Figure 1b - HIV prevalence in Malawi: tree diagram for women in union disruption



196x90mm (300 x 300 DPI)

Figure 1c - HIV prevalence in Malawi: tree diagram for never married women



167x90mm (300 x 300 DPI)