Web appendix

Table of Contents

Methods

Sampling methodology

We used multi-stage sampling to randomly select households and recruit a household-representative sample of men and women. The enumeration area was the primary sampling unit. The sampling frame for the number of households and the number of persons in each enumeration area was created from three sources; a) the Census undertaken in 2011, b) a Community survey undertaken in 2007 and c) aerial imaging of dwellings supplied by Geo Terra Image.

The study area consists of an estimated 100,018 households with a total of 243,115 females and 201,870 males. A total of 244,699 individuals in the age range 15-49 years are estimated to live in the study area (based on the 2011 Census).

From a total of 600 enumeration areas, all 591 enumeration areas with more than 50 households were included in the sample. Of these, 221 enumeration areas were drawn randomly. Within an enumeration area the households were drawn systematically with a random start. Study staff identifed households using the Global Positioning Systems receiver to record the geographic coordinates of each randomly selected household. Sampling was designed to obtain approximately 10,000 individuals. In instances where a selected household was abandoned, refused to complete the composition form or the members away for an extended period of time, then the household on the right side of the selected house, when facing the entrance of the selected household, was used as a replacement.

Once a household was selected, and the head of the household provided consent for the household to participate, a handheld personal digital assistant was used to compile a list of the individuals residing in the household. The eligibility of each individual was determined by the personal digital assistant. These individuals were numbered and the handheld device selected one of these individuals at random to be included in the study. If this individual refused participation a second individual was selected randomly. Only one individual per household was selected and enrolled in the study. Should the second individual selected also refuse the household was replaced.

Eligibility criteria

Men and women aged 15 to 49 years (inclusive) who resided in the selected household, who were willing to participate and provide written informed consent (or parental consent with child participant assent if <18 years of age) in either English or Zulu, and were willing to undergo all study procedures including provision of clinical specimens (peripheral blood, urine, sputum-if indicated and for females, self-collected vulvo-vaginal swabs) were eligible. Those who refused to participate and/or provide samples, who could not consent, were mentally challenged, or who were planning to relocate within 12 months were excluded.

Weighting

Weights were calculated taking into account the probability of selecting the enumeration area, the probability of selecting the household in the enumeration area and the probability of selecting the individual in a household, and was adjusted for non-response. The weights were then revised to reflect the size of the population in the study area.

The weights were calculated in three stages. In the first stage the probability of the enumeration area being selected was calculated. The probability of selecting an enumeration area was the number of enumeration areas selected divided by the total number of enumeration areas.

In the second stage the probability of a household being selected was calculated. The probability of a household being selected was the number of households selected in the enumeration area divided by the total number of households in the enumeration area. Within each enumeration area the number of households that were found to be not eligible or who refused participation or could not be contacted was calculated. The weight for a household was then calculated as the reciprocal of the probability of selection multiplied by the probability of responding, thus inflating weights for non-response. Household weights above the 97.5th percentile and below the 2.5th percentile were truncated to remove extreme weights.

The third stage involved calculating the probability of selecting an individual within a household. The probability of selecting an individual is the reciprocal of the number of eligible household members. The proportion of individuals of a certain age and gender who refused participation was calculated. The individual weight was then calculated as the reciprocal of the probability of selection multiplied by the probability of responding. This inflated the weights for individuals of the same age and gender as individuals who refused participation.

The weight for each individual was calculated as the product of the three weights calculated in the three stages (enumeration area weight, household weight and individual weight). Individual weights above the 97.5th percentile and below the 2.5th percentile were truncated to remove extreme weights.

The final step was to benchmark these weights against census data by age group and gender to ensure that the weights of the participants sampled added up to the population estimates for the area. This means that the weight of an individual can be thought of as the number of people in the community that the selected individual represents.

Sample size

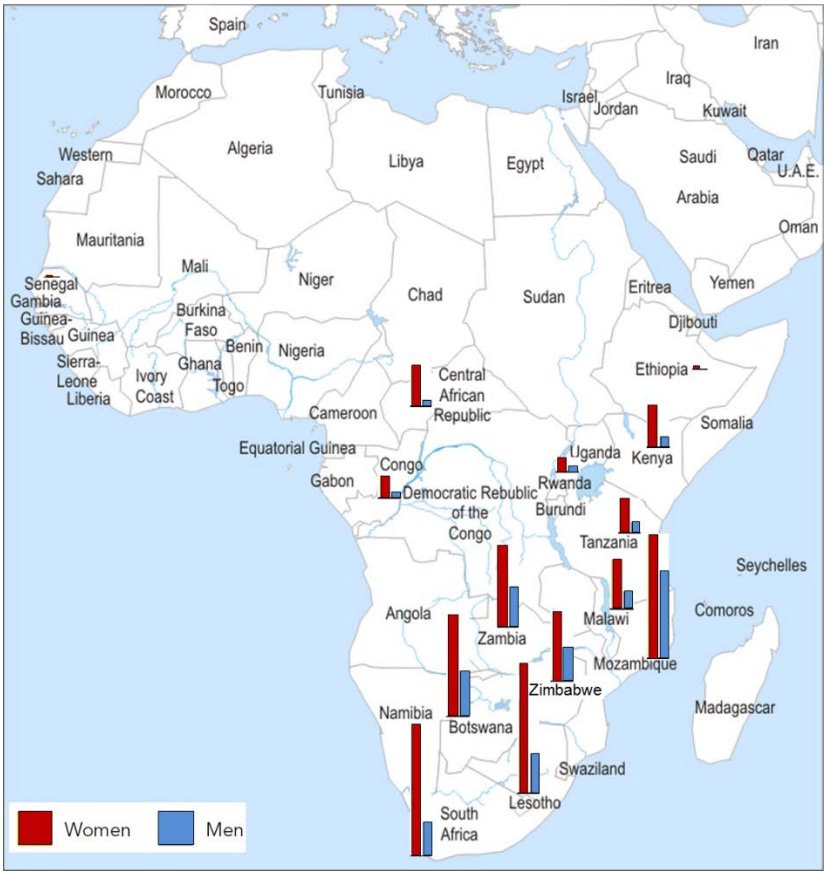
The sample size of this cross-sectional survey was derived from the sample size for the primary HIV incidence endpoint of the parent study. The detailed calculation of the sample size has been previous described[1]. In summary, the sample size was based on recently conducted longitudinal studies that reported HIV incidence rates in women of 6.5[2] and 6.3[3] per 100 person years in Vulindlela and Greater Edendale respectively. Little or no data on HIV incidence on men from these districts were available. However, cross sectional surveys have shown that HIV prevalence is at least 5 times higher in young girls compared to young boys in the 15-24 year age groups and assume a slightly lower population based HIV incidence rate. The sample size of 10,000 was selected for this survey so that the parent study had 84% power to detect a 30% reduction in HIV incidence rate at a 5% significance level, given an HIV prevalence of 20%, loss-to-follow-up of 15% per annum and an initial HIV incidence rate of 3 per 100 person years.

A total of 14,618 households were visited. The response rate at household level was 85.3% (12,474/14,618). A total of 11,289 individuals were approached for participation in the study; 577 (5.1%) refused, 488 (4.3%) were still in the process of completing enrolment when this study ended, 398 (3.5%) were not eligible and 14 (0.1%) had no HIV results available. Of the 577 individuals who refused participation, 246 were men and 331 were women. A total of 9,812 individuals were analysed for this study.

Results

Supplementary Table 1: Cluster size distribution by gender composition of the clusters

Number of individuals in the cluster	Male-female clusters (n=90)	Male only clusters (n=21)	Female only clusters (n=91)	All clusters (n=202)
2	69	19	80	168
3	13	0	9	22
4	5	1	2	8
5	1	0	0	1
6	1	1	0	2
18	1	0	0	1



Source: Adapted from UNAIDS 2012

Supplemental Figure 1: HIV prevalence in 15–24 year men and women in the years 2008 to 2011 in southern and eastern African countries

References

1. Kharsany AB, Cawood C, Khanyile D, Grobler A, McKinnon LR, Samsunder N, et al. Strengthening HIV surveillance in the antiretroviral therapy era: rationale and design of a longitudinal study to monitor HIV prevalence and incidence in the uMgungundlovu District, KwaZulu-Natal, South Africa. BMC public health. 2015; 15: 1149.
2. Abdool Karim Q, Kharsany AB, Frohlich JA, Werner L, Mashego M, Mlotshwa M, et al. Stabilizing HIV prevalence masks high HIV incidence rates amongst rural and urban women in KwaZulu-Natal, South Africa. Int J Epidemiol. 2011; 40(4): 922-30.
3. Nel A, Mabude Z, Smit J, Kotze P, Arbuckle D, Wu J, et al. HIV incidence remains high in KwaZulu-Natal, South Africa: evidence from three districts. PLoS ONE. 2012; 7(4): e35278.