

# Identifying HIV most-at-risk groups in Malawi for targeted interventions. A classification tree model

Jacques B O Emina,<sup>1,2,3</sup> Nyovani Madise,<sup>4</sup> Mathias Kuepie,<sup>2</sup> Eliya M Zulu,<sup>5</sup> Yazoume Ye<sup>6</sup>

**To cite:** Emina JBO, Madise N, Kuepie M, *et al*. Identifying HIV most-at-risk groups in Malawi for targeted interventions. A classification tree model. *BMJ Open* 2013;**3**:e002459. doi:10.1136/bmjopen-2012-002459

► Prepublication history for this paper are available online. To view these files please visit the journal online (<http://dx.doi.org/10.1136/bmjopen-2012-002459>).

Received 6 December 2012  
Revised 8 April 2013  
Accepted 9 April 2013

This final article is available for use under the terms of the Creative Commons Attribution Non-Commercial 2.0 Licence; see <http://bmjopen.bmj.com>

For numbered affiliations see end of article.

## Correspondence to

Dr Jacques B O Emina;  
[jacques.emina@gmail.com](mailto:jacques.emina@gmail.com)

## ABSTRACT

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

**Design:** A cross-sectional survey.

**Setting:** Malawi

**Participants:** The study used a sample of 6395 women aged 15–49 years from the 2010 Malawi Health and Demographic Surveys.

**Interventions:** N/A

**Primary and secondary outcome measures:**

Individual HIV status: positive or not.

**Results:** Findings from the Pearson  $\chi^2$  and  $\chi^2$  Automatic Interaction Detector analyses revealed that marital status is the most significant predictor of HIV. Women who are no longer in union and living in the highest wealth quintiles households 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 areas.

**Conclusions:** In the light of these findings, this study recommends: (1) that the design and implementation of targeted interventions should consider the magnitude of HIV prevalence and demographic size of most-at-risk groups. Preventive interventions should prioritise couples and never married people aged 25–49 years and living in rural areas because this group accounts for 49% of the study population and 40% 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 aged 30–49 years and living in urban areas; (3) community health workers, households-based campaign, reproductive-health services and reproductive-health courses at school could be used as canons to achieve universal prevention strategy, testing, counselling and treatment.

## 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

## ARTICLE SUMMARY

### Article focus

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

### 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% of women are HIV positive.
- Our findings revealed that the richest and formerly in union women constitute the most-at-risk group.
- To achieve zero new infection as part of the Millennium Development Goal 6, there is a need for a more comprehensive policy to combat HIV because of the complexity of the HIV-socioeconomic profile in Malawi. There are several groups built from several socioeconomic categories depending on the individual marital status, wealth index, age, place of residence and relationship to the head of the household.

declining, HIV/AIDS has remained the leading cause of death in women of reproductive age in low-income and middle-income countries, particularly in sub-Saharan Africa.<sup>1</sup> 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 2 years remaining to the end-date of the Millennium Development Goals (MDG) target, HIV/AIDS remains a long-term global challenge.<sup>1</sup>

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 the individual cost of US\$4707

## ARTICLE SUMMARY

## Strengths and limitations of this study

- To 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  $\chi^2$  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 themselves to graphical displays.
- This study has two major limitations. First, it used cross-sectional data from the Demographic and Health Surveys, which do not permit one to draw causal associations between HIV status and the associated factors. For instance, whether HIV infection has occurred before, during or after the union. Last, the CHAID model ignores the hierarchical structure of the Demographic and Health Survey data and needs a large sample size.

over a lifetime to reach global targets,<sup>2, 3</sup> targeted interventions and evidence-based prevention programmes have been advocated as a cost-effective strategy to combat HIV/AIDS. Such a strategy reduces the levels of vulnerability and risk as well as allowing HIV interventions to optimise coverage, reduce costs and lower the number of new infections.<sup>4</sup> In the United States 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.<sup>5</sup> A prevention of Mother-to-Child Transmission intervention in Cape Town, South Africa, revealed that a programme at a scale sufficient to prevent 37% of paediatric HIV infections would cost about US\$0.34/person in South Africa and would be affordable to the healthcare system.<sup>6</sup>

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

Likewise, an analysis of targeting Voluntary HIV Counseling and Testing in Kenya and in Tanzania showed that increasing the proportion of couples to 70% 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 an HIV-1

prevalence of 45% decreased the cost per DALY saved to \$8.36 in Kenya and \$11.74 in Tanzania.<sup>9</sup>

However, despite the 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.<sup>10–15</sup> Indeed, whereas in countries with concentrated HIV/AIDS epidemics (Latin 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<sup>16, 4, 3</sup> account for a large proportion of new infections, in countries with high prevalence, they account only for a smaller share of new infections.<sup>10</sup>

Against this background, this study aims to identify HIV-socioeconomic predictors as well the most-at-risk groups among women in Malawi. With an HIV prevalence of about 13.6% among women of reproductive age,<sup>17</sup> HIV/AIDS constitutes a drain on the labour 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<sup>2</sup> with an estimated population of about 16 million.<sup>17</sup> 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's Gross National Income per capita at purchasing power parity is estimated at \$860 while the world average is estimated at \$10 780.<sup>17, 18</sup> Ninety-one per cent of Malawians live on an income of below \$2 (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.<sup>18</sup>

Malawi has a low life expectancy (53 years) and high infant mortality (66 deaths/1000 live-births) compared with the world average (70 years and 41 deaths/1000 live-births). Averages for sub-Saharan Africa are estimated at 55 years and 80 deaths, respectively, for 1000 live-births. There is a high prevalence of HIV/AIDS, especially among women with about 13.6% HIV positive.<sup>17</sup>

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 2% of their recurrent budget to support the HIV and AIDS programme.<sup>19</sup> The HIV national commission budget has increased from US\$98.1 million in 2010 to US\$113.51 million in 2011.<sup>19</sup> According to the Malawi 2012 Global AIDS Response progress report:

- ▶ Distribution of leaflets and HIV radio and TV programmes. During the 2010–2011 financial year, 1477 radio and 429 television (TV) programmes were produced.

- ▶ In 2010 and 2011, around 3.8 million young people (50% men and 50% women) have been trained on life skills education 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, a cumulative total of 26 461 079 condoms were distributed.
- ▶ The number of sites providing Prevention of Mother to Child Transmission services has also been increased from 152 facilities in 2006 to 544 sites in 2011.
- ▶ Antiretroviral therapy has been provided free of charge in the public sector since 2004. The 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 the socioeconomic profile of women aged 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 aged 15–49 years. Ninety per cent of all 2010 MDHS women who were eligible (8174) 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 6395 women who ever had sexual intercourse. Details on the sample design are provided elsewhere.<sup>20 21</sup>

### Variables

The dependent variable for this analysis is HIV status, characterised as a positive or negative blood test. The independent variables include 12 main variables grouped into two major types including: demographic and reproductive behaviour 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 the 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.<sup>3–8</sup> Most-at-risk populations refer to a combination of several factors because socioeconomic factors associated with HIV are not mutually exclusive.

### Statistical analyses

Statistical analyses used Pearson's  $\chi^2$  and  $\chi^2$  Automatic Interaction Detector (CHAID) using SPSS V.21. We used weighted data to take into account the complexity of the DHS design. We performed Pearson's  $\chi^2$  to identify

associations between the HIV status (positive or negative) and demographic and reproductive behaviour variables as well as the socioeconomic and contextual variables.

We used CHAID to identify HIV predictors and the most-at-risk groups among women living with HIV.<sup>22</sup> CHAID is a non-parametric technique that makes no distributional assumptions on outliers, collinearities, heteroscedasticity 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 on earlier ones, and not simultaneous as would be the case in a regression model or analysis of variance where all effects are fitted simultaneously. CHAID solves the problem of simultaneous inference using a Bonferroni multiplier. It automatically tests for and merges pairs of homogeneous 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 differently 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 also permits the identification of women who are likely to be members of a particular group (Segmentation), and assigns cases into one of several categories, such as high-risk, medium-risk 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); Identifying relationships that pertain only to specific subgroups and specifying 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 discretising continuous variables).

The output of the CHAID prediction model is displayed in a 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. The 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. They are the last categories of the CHAID analysis tree.

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

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 the number of women living with HIV in a selected node divided by the total of women living with HIV.
3. Response can be defined as HIV prevalence among women belonging to each terminal node. The number of women living with HIV in a selected node is divided by the total of women in 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 by the proportion of records presenting category  $j$  in the total sample.

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

However, CHAID has two major shortcomings. First, the method needs large sample sizes to work effectively because it uses multiway 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%) of the population is aged less than 30 years. The average age of the sample is estimated at 29 years. Women who are in union (ie, currently married or living with a man) constitute about 77%. The proportion of women who have never been married is estimated at 8%. Regarding the relationship to the head of household, the majority of women are spouses (63%).

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

### HIV prevalence by selected background characteristics

Table 2 describes HIV prevalence in Malawi by women's selected background characteristics. Overall, 14% 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%) among women aged 30–39 years. Women who are no longer in union (widowed, divorced and separated) had a significantly higher prevalence (32%) compared with those who had never been in a marital union (single; 8%) or those living in union (11%). HIV prevalence was high among the heads of households (25%). Furthermore, while 25% of women in urban areas were HIV positive, the prevalence was less than half (12%) compared with their counterparts from the rural areas. The HIV epidemic shows regional heterogeneity with a higher prevalence (20%) observed in the Southern region. Women with secondary education had a higher HIV prevalence compared with those who never attended school (17% vs 14%). 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 behaviour, HIV prevalence was higher among women who had their first sexual intercourse before their 15th birthday or from their 25th birthday, and/or who had experienced premarital childbearing.

### HIV predictors in Malawi: results from the CHAID analysis

The tree diagram depicted in figure 1 shows that 'Marital status' is the best predictor of HIV status among women in Malawi ( $\chi^2=313.22$ ,  $p<0.0001$ ).

The tree is split into two main nodes. 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 ( $\chi^2=56.30$ ,  $p<0.001$ ). The group is further split into four subage groups: 15–24, 25–29, 30–34 and 40–49, and 35–39.

Among women aged 30–34, 40–44 and 45–49 (node 3) with an HIV prevalence of 37%, household wealth quintiles are the best predictor of HIV infection ( $\chi^2=29.81$ ,  $p<0.001$ ). Indeed, in this group, women in the highest wealth quintile (node 9) have an HIV prevalence about three times higher than their counterparts from the

**Table 1** Description of the sample

Socioeconomic and demographic Characteristics	Per cent		Number†
	Weighted*	Unweighted	
<b>Age</b>			
15–19	11.5	11.6	744
20–24	21.5	20.8	1327
25–29	22.7	21.9	1402
30–34	15.2	15.7	1001
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	–
<b>Age at first sex</b>			
<15	19.1	19.2	1230
15–19	68.5	68.4	4376
20–24	11.2	11.1	708
25+	1.3	1.3	81
Average	16.6	16.6	16.6
<b>Marital status</b>			
Single	7.5	7.6	484
In union	77.4	77.1	4929
No longer in union/ever married	15.1	15.4	982
<b>Number of ever born children</b>			
0	9.9	9.6	617
1+	90.1	90.4	5778
<b>Age at first birth</b>			
Never give birth	10.3	10.6	660
<20 years old	64.8	64.2	4144
20 +	24.9	25.2	1591
<b>Ever had premarital child</b>			
No	88.7	88.4	5652
Yes	11.3	11.6	743
<b>Relationship to the head of household</b>			
Head of household	19.4	19.0	1213
Spouse	62.6	62.4	3992
Daughter and grand daughter	11.0	11.7	748
Others	7.1	6.9	442
<b>Province of residence</b>			
Northern	11.1	17.5	1122
Central	42.2	34.1	2181
Southern	46.7	48.4	3092
<b>Place of residence</b>			
Urban	19.2	13.1	837
Rural	80.8	86.9	5558
<b>Religion</b>			
Catholic	21.2	20.6	1316
Protestant	24.3	25.2	1610
Other Christians	39.7	42.3	2708
Muslim	13.5	10.9	695
Others	1.3	1.0	66
<b>Education</b>			
None	17.5	16.6	1060
Primary	63.8	66.4	4246
Secondary+	18.7	17.0	1089

Continued

**Table 1** Continued

Socioeconomic and demographic Characteristics	Per cent		Number†
	Weighted*	Unweighted	
<b>Household wealth quintiles</b>			
Poorest	17.6	19.0	1215
Poorer	20.1	20.6	1319
Middle	19.7	20.9	1334
Richer	19.3	20.7	1323
Richest	23.3	18.8	1204
Total			6395

\*Interpretations are based on weighted percentage.

†Unweight crude numbers.

lowest wealth quintile—node 11 (60% vs 22%). For women in the age groups of 15–19 and 20–24 years (node 4) with an HIV prevalence of 13%, the relationship to the head of the household is the best predictor of HIV infection ( $\chi^2=11.1$ ,  $p<0.003$ ). Women head of a household (node 12) have a higher HIV prevalence compared with other women with a different relationship to the head of the household—node 13 (24% vs 7%). The region of residence is the best predictor of HIV infection among women aged 35–39 years ( $\chi^2=11.5$ ,  $p<0.002$ ), with women living in the Southern region (node 14) having an HIV prevalence about twice that of the women from the Central and Northern regions—node 15 (59% vs 31%). Among women aged 25–29 years (node 6) accounting for 3% of the study population with HIV prevalence of 27%, age remains the only significant and final predictor of HIV prevalence.

#### Node 2: women in union and never married women

This group includes women in union (married or living together) and those who have never been in union, representing 85% of the study population and have HIV prevalence of 10%. Place of residence (rural or urban) is the best predictor of HIV infection with a higher prevalence in urban areas (node 8) compared with rural areas—node 7 (21% vs 9%,  $\chi^2=89.8$ ,  $p<0.001$ ).

For women living in rural areas (node 7) and representing 74% of the population, the best predictor for HIV infection is age ( $\chi^2=86.0$ ,  $p<0.001$ ) with the highest prevalence among women aged 30–44 years (13%) followed by the age group 25–49 (node 18: 9%) and the age group 15–24 (node 18: 4%). Similarly, age is a strong predictor of HIV infection ( $\chi^2=86$ ,  $p<0.001$ ) among women living in urban areas (node 8), where the age group 30–49 (node 19) has a prevalence about twice that among the age group 15–29—node 20 (29% vs 15%; Table 3).

#### HIV risk groups in Malawi

There are in total 13 homogeneous subgroups or terminal nodes. Table 4 describes these 13 subgroups (terminal

**Table 2** HIV prevalence by selected socioeconomic and demographic characteristics

Socioeconomic and demographic characteristics	HIV+ prevalence (%)	Total (N)	$\chi^2$	p Value
Age				
15–19	5.0	744		
20–24	6.9	1327		
25–29	12.4	1402		
30–34	19.8	1001	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	1230		
15–19	13.1	4376	9.13	0.028
20–24	12.4	708		
25 and +	18.5	81		
Marital status				
Single/never married	7.9	484		
In union	10.7	4929	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 and +	14.2	5778		
Age at first birth				
Never give birth	9.1	660		
<20 years old	14.1	4144	12.96	0.002
20+	14.3	1591		
Ever experience premarital childbearing				
No	12.8	5652	26.99	<0.001
Yes	19.8	743		
Relationship to the head of household				
Head of household	25.0	1213		
Spouse	10.2	3992	179.93	<0.001
Daughter and grand daughter	11.9	748		
Others	17.0	442		
Region of residence				
Northern	10.0	1494		
Central	9.5	3062	184.90	<0.001
Southern	20.0	4444		
Place of residence				
Urban	24.7	1156	157.00	<0.001
Rural	12.3	7844		
Religion				
Catholic	12.6	1316		
Protestant	14.3	1610		
Other Christians	13.4	2708	2.66	0.616
Muslim	14.8	695		
Others	13.6	66		
Education				
None	13.9	1060		
Primary	12.8	4246	10.73	0.005
Secondary+	16.6	1089		
Household wealth quintiles				
Poorest	10.3	1215		

Continued

**Table 2** Continued

Socioeconomic and demographic characteristics	HIV+ prevalence (%)	Total (N)	$\chi^2$	p Value
Poorer	10.9	1319		
Middle	11.6	1334	88.34	<0.001
Richer	14.3	1323		
Richest	21.5	1204		
Total	13.6	6395		

nodes) in terms of their composition, demographic weight in the sample (columns 1 and 2), their share in the HIV burden (columns 3 and 4) and their corresponding HIV prevalence (column 5). The 13 homogeneous subgroups could be grouped into five major groups.

*Group 1* represents 3% of the sample with an overall HIV prevalence of 59%. This group accounts for 11% of all the women who are HIV positive. Group 1 includes two subgroups: (1) women in union disruption living in the richest household and aged 30–34 or 40–49 years and (2) women in union disruption living in the Southern region and aged 35–39 years.

*Group 2* represents 5% of the sample with an overall HIV prevalence of 35% and accounts for 12% 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) aged 30–34 or 40–49 years and women in union disruption aged 35–39 years and living in the Northern or Central region.

*Group 3* represents about 10% of the study population with an overall HIV prevalence of 27% and accounts for 20% of all HIV-positive women. This group is divided into four subgroups: (1) never married and women in union, living in an urban area and aged 30–49 years; (2) formerly in union (widowed or divorced) women aged 25–29; (3) young women (15–24) formerly in union who are the head of the household and (4) formerly in union women living in the poorest household and aged 15–24, 30–34 or 40–49 years.

*Group 4* represents about 33% of the study population with an overall HIV prevalence of 14% and accounts for 33% of all the HIV-positive women. This group includes adolescent (15–19), never married women or women in union living in urban areas; and never married or women in union living in rural areas aged 30–44 years.

*Group 5* represents 50% of the study population and has the lowest HIV prevalence of 7%, but accounts for 23% of all the HIV-positive women. This group includes three subgroups: (1) never married or women in union living in rural areas and aged 25–29 or 45–49 years; (2) young women aged 15–24 years who are no longer in union and are not the head of the household and (3) young women (15–24) who are never married or in union and are living in rural areas.

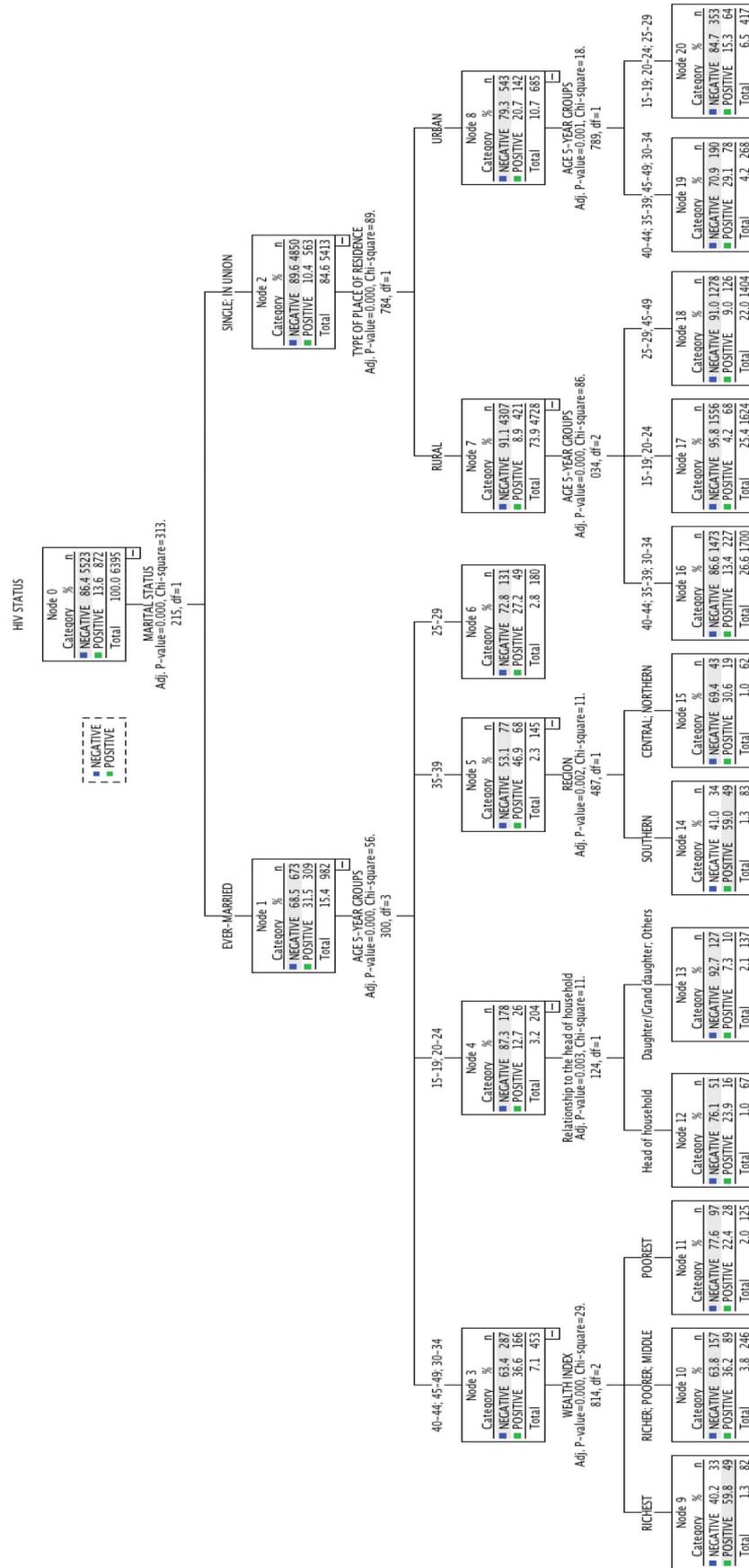


Figure 1 HIV prevalence in Malawi: (A) tree diagram for women in union, (B) for women in union disruption and (C) for never married women.

**Table 3** Summary information on the specifications used to build the CHAID model and the resulting model

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	<i>Marital status</i> , 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

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 CHAID,  $\chi^2$  Automatic Interaction Detector.

Table 4 also reports 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 a group with high HIV prevalence but with small demographic weight in the population (categories 1–3). 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.

**DISCUSSIONS**

This paper aimed to describe and profile HIV prevalence among women in Malawi. The study used  $\chi^2$  and CHAID techniques to analyse data from the Malawi 2010 DHS.

The analyses suggested three keys findings. First, consistent with previous studies,<sup>6 23</sup> findings from bivariate analysis and a  $\chi^2$  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 also confirmed 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 a high education level, and belonging to a high socioeconomic class of the society.<sup>24 25</sup> In parallel, relatively rich and better-educated men have higher rates of partner change because they have greater personal autonomy and spatial mobility.<sup>26–29</sup> Women’s economic dependence on their partners may also make it difficult for them to insist on safer sex (eg, condom use). Concentration (about 50%) of the most educated, richest

and urban women in the Southern region may explain the 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) and aged 30–34 or 40–49 years have a significantly higher HIV prevalence. This may be because: (1) husbands from the highest quintile or a male partner may have more access to transactional sex and other risk behaviours such as polygamy which may increase women’s vulnerability to HIV; (2) wealthier HIV-positive widowed women may have a better quality of life as well as better access to treatment and survive longer.<sup>30</sup> Furthermore, divorced and separated women are more frequent among the most educated women with economic autonomy.<sup>31</sup> Their causes (polygyny and/or infidelity) as well as consequences (multiple sexual partnerships) are also factors associated with HIV prevalence.<sup>32 33</sup>

Third and last, the CHAID models also depicted different interactions between risk factors and the profiled HIV risk groups in Malawi. For instance, while overall HIV prevalence among women living in urban areas (25%) is twice the prevalence observed among women living in rural areas (12%), HIV prevalence is estimated at 15% among never married women or women in union living in urban areas aged 15–29, and at 13% among never married women or women in union living in the rural areas aged 30–44 years. Likewise, whereas in general, HIV prevalence is low among never married women and women in union (10%), CHAID results revealed a higher HIV prevalence (29%) among never married women and women in union aged 30–49 years who live in urban areas compared

**Table 4**  $\chi^2$  Automatic Interaction Detector gains for nodes

Group/node number	Node description	Node		Gain		Percentage	
		N*	%†	N‡	%§	of HIV¶	Index**
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		2117	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	1700	26.6	227	26.0	13.4	97.9
Group 5		3165	49.5	204	23.3	6.5	42.3
18	Never married/in union, rural area, 25–29 and 45–49	1404	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	1624	25.4	68	7.8	4.2	30.7
Total		6395	100	872	100	13.6	–

\*Number of cases per node (demographic size in the sample).

†Demographic size in percentage= $(0.1/\Sigma.1) \times 100$ .

‡Number of HIV women.

§Demographic size among HIV-positive women in percentage= $(0.3/\Sigma.3) \times 100$ .

¶HIV prevalence in percentage= $(0.3/\Sigma.1) \times 100$ .

\*\*Node index= $((0.3/\Sigma 3)/(0.1/\Sigma.1)) \times 100$ .

with: (1) women in union disruption aged 15–24 (7% if they are not the head of the household and 23% for the head of the household); (2) women in union disruption aged 25–29 (27%) and (3) women in union dissolution aged 30–34 and 40–49 years who live in the poorest households (22%).

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 on earlier ones and not simultaneous as would be the case in a regression model or analysis of variance where all effects are fitted simultaneously. Furthermore, CHAID allows automatic detection of interaction between variables.

In the light of these findings, it is noteworthy that to reduce the number of new infections, interventions should be targeted and prioritised according to the prevalence and demographic size of different risk groups. Furthermore, policy makers' prioritisation of interventions may also depend on the preference for preventive interventions compared with 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 with care and treatment.<sup>34</sup>

Regarding preventive interventions, the findings suggested that:

1. Couples (men and women in union) and never married people aged 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% of the study population, among whom the HIV prevalence is estimated at 11% on average. About 40% of women living with HIV in Malawi belong to this category.
2. Young aged 15–24 years and living in rural areas (node 17) and urban adolescents aged 15–19 (node 20) are the second most important target. This group accounts for 32% of the studied population and 15% of women living with HIV. Besides, a majority of adults living with HIV may be infected during adolescence. Unfortunately, the available dataset could not provide information on the 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% on average among young women aged 15–24 years (table 2), that prevalence is estimated to be above 20% among young women formerly in union and among young women who ever experience premarital childbearing and live in urban areas. Likewise, HIV prevalence is very high among women in union disruption (32% on average) compared with other groups (10%). 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 aged 30–49 years and living in urban areas. Indeed, formerly in union women represented only about 13% of women of reproductive age in Malawi; they have the higher HIV prevalence ranges between 22% observed among the poorest and 60% among the richest.

Nevertheless, to achieve zero new infection as part of MDG6, there is a need for a more comprehensive policy to combat HIV because of the complexity of the HIV-socioeconomic profile in Malawi. There are several groups built from several socioeconomic categories depending on the individual marital status, wealth index, age, place of residence and relationship to the head of the household. In South Africa, Bendavid *et al*<sup>35</sup> 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% 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 do not permit one to draw causal associations between HIV status and the associated factors. For instance, whether HIV infection has occurred before, during or after the union. Last, the CHAID model ignores the hierarchical structure of the Demographic and Health Survey data and needs a large sample size.

In conclusion, this study recommends: (1) the 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 (FP) 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 campaigns, reproductive health services and reproductive health courses at school could be used as canons to achieve a universal prevention strategy, testing, counselling and treatment.

#### Author affiliations

<sup>1</sup>Department of Population and Development Studies, University of Kinshasa, Kinshasa XI, Democratic Republic of Congo

<sup>2</sup>AFRILUX, CEPS/INSTEAD 3, Esch-sur-Alzette, Luxembourg

<sup>3</sup>Research Coordination, INDEPTH Network, Accra, Ghana

<sup>4</sup>Faculty of Social and Human Sciences, University of Southampton, Southampton, UK

<sup>5</sup>African Institute for Development Policy (AFIDEP), Westlands, Nairobi, Kenya

<sup>6</sup>ICF International, Calverton, Maryland, USA

**Acknowledgements** The authors thank Macro International for providing the free Malawi 2010 DHS datasets including the HIV data. We thank Professor Guillaume Wunsch for his comments. Furthermore, we have appropriate approval and authorisation from the DHS datasets research ethics board for conducting HIV secondary data analysis.

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

**Funding** This research was supported by the Luxembourg National Research Fund (FNR).

**Competing interests** None.

**Patient consent** Obtained.

**Ethics approval** The DHS questionnaire and the testing protocol undergo a host country ethical review as well as an ethical review at ICF Macro. Furthermore, participation in individual survey and in HIV testing is voluntary. Participants should sign the consent form before interview and before blood collection.

**Provenance and peer review** Not commissioned; externally peer reviewed.

**Data sharing statement** This study is based on the Malawi 2010 Demographic and Health Surveys. These data are available on <http://www.measuredhs.com>. Access to individual HIV status as well as individual background information required authorisation from MACRO ORC (<http://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/](http://www.unaids.org/en/resources/campaigns/20121120_globalreport2012/globalreport/) (accessed 12 Feb 2013).
3. 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.
4. UNAIDS. *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 Infect Dis* 2008;8:174. <http://www.biomedcentral.com/1471-2334/8/174> (accessed 12 Feb 2013).
6. Roux P, Henley L, Cotton M, *et al*. 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. S Afr Med J* 2000;90:1008–11.
7. Kumar R, Mehendale SM, Panda S, *et al*. 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> (accessed 12 Feb 2013).
8. Marseille E, Kahn JG, Billingham K, *et al*. Cost-effectiveness of the female condom in preventing HIV and STDs in commercial sex workers in rural South Africa. *Soc Sci Med* 2001;52:135–48.
9. Sweat M, Gregorich S, Sangiwa G, *et al*. Cost-effectiveness of voluntary HIV-1 counselling and testing in reducing sexual transmission of HIV-1 in Kenya and Tanzania. *Lancet* 2000;356:113–21.
10. Durevall D, Lindskog A. *HIV/AIDS, adult mortality and fertility: evidence from Malawi. Working papers in economics, No 284*. Göteborg: School of Business, Economics and Law, University of Gothenburg, 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. Ojteg K. *Socio-economic determinants of HIV in Zambia. A district level analysis*. Ph.D. 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;17:1067–83.
14. Magadi M. Understanding the gender disparity in HIV infection across countries in sub-Saharan Africa: evidence from the Demographic and Health Surveys. *Social Health Illn* 2011;33:522–39.

15. Asiedu C, Asiedu E, Owusu F. The Socio-economic determinants of HIV/AIDS infection rates in Lesotho, Malawi, Swaziland and Zimbabwe. *Dev Policy Rev* 2012;30:305–26.
16. Green E. *Rethinking AIDS prevention. Learning from successes in developing countries*. Westport, CT: Praeger publishers, 2004.
17. Population Reference Bureau. *2012 World Population Data sheet*. Washington DC: PRB, 2012.
18. United Nations Development Programme/UNDP. *Sustainability and equity: a better future for all, human development report 2011*. New York: UNDP, 2011.
19. Malawi, HIV National Commission. *National 2012 global AIDS response progress report: Malawi country report*. Malawi: HIV National Commission, 2012;67.
20. Malawi National Statistical Office (NSO); ORC Macro. *Malawi demographic and health survey 2004*. Calverton, MD: NSO and ORC Macro, 2005.
21. Malawi National Statistical Office (NSO); ICF Macro. *Malawi demographic and health survey 2010*. Zomba, Malawi and Calverton, MD: NSO and ICF Macro, 2011.
22. Kass GV. An exploratory technique for investigating large quantities of categorical data. *Appl Stat* 1980;29:119–27.
23. 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.
24. Cleland J, Van Ginneken J. Maternal education and child survival in developing countries: the search for pathways of influence. *Soc Sci Med* 1988;27:1357–68.
25. Barrett H, Brown A. Health, hygiene, and maternal education: evidence from Gambia. *Soc Sci Med* 1996;43:1579–90.
26. Ainsworth M, Semali I. Who is most likely to die of AIDS? Socioeconomic correlates of adult deaths in Kagera Region, Tanzania. In: Ainsworth M, Fransen L, Over M, eds. *Confronting AIDS: evidence for developing world*. Washington, DC: The World Bank, 1998:95–110.
27. Gregson S, Garnett GP, Nyamukapa CA, et al. HIV decline associated with behaviour change in eastern Zimbabwe. *Science* 2006;311:664–6.
28. Hargreaves JR, Glynn JR. Educational attainment and HIV-1 infection in developing countries: a systematic review. *Trop Med Int Health* 2002;7:489–98.
29. Gillespie S, Kadiyala S, Greener R. Is poverty or wealth driving HIV transmission? *AIDS* 2007;21(Suppl 7):S5–16.
30. Rodrigo C, Rajapakse S. HIV, poverty and women. *Int Health* 2010;2:9–16.
31. Takyi B, Broughton C. Marital stability in sub-Saharan Africa: do women's autonomy and socioeconomic situation matter? *J Fam Econ Issues* 2006;27:113–32.
32. Reniers G. Divorce and remarriage in rural Malawi. *Demogr Res* 2003; Special Collection 1, article 6:175–206. <http://www.demographic-research.org/special/1/6/S1-6.pdf> (accessed 12 Feb 2013).
33. Reniers G. Marital strategies for regulating exposure to HIV. *Demography* 2008;45:417–38.
34. Youngkong S, Baltussen R, Tantivess S, et al. Criteria for priority setting of HIV/AIDS interventions in Thailand: a discrete choice experiment. *MC Health Serv Res* 2010;10:197. <http://www.biomedcentral.com/1472-6963/10/197> (accessed 12 Feb 2013).
35. Bendavid E, Brandeau ML, Wood R, et al. Comparative effectiveness of HIV testing and treatment in highly endemic regions. *Arch Int Med* 2010;170:1347–54.