# Annex 1. Multivariate analysis of South African survey data on HIV status and sexual behaviour

(Emma Slaymaker, LSHTM & Vicky Hosegood, Africa Centre)

## Section 1A. Introduction

This annex describes the methods and results of survey specific and pooled analyses of nationally representative surveys on HIV related behaviours. The five surveys included are all nationally-representative household surveys which collected information on sexual behaviour, HIV knowledge and awareness, and exposure to HIV prevention activities, in addition to socio-demographic background information. They are:

By the Human Sciences Research Council (HSRC), the "*National HIV prevalence, behavioural risks and mass media household survey 2002*", the "*National HIV Prevalence, Incidence, Behaviour and Communication Survey 2005*" and the "*National HIV Prevalence, Incidence, Behaviour and Communication Survey 2008*", subsequently referred to as HSRC 2002, HSRC 2005 and HSRC 2008.

By a team from Health & Development Africa (HDA) and Johns Hopkins Health and Education in South Africa (JHHESA), the "*National Communications Survey 2006*" and the "*National Communications Survey 2009*", subsequently referred to as NCS 2006 and NCS 2009.

In addition to interviews, the HSRC surveys collected samples from consenting respondents which were tested for HIV. In 2002 saliva-based tests were used; in 2005 and 2008 blood samples were taken. HIV status was not collected in either of the NCS.

A conceptual framework for the factors associated with HIV infection (show in **Figure 1.1.1**) was used to guide analysis. This was based on the proximate determinants framework[1] and separates factors into three levels, the background (most distal) factors, those with determine exposure to an infected partner and those which affect transmission (most proximate). In the framework the socio-demographic factors encompass the time and place of each individual, which describes the prevailing HIV prevalence. Individual factors such as mobility, education, socio-economic status (SES), religion and marital status can influence HIV risk by modifying the opportunity for, motivation for and choice of sexual partners and can also influence the risk of transmission via condom use and STI prevention and treatment. These datasets contained information on prevalent, rather than incident HIV cases. Analysis of HIV status therefore focussed on describing who was infected, and what their current behaviour was, rather than trying to understand what factors led to their infection.

Exposure to an HIV-infected partner is determined by the rate at which partners are acquired and the duration of partnerships. Greater numbers of partners present more risk, whether the risk is increased more by many new partners, or even just a small number of long term partners will depend on the epidemic stage and incidence rate. New, short term partners, are most risky if already infected because in relationships of very short duration (i.e. a few days) there is little chance for a new partner to sero-convert and transmit HIV. However, longer term partners have more opportunity to acquire HIV-infection from a third partner, sero-convert and infect the first partner (the survey respondent in our data). The relative risks of short- and long-duration partners therefore depend on HIV incidence rates and HIV prevalence. Where the HIV status of partners is unknown (as in these data), it can be approximated by other characteristics: age, how the relationship is described by the respondent, where the index person met the partner, the partner's age and marital status.

At the level of individual sex acts, HIV transmission is influenced by coital frequency, male circumcision, presence of other STI in either partner and condom use. This is the area with the least available information, not least because to truly capture this one would need information on each sex act with every partner.

The framework was used to derive variables and to structure the regression models for analysis of HIV status and behavioural outcomes.

---

1 Boerma JT & Weir SS (2005). Integrating demographic and epidemiologic approaches to research on HIV/AIDS: the proximate determinants framework. Jour Inf Dis; 191(Suppl 1):S61-S67.

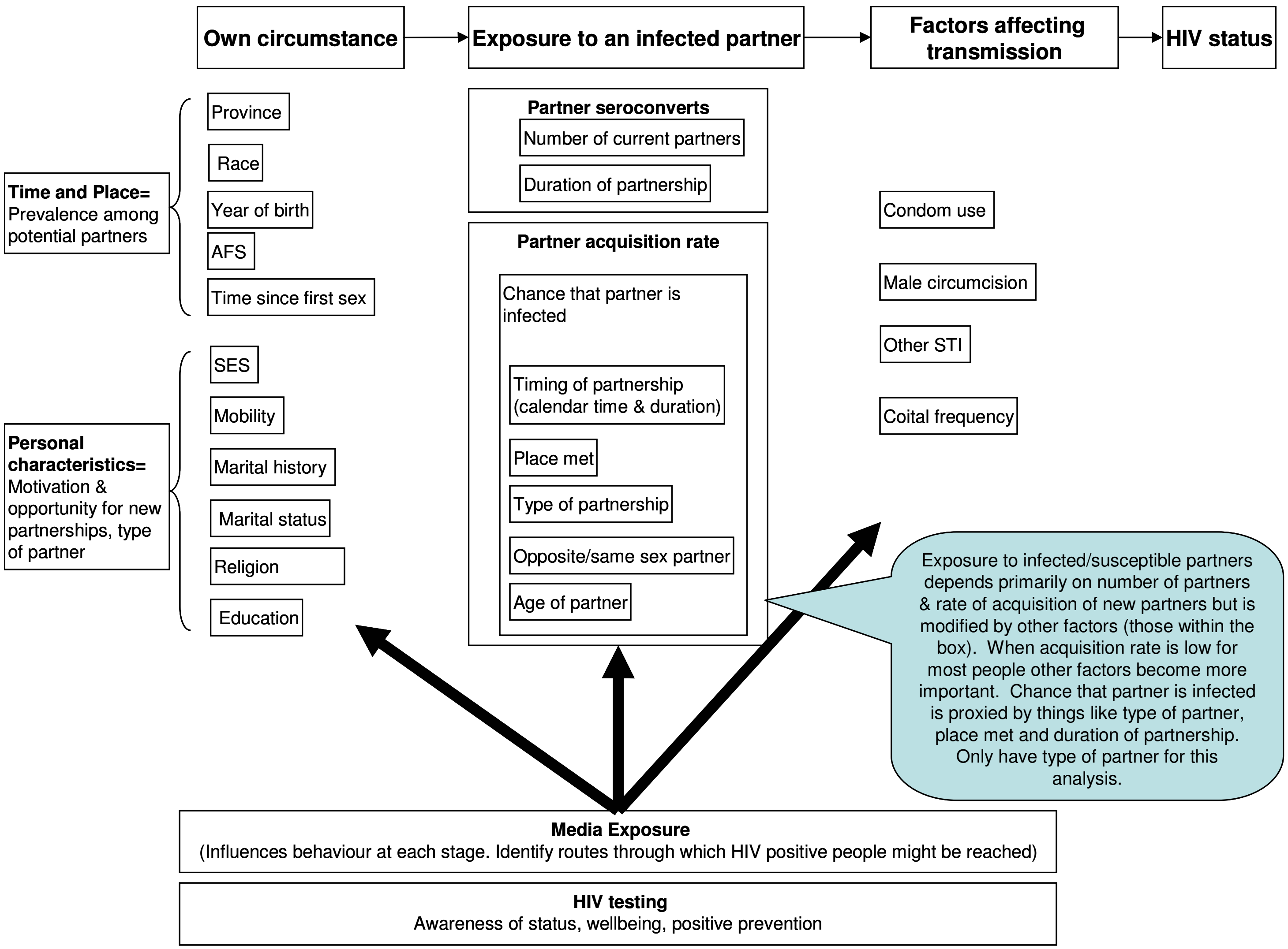



**Figure 1.1: Conceptual framework used for analysis**

## Aims of the multivariate analysis

1. To identify and characterise HIV positive individuals, including any current behaviours which might facilitate onwards transmission.
2. To describe factors associated with any risky sexual behaviours that were found to be associated with HIV status at survey.
3. To compare the proportion reporting multiple partners in the five different surveys, to look for evidence of a trend over time.
4. To compare the proportion HIV positive in the five different surveys, to look for evidence of a trend over time.
5. To describe acquisition rates for new partners, and factors associated with this.

The first aim was focussed on finding factors which were more common among HIV positive people at survey, and which might contribute to onwards transmission of infection, rather than finding factors that might explain their infection. With cross-sectional data, and prevalent rather than incidence cases, it is impossible to establish to order of events and highly unlikely that the survey captured the behaviour during the period prior to infection. It is therefore unwise to use these data to try and explain the current HIV status of respondents.

# Section 1B. Methodology

Datasets for 2002, 2005 and 2008 surveys were supplied by HSRC for respondents aged 15 and over. Different questionnaires had been used in each survey and so there is some difference between the years in the information available. Standard variables were derived from the datasets (see **Table 1.2**) based on the conceptual framework and the bivariate results carried out by HSRC for all respondents in the 2008 survey.
Datasets for the 2006 and 2009 NCS were supplied by HDA for respondents aged 15 to 65 and 16 to 56, respectively. Questionnaires were different to each other, and to the HSRC surveys. The same set of standard variables were derived, so far as was possible, and these are also shown in **Table 1.2**.

Each survey used a different approach to collect information on sexual partners in the year before the survey. In HSRC 2005 and HSRC 2008 respondents were asked to classify their sexual partners in the last year as regular, non-regular (had sex only a few times) and commercial sex workers (CSW) and these categories were used for analysis. In the 2009 NCS the relationship to the partner was collected in the partner history. The categories reported were combined into spouse or cohabiting partner, boyfriend or girlfriends (those described as "main partners") and casual (those described as "friends", "someone I've known for a while", "Someone I just met", "One night encounter"). Respondents were asked if they had given money or resources to any of their partners in the last year. Those described as casual, and who were given money or resources were identified as CSW for this analysis. Therefore the partner types from the HSRC 2005 and HSRC 2008 are not directly comparable with those from NCS 2009.

**Missing values**
Values missing by design because of skips in the questionnaire were replaced with the appropriate code. For example, respondents who had never had sex were not asked the questions about sexual behaviour and had been coded as missing on those variables. For this analysis those respondents were coded as "never had sex" for all the sexual behaviour variables to maintain the same denominator throughout the analysis.
For many variables, there were also some missing values where there should have been a response. These were coded as "missing" to ensure a consistent denominator, and to explore whether there were any patterns to the missing values.
Complete information was not available for all respondents. Multivariate analysis is only possible for complete records and we examined two approaches to handling missing data in the analyses:
1) Give missing values a numerical code and include these as a category in the model. This means that only categorical variables could be included, because it is not possible to distinguish a missing value from a real response in a continuous variable.
2) Use only complete data for analysis. The variables identified for the regression model were used to define a respondent's data as complete or not.

The second option was chosen because, although it meant losing about a substantial proportion of respondents from some surveys, the resulting models are be easier to interpret. The missing values were present for all variables which meant the addition of an extra parameter to the model for each variable and, for most variables, this extra category contained only a small number of respondents which made the coefficient for the group difficult to interpret and reduced the statistical power.

**Methods for Analysis**

Analysis was conducted on respondents with complete data and, for HSRC, only those who were also tested for HIV. For the HSRC data, the weights used were those for respondents tested for HIV which included an element to adjust for non-response to the HIV testing component of the survey. For these reasons, the results presented here differ from those in the 2008 HSRC report, which included all respondents (included those not tested for HIV) and used a different weighting scheme for the non-HIV outcomes. All standard errors, confidence intervals and statistical tests were calculated accounting for the complex survey design using Stata 11's svy commands and using the survey design variables listed in **Table 1.1.1.**

**Table 1.1: Survey design information in each dataset. EA stands for Enumeration Area and refers to the standard enumeration areas used for the most recent South African census.**

| | Strata (variable name) | PSU (variable name) | weight (variable name) |
|---|---|---|---|
| **HSRC 2008** | Province & age group (prov & agegrp) | EA (ea) | ibreal1 |
| **HSRC 2005** | Province (province) | EA (eanumber) | ibreal1 |
| **HSRC 2002** | Province (cprov) | EA (ceanum) | cweight |
| **NCS 2009** | Province (province) | EA (psu) | weight_stdpost |
| **NCS 2006** | Province (provincs) | n/a | weight |

**Univariate Analysis**

The percentage distributions of respondents by each variable listed in **Table 1.1** were tabulated, accounting for the complex survey design, separately by sex and year of survey.

**Bivariate Analysis (HSRC data only)**

The percentage HIV positive by each variable was tabulated and chi-squared tests calculated, allowing for the survey design using Stata 11's svy commands.

**Bivariate Analysis (2009 NCS data only)**

The NCS 2009 collected information on partners with whom the respondent had had sex during the 12 months before the survey. The data from this partnership history were rearranged to form a sequence of events (partner acquisition, partner loss, censoring at survey date) by date for each respondent. The number of partnerships in which each respondent was engaged was calculated for the time of each reported event. This makes it possible to distinguish respondents who embarked on a new partnership when they already had at least one current partner. Survival analysis was used to calculate partner acquisition rates (PAR) where the outcome event was starting a partnership, defined by the date of first sex with a partner.

**Multivariate analysis**

*HIV status (HSRC data)*

Logistic regression models were fitted separately for men and women to identify the factors associated with HIV status at survey.

All variables identified as important in the univariate analysis were put into a fixed effects logistic regression model adjusted for survey design. Variables with a p-value on the Wald test of <=0.01, for at least one category, were retained. Those with large p-values were removed, one by one, starting with those on the left of the conceptual framework.

Once unimportant variables had been removed, interactions were explored for selected variables.

*Media Exposure*

Respondents were asked how often they listened to, saw or read the: radio, TV, newspapers, magazines, internet. Response categories were: Never, once a week, 2-6 times a week, every day. These variables were examined separately in logistic regression models for the HIV outcome and then all combined into one model. Variables which showed a non-linear association with HIV were treated as categorical variables, the others were treated as continuous variables for the multivariate analysis.

*Province level effects (HSRC 2008)*
After establishing the fixed effects for province (as covariates in the logistic regression models) and looking for evidence of effect modification ( by fitting interaction terms) the crude odds ratios for HIV infection were calculated for each factor separately for each province. These results were used to assess whether any of the variables revealed to be important in the regression models was operating differently at provincial levels.

*Partnership behaviours: HSRC 2008*
A regression model was fitted to explore the factors associated with sex with a commercial sex worker among men in the 2008 HSRC survey because this was the only behavioural factor to remain associated with HIV once other variables were controlled for.

*Partnership behaviours: NCS*
Piecewise exponential models, adjusted for survey design, were fitted to the NCS 2009 data to estimate the effects of covariates on the hazard of acquiring of a new partner. Analysis time was split into pieces corresponding to partner status categories (no partner; at least one non-spousal partner; at least one spousal partner; at least one non-spousal plus one or more spousal partners) for the respondent during the year prior to the survey. A crude model was fitted showing the effect of partner status on partner acquisition. An adjusted model was then fitted to explore the effects of partner status and other characteristics. This was done separately for men and women.

**Pooled Analysis**
Datasets from the three HSRC and two NCS surveys were combined using a set of core variables available in most of the surveys (see **Table 1.2**). Only respondents aged 16 to 55 years were used for these analyses, although those outside the age range were retained in the dataset since they were required for the correct calculation of standard errors adjusted for the survey design. Information from each survey regarding the strata and primary sampling units was recoded to ensure that these were unique across the datasets. The combined dataset was used to estimate regression models for:

- Trend in HIV prevalence across the surveys, accounting for changes in the survey sample (HSRC data only);
- Trend in the proportion of all men reporting more than one partner in the year before the survey.

## Limitations

**1. Differential participation.** The conclusions drawn from the results in this report rest on the assumption that the survey samples are representative of the population of South Africa, and are therefore free from selection, participation and response bias. Since the samples were drawn from a good quality national sample frame a noticeable selection bias seems unlikely. Participation bias is more difficult to detect or quantify. **There is evidence that participation was lower among certain population groups** (white, urban, wealthier) than among the poorer, rural African population. Levels of non-response were adjusted for in the survey weighting but this would not correct for any systematic differences in participation introduced by this pattern of non-response.

**2. Incomplete HIV test results.** In addition to concerns about the survey response rates, the validity of the results for HIV status estimates is undermined if HIV test non-response has introduced a bias. **Response to the HIV testing component was around 75% in the 2005 and 2008 HSRC surveys**. If the 25% who did not consent to HIV testing were systematically different to other respondents this could undermine the validity of the results presented here. The report on the HSRC 2008 survey found little evidence for a systematic difference in the characteristics of those who were interviewed only and those who were interviewed and tested for HIV. However, they **did not separate non-response to the HIV test component by reason** and it is possible that there may have been an effect once those who refused (the majority of non-responders) were separated from the other non-responders. This information is not in the dataset and could not be investigated here.

**3. Missing information.** A substantial number of respondents were missing information for a few variables each. This entailed the **exclusion of a substantial proportion of respondents**. This may have introduced a bias if those with incomplete data were systematically different to other respondents. For example, if respondents who have had a lot of partners had difficulty reporting accurately on all their partners and therefore did not give answers to some questions. Comparison of the characteristics of those with complete data for analysis and other respondents did not reveal any systematic differences, which suggests that most of the missing information was essentially missing at random. Therefore exclusion of the incomplete respondents should not have introduced a bias.

**4. Reduced power.** Exclusion of respondents with missing key data greatly reduced the power of the analysis since the multivariate analysis of the HSRC surveys was based on around two thirds of the total sample. It is therefore possible that some associations might not have been apparent in this reduced sample.

**5. Differences between the surveys (design, implementation, survey instruments).** All these differences undermined comparisons over time. Each of the five surveys had a different target population, sampling strategy and questionnaire. It is impossible to distinguish real changes over time from changes in response to the differences in the surveys. Where trends over time are consistent, then changes due to the survey methods, which should be non-directional, would seem a less likely explanation than a real change over time. There was an apparent trend over time for men reporting multiple partner in the last year but this was not significant. However it is possible that this trend was muted by the differences in the survey designs. Subsequent, comparable surveys would make it easier to detect trends.

**6. Complex survey designs.** The impact of the complex survey designs used, especially in HSRC 2008, have not been fully explored in the multivariate analysis. Stratification for this survey was on several variables known to be strongly associated with HIV prevalence: province, race, type of settlement. In order to fully understand the processes it would have been necessary to break down the individual sample weights to the components for each level of stratification and the information needed to do this was not available.

**7.** Correct standard errors could not be calculated for NCS 2006 because the PSU variable was missing from the dataset. This is needed to properly correct the standard errors for the clustering in the data.

**8.** Information on religion was available in the HSRC 2005 survey but could not be used in this analysis because it was stored as un-coded, open-ended responses. Socio-economic had been collected in the HSRC 2008 survey, but was not available for analysis.

**Table 1.2: Definitions of variables used in the analysis**

| Variable | varname | Categories | HSRC 2002 | HSRC 2005 | HSRC 2008 | NCS 2006 | NCS 2009 |
|---|---|---|---|---|---|---|---|
| **Age group** | agegrp | 15-19 to 60+ in 5-year groups | ✓ | ✓ | ✓ | ✓ | ✓ |
| **Birth cohort (5 year)** | birth_cohort | pre-1940, 1940-49, 1950-54 to 1990-1994 in 5-year groups | ✓ | ✓ | ✓ | ✓ | ✓ |
| **Birth cohort (10 year)** | bc10 | pre-1940, 1940-49, 1950-59 to1980-89 in 10-yr groups, 1990-94 | ✓ | ✓ | ✓ | ✓ | ✓ |
| **Education** | educ | None, Primary, Secondary, Tertiary | ✓ | ✓ | ✓ | ✓ | ✓ |
| **Marital status** | mstat | Single, Married, Unmarried cohabiting, Ex-married, missing | ✓ | ✓ | ✓ | ✓ | ✓ |
| **Race** | african | Other, African | ✓ | ✓ | ✓ | ✓ | ✓ |
| **Province** | province | | ✓ | ✓ | ✓ | ✓ | ✓ |
| | | | | | | | |
| **Radio** | radio | Never, once a week, 2-6 times a week ,every day | ✓ | ✓ | ✓ | | ✓ |
| **TV** | tv | Never, once a week, 2-6 times a week ,every day | ✓ | ✓ | ✓ | | ✓ |
| **Newspaper** | newspaper | Never, once a week, 2-6 times a week ,every day | ✓ | ✓ | ✓ | | ✓ |
| **Magazine** | magazine | Never, once a week, 2-6 times a week ,every day | ✓ | ✓ | ✓ | | ✓ |
| **Internet** | internet | Never, once a week, 2-6 times a week ,every day | ✓ | ✓ | ✓ | | ✓ |
| | | | | | | | |
| **Ever had sex** | sexever | No, yes, no response,missing | ✓ | ✓ | ✓ | ✓ | ✓ |
| **Sex in last year** | sexuallyactive | No, yes, no response, missing | ✓ | ✓ | ✓ | ✓ | ✓ |
| **Sex in last month** | active30 | No, yes, no response, missing | | ✓ | ✓ | | ✓ |
| **Lifetime partners** | lifetime_ptnrs | Total number of opposite sex partners in lifetime | | | ✓ | | |
| | lifesamesex | Total number of same sex partners in lifetime | | | ✓ | | |
| **>1 lifetime partner** | multilife | No, yes, missing | | | ✓ | | |
| **Age at first sex** | lt_agefirstsex | reported age at first sex, current age for virgins | ✓ | ✓ | ✓ | ✓ | ✓ |
| | afscat | Never had sex, <16, 16-20, 21+ | ✓ | ✓ | ✓ | ✓ | ✓ |
| **Time since first sex** | tfs | Time since first sex in years, 0 for virgins | ✓ | ✓ | ✓ | | |
| **Partners in last year** | num_ptnrs | Total number of opposite sex partners in last year | ✓ | ✓ | ✓ | ✓ | ✓ |
| **>1 partner in last year** | multiyear | No, yes, missing | | ✓ | ✓ | ✓ | ✓ |
| **Parners in last month** | ptnrs30 | Total number of partners in last month | | | ✓ | ✓ | ✓ |
| **>1 partner in last month** | multimonth | No, yes, missing | | | ✓ | ✓ | ✓ |
| **Partners at interview** | ptnrsnow | Total number of current partners | ✓ | ✓ | | | |
| **>1 partner at interview** | multinow | No, yes, missing | ✓ | ✓ | | | |
| **Types of partner in last year** | regular | No, yes, missing | | ✓ | ✓ | | ✓ |
| | nonreg | No, yes, missing | | ✓ | ✓ | | ✓ |
| | csw | No, yes, missing | | ✓ | ✓ | | ✓ |
| **Age mixing** | biggestgap | No partner, 0-4 years, 5-9 years, 10+ years | ✓ | ✓ | ✓ | | ✓ |
| **HIV testing** | hiv_test_when | Never, <1 year, 1-2 years, 2+ years, missing | | ✓ | ✓ | | |
| | hiv_test_ever | No, yes, missing | ✓ | ✓ | ✓ | | |

| | | | | | |
|---|---|---|---|---|---|
| **Reason for HIV test** | hiv_test_why | | ✓ | ✓ | ✓ |
| **Forced sex in last year** | forced_sex | No, yes, missing | ✓ | ✓ | ✓ |
| | | | | | |
| **Going without** | gw_shelter | Shelter: Never, Rarely, Sometimes, Often | | ✓ | |
| | gw_fuel | Fuel: Never, Rarely, Sometimes, Often | | ✓ | |
| | gw_water | Water: Never, Rarely, Sometimes, Often | | ✓ | |
| | gw_medicine | Medicines: Never, Rarely, Sometimes, Often | | ✓ | |
| | gw_food | Food: Never, Rarely, Sometimes, Often | | ✓ | |
| | gw_cash | Cash income: Never, Rarely, Sometimes, Often | | ✓ | |
| **Has disability** | disability | No, yes, missing | | | ✓ |
| **Has dependents** | dependents | No, yes, missing | | ✓ | ✓ |
| **Mobility** | mobileyear | Been away in last year | | ✓ | |
| | mobileweek | Been away 1+ nights in last week | | ✓ | |
| **Male circumcision** | circumcised | Men only: No, yes, missing | ✓ | | ✓ |
| **Circumcised before first sex** | circbeforesex | Men only: No, yes, unclear, missing | ✓ | | ✓ |
| **Partners' fidelity in last month** | punfaith | No, yes, don't know, no partner(s), missing | | | ✓ |

## Section 1C. Results from multivariate analysis

### Data available for multivariate analysis

**Table 1.3** describes the numbers of male and female respondents who were interviewed, interviewed and tested for HIV (HSRC only) and who had **sufficiently complete data for multivariate analysis** (i.e. who did not have missing responses on key variables). The low proportion of respondents from HSRC 2008 available for multivariate analysis was due largely to missing information on age at first sex and details of partners in the year prior to the survey. However, most variables had missing values for between 6% and 10% of respondents which HIV status, and many respondents were missing some information.

In HSRC 2008, information on sexual activity (sex ever or sex in the last year) was missing for 7% of the respondents who were both interviewed and tested for HIV. Age at first sex was missing for 14.5% of these respondents. Around a fifth of respondents were not tested for HIV and, as a result of excluding these respondents with missing values, multivariate analysis was conducted on just over half of the original sample. In HSRC 2005, a smaller proportion of respondents were missing information on sexual activity (3.5% and 4.5% missing sex ever and sex in the last year respectively), however, 16% of respondents were missing age at first sex. Thus just over 60% of respondents could be included in the multivariate analysis. For both HSRC 2005 and HSRC 2008 the HIV status of those excluded from analysis was compared with those included in the multivariate analysis. For both surveys there was no difference in HIV status between those without missing values and those excluded from the analysis due to missing data. The data from the HSRC 2002 survey was more complete: less than 1% of respondents were missing information on sexual activity and for only 5% was age at first sex not available. The NCS datasets were essentially complete and the few respondents with missing values were excluded from analysis.

**Table 1.3: Number of respondents interviewed, tested for HIV (HSRC only) and available for multivariate analysis, by survey and sex**

| Survey | | Number of respondents | Number interviewed & tested for HIV (% of total) | Number with complete data for analysis (% of total) | Number complete & aged 16-55 (% of total) |
|---|---|---|---|---|---|
| **HSRC 2002** | *Men* | 2,704 | 2,591 (95.8) | 2,016 (74.6) | 1,731 (64.0) |
| | *Women* | 3,667 | 3,497 (95.4) | 2,666 (72.7) | 2,194 (59.8) |
| **HSRC 2005** | *Men* | 6,338 | 4,529 (71.5) | 3,828 (60.4) | 3,146 (49.2) |
| | *Women* | 10,057 | 7, 503 (74.6) | 6,404 (63.7) | 5,140 (51.5) |
| **NCS 2006** | *Men* | 3,505 | n/a | 3,411 (97.3) | 3,137 (89.5) |
| | *Women* | 3,501 | n/a | 3,382 (96.6) | 3,104 (88.7) |
| **HSRC 2008** | *Men* | 5,501 | 4,229 (76.9) | 2,944 (53.5) | 2,424 (44.1) |
| | *Women* | 8,327 | 6,579 (79.0) | 4,751 (57.1) | 3,801 (45.6) |
| **NCS 2009** | *Men* | 4,437 | n/a | 4,419 (99.6) | 4,413 (99.5) |
| | *Women* | 5,291 | n/a | 5,266 (99.5) | 5,263 (99.5) |

### Characteristics of included respondents

Characteristics of respondents, by survey and sex, are shown in **Section 1F** using only those respondents available for the multivariate analysis and aged 16-55.

The sampling and survey instruments differed between the HSRC and NCS surveys, which might explain differences between the two types of survey. However, slightly different methods were used for each of the three HSRC surveys, and for the 2006 and 2009 NCS. Therefore comparisons were made across all five surveys, with the caveat that some differences may be due to survey design. At this stage, no formal statistical tests were conducted to see if differences between surveys were due to chance (see sections on trends over time).

*Socio-demographic characteristics*

Distribution by provinces has changed little over the three HSRC surveys and was broadly similar between the HSRC and NCS 2009 surveys. The HSRC 2008 sample has fewer respondents from rural/tribal areas and more from urban formal areas than the previous HSRC surveys. The NCS 2009 has more respondents from urban informal areas than the other surveys, and fewer from rural areas. At least three quarters of respondents in all surveys were African.

The surveys covered different age ranges but, once restricted to the range common to all surveys (16-55) the distribution did not change markedly between the surveys. Male respondents tended to be younger than female respondents (mean ages 34.5 and 36.8 respectively). The HSRC surveys had no upper age limit but the distribution of older respondents was similar across the surveys (not shown). The proportion reporting secondary education has increased between 2002 and 2009. The proportion of men and women reporting tertiary education in NCS 2009 (6% and 5% respectively) was around half of the level in the other surveys.

Comparing the five surveys, the proportion of respondents who were married had decreased over time and the proportion single had increased. The proportion of respondents cohabiting with a non-spousal partner had increased steadily; the proportion formerly married had not changed.

**Table 1.4** gives the number of respondents with complete HIV test data and the proportion HIV positive in each HSRC survey, by sex.

**Table 1.4: Respondents with complete HIV test data and HIV status (2002, 2005, 2008)**

| | MEN | | | WOMEN | | |
|---|---|---|---|---|---|---|
| **Variable** | **2002** | **2005** | **2008** | **2002** | **2005** | **2008** |
| N respondents with complete data, age 15+ | 2016 | 3828 | 2944 | 2666 | 6404 | 4751 |
| **HIV Status** | | | | | | |
| Positive, ages 15+ | 11.3 | 10.7 | 10.1 | 13.9 | 16.8 | 17.0 |
| Positive, ages 16-55 only | 12.2 | 12.4 | 11.3 | 16.1 | 19.6 | 20.0 |

## HIV Status Models (see tables in Section 1G)

### Regression model for HIV status: Women 2008

After adjusting for other factors there was still a strong association between race and HIV status.

The geographical variables, province and geotype, remained important but only certain categories were significantly different from the others. Living in an urban informal area was associated with higher odds of HIV infection compared to other types of area. Respondents living in KwaZulu Natal and Mpumalanga were more likely to be HIV infected than others.

In the same way, education was only discriminatory for women with tertiary education compared to others.

Married women had lower odds of HIV infection than other women and showed an interaction with birth cohort so a combined variable was created. The baseline for the combined variable is single women born 1980-89.

Unmarried women in the youngest cohort were less likely to be HIV positive women that unmarried women born 1980-89. The odds of infection for single women born in the 1960s and 1970s were no different to those of women born in the 1980s. Single women born before 1960 had lower odds of infection than single women born in the 1980s.

Married women born before 1980 had lower odds of infection than single women born in the 1980s, the odds of infection declined steadily for older cohorts. Very few women born in the 1990s and 1980s were married, but their HIV status did not differ from that of single women born in the 1980s. Therefore marriage appears to be negatively associated with HIV infection in the middle age range, where HIV prevalence peaks. The reduced odds found for the older cohorts (born before 1960) is probably a selection effect: women from these cohorts who became infected probably acquired HIV infection in the 1980s and 1990s and, without access to ART would have died before these surveys were fielded. Older women (born before 1960) are less likely to be infected regardless of marital status- because when they were younger HIV prevalence was much lower and any women from that cohort who were infected probably died before their fifties.  This assumes that the force of infection was greatest for these women when they were younger and has declined as the women have aged.

Reason for having an HIV test was associated with HIV, compared to those who had not had a test. By and large this shows what is expected- respondents who waited until they were ill to go and get a test are more likely to be positive.

HIV infection was overwhelming associated with sexual activity, and there was a slightly higher risk among those active in the last year, though this difference was not significant.

## Regression model for HIV status: Men 2008

Most of these factors from the bivariate analysis were not significant once adjusted for the other variables.

As for women, race (African or not) was the most important predictor of HIV status. Geotype was not important for men but most provinces were different to Western Cape, reflecting the relatively low prevalence in that province. Three provinces were not different: Northern Cape, Free State and Limpopo.

Higher odds of infection were seen for men born between 1960 and 1979, compared to men born 1980-89.  Marriage was not important and there was no interaction effect observed.

Men who used the internet regularly, compared to those who never did, were much less likely to be infected .

The only sexual behaviour variable to retain an effect once adjusted for other variables was sex with a CSW in the year before the survey, which trebled the odds of HIV infection, though the p-value was quite large.

Reason for having had an HIV test was important. Compared to respondents who had never had a test, men who had a test because they were feeling ill, were referred by a health worker or because a partner or child was ill, diagnosed with HIV or had died, had higher odds of infection. Men who cited curiousity as the main reason for their test were also more likely to be infected.

## Regression model for HIV status: Women 2005

Geographic variables, province, geotype and race (African or other) all retained a large and significant effect in the adjusted model. Four provinces (Free State, KZN, Gauteng & Mpumalanga) were associated with higher HIV prevalence compared to Western Cape. Being African had an adjusted OR for HIV of 8.7 (95% CI 5.1-14.8) compared to other races.

Geotype showed some effect, with a borderline protective association for rural formal areas compared to urban formal areas.

Women born in the 1970s were at increased likelihood of being HIV positive compared to those born later. Women born before 1960 were less likely to be HIV positive. There was no interaction between birth cohort and marriage in this model. Married and cohabiting women were less likely to be HIV positive than single women.

Respondents who said they ever went without cash were more likely to be HIV infected. Respondents who said they rarely went without shelter were less likely to be infected.

Women who had ever had sex were much more likely to be HIV positive than virgins; the risk was only slightly higher in those who had had sex in the year before the survey than in those who had not had sex within the last year. Women with more than one partner at the time of the survey, those who reported a non-regular partner in the last year and those who had a partner more than 10 years older all had higher odds of infection.

Women who reported having had an HIV test because they were sick, they were advised by a health worker to take a test or because a partner or child was sick, diagnosed with HIV or died were more likely to be infected than women who had never taken a test. Other test reasons were not associated with infection.

Watching television was protective and there was a borderline protective association with internet use.

## Regression model for HIV status: Men 2005

In 2005 province and race retained their strong effects.  Men born in the 1960s and 1970s were much more likely to be HIV positive than men born in the 1980s. Men born in the 1990s were less likely to be infected in 2005. Tertiary education, compared to no education, was protective.

Men who reported at least one partner who was more than a 10 years younger (or older) than them were more likely to be infected. (NB. the age gap can go either way but the vast majority of men's partners were younger, only a handful were older than the respondents.) This suggests that men who choose much younger partners are more likely to be HIV positive than men whose partners are more similar in age, but which comes first the HIV infection or the younger partner.

Having had an HIV test prompted by feeling sick, referral by a health worker or because a partner or child was ill, diagnosed with HIV or died, was associated with increased odds of infection. Tests for other reasons showed no association.

## Regression model for HIV status: Women 2002

Province was not important for women in 2002 but race and type of settlement were both associated with HIV. Women living in tribal areas were less likely to be infected than women in urban formal areas. In this model there was no significant difference between women in formal and informal urban areas but the OR was higher for informal urban areas, consistent with the models for 2005 and 2008.

The effect of age was also less marked in this model, only women born in the 1970s were at significantly higher odds of infection compared to those born in the 1980s. Marriage was protective compared to being single. However being sexually active was associated with much higher odds of infection, and having had sex in the last year was associated with even higher odds, which outweighed the effect of marriage. Therefore, as in all the models, sexually active respondents were much more likely to be HIV positive and this was only partially offset by marriage.

HIV test history was not significantly associated with HIV status, there was a borderline protective effect with taking a test for administrative purposes and the OR for the other reasons were in line with those in the 2005 and 2008 models.

Internet usage and reading newspapers appeared to be protective in 2002.

## Regression model for HIV status: Men 2002

Effects of province and race were evident for men in 2002 but not as marked as in the later surveys. Men born in the 1960s and 1970s were more likely to be infected than men born in the 1980s.

Men who reported a partner 10 or more years younger or older were much more likely to be HIV infected than men whose partner was of a similar age.

Men who were circumcised before they first had sex were much less likely to be HIV infected.

Men who had had an HIV test for administrative purposes were less likely to be infected than men who had never had a test. Men who had been tested because of illness were also less likely to be infected than those who had never tested, but this effect was borderline. If true the OR has changed dramatically between 2002 and 2008 from 0.25 to 4.9.

## HSRC 2008: Interactions by race and province

Race and province were strongly associated with HIV status. Both variables describe the local prevalence of HIV and, because there are such marked variations, may drown out the effects of individual level characteristics. Certain characteristics may be differently associated with HIV status depending on province of residence or race i.e. these variables may be effect modifiers. Interaction terms were added to the main models to explore these effects but the sample size was insufficient for the resultant models. Regression models with just two factors were suggestive of interactions between race & province, and province & birth cohort for men and between race and birth cohort for women.

## Trend over time in HIV prevalence (see table in Section 1H)

Results of logistic regression models to compare the HIV status of respondents in the three HSRC surveys are given in **Section 1H**).

For men, there was little evidence of a trend over time in HIV prevalence. Compared to 2008, the crude odds ratio for HIV positivity was 1.1 for both 2002 and 2005 (95% CI 0.8-1.5 and 0.8-1.5 for 2002 and 2005 respectively). For women there was no evidence for a difference between 2005 and 2008 (crude OR 0.98, 95% CI 0.8-1.2) but some evidence for lower HIV prevalence in the 2002 survey compared to 2008 (crude OR 0.77, 95% CI 0.61-0.97).

Differences between the surveys may be amplified, or disguised by, differences in the characteristics of survey respondents in the different years. Therefore adjusted logistic regression models were fitted to examine this.

Models included age group, educational attainment, province, type of settlement, marital status and race. Once adjusted for these factors, there was still no trend over time for men and the p-value for the crude association observed for women increased from 0.03 to 0.07. This suggests that the lower prevalence for women in 2002 compared to 2008 could be due to chance alone.

## Characteristics associated with reporting sex with a CSW in 2008 (see table in Section 1I, men only)

Since sex with a sex worker was the only sexual behaviour to show an association with HIV status in any of the models for men a separate model was constructed to look at the characteristics associated with this behaviour.

The adjusted OR for being HIV positive by whether or not the respondent had a CSW partner in the last year was 2.9 (95% CI 0.9-9.4). Although only 0.5% of men (24 respondents) reported this behaviour, HIV prevalence was 22% among those men. The population impact may be relatively small but this nonetheless suggests that around 15,000 HIV positive men visit sex workers each year.

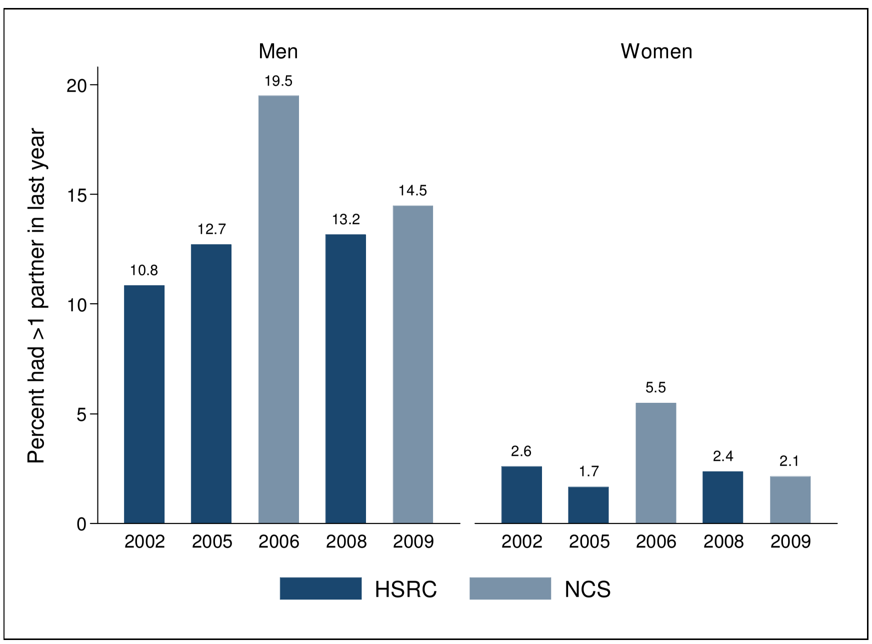
The logistic regression model is given in **Section 1I**). There was no association between CSW partners and place of residence, age, education or marital status. Men who had been sexually active recently were more likely to have had a CSW partner (adjusted OR 23.5), as were men who reported a non-regular partner in the last year (adjusted OR 10.1). HIV positive men were almost four times more likely to report CSW partners that HIV negative men (adjusted OR 3.8). Whilst the CSW partner may have been the source of HIV infection for some of the men, many will have been infected more than 12 months before the survey. Therefore HIV positive men continue to visit sex workers. Two factors were negatively associated with CSW partners: men who listened to the radio were less likely to have had a CSW partner than men who did not listen to the radio (adjusted OR 0.2) . Men who had ever been tested for HIV were also less likely to have had a CSW partner in the last year than men who had never been tested (adjusted OR 0.07).

## Trend over time in proportion of men reporting multiple partners in the last year (see table in Section 1J)

As **Figure 1.2** shows, there was some indication of an increase over time in the proportion of 16-55 year old men who reported more than one partner in the year before the survey. No such trend was apparent for women.  The 2006 NCS is an outlier for both men and women.

The crude OR for having had more than one partner in the year before the survey, comparing each year to 2008 were, or men, 0.8 (for 2002), 0.96 (for 2005), 1.6 (for 2006) and 1.1 (for 2009) .  Only the OR for 2006 was significantly different from 2008, 95% CI: 1.3-2.0.

**Figure 1.2: Proportion of 16-55 year olds reporting more than one partner in the year before the survey, by survey and sex.**




A logistic regression model was fitted for men to see if changes in respondent characteristics might explain the differences over time. The model included age group, educational attainment, province, type of settlement, marital status and race.  The results of this model are in **Section 1J**.

The difference between 2006 and 2008 remains once other factors are controlled for.  Education was not important and was not retained in the model. All other factors showed an independent association with the outcome but none appeared to confound the association between year of survey and multiple partners in the last year.

## Section 1D. Discussion

### HIV status at survey

The logistic regression models for HIV status, fitted separately for men and women in each of the HSRC surveys showed that the macro level variables (such as province, race and birth cohort) were more important than individual level factors. These variables describe the "when" and the "where" of infection and reflect the marked geographical differences in HIV prevalence and the sharp contrasts over time. This finding echoes that of an analysis of cohort data from Zimbabwe[2].

Some behavioural factors were associated with HIV status, and these are discussed more fully below.

#### Women

In the HIV status models for women, in each year, women who had been sexually active were more likely to be HIV positive. This effect is due to the low prevalence among those who have never had sex, reflecting that fact that most infections are sexually transmitted and therefore complete abstinence is very protective. However, it raises the question of the extent to which HIV positive women continue to have active sex lives, and with whom. The OR for having had sex in the last year, comparing HIV positive women to HIV negative was 2.6 in 2002 (95% CI 1.7-3.8), 2.0 in 2005 (95% CI 1.6-2.5) and 1.9 in 2008 (95% CI 1.5-2.5). After controlling for age, this effect disappears for 2005 and 2008 and the adjusted OR for 2002 decreases to 1.8 (95% CI 1.1-2.7). The OR for having had sex in the month before the survey, again comparing HIV positive to HIV negative women, was 1.5 in 2005 (95% CI 1.2-1.9) and 1.8 in 2008 (95% CI 1.4-2.3). This information was not available in HSRC 2002. Again, this effect disappeared once controlled for age.

Marriage is associated with lower odds of infection in each year; this may be because women who married, and were HIV negative when they got married, were subsequently less likely to acquire infection, or because women who are HIV positive are less likely to remain married. It is unfortunate that it is not possible to distinguish first and subsequent marriages in these data.

In 2005, women who had had more than one partner in the year before the survey were more likely to be HIV positive. If HIV positive women are those most likely to have multiple partners, this would be a cause for concern regarding onwards transmission of infection. Comparing all women, including those who were not sexually active or who had never had sex, in 2002, there was no association between HIV status and multiple partners in the last year. In 2005, the crude OR for having had multiple partners, comparing HIV positive women to HIV negative was 3.9 (95% CI 2.2-6.9) and in 2008 the crude OR was 2.3 (95% CI 1.2-4.4). Once adjusted for age, the association disappeared in 2008 but was remained in 2005: adjusted OR 3.0 (95% CI 1.7-5.4).

In all three surveys, HIV positive women were significantly more likely to be single compared to HIV negative women. The proportion of HIV positive women who were single was 60%, 61% and 67% in 2002, 2005 and 2008, respectively, compared 42%, 41% and 47% of HIV negative women. This difference is not explained by the differences in age distribution between HIV positive and HIV negative.

#### Men

For men, few behavioural variables were associated with HIV status. In HSRC 2002 and HSRC 2005, having had an age gap of 10 years or more with a partner in the last year was associated with increased odds of infection. In 2008, having had sex with a CSW in the year before the survey was associated with increased odds of infection.

### HIV status and sex with a sex worker: men

The logistic regression model, based on men in HSRC 2008, for sex with a CSW in the last year found that HIV status was associated with this outcome. Men who had had sex in the month before the survey, and men who had had a non-regular partner in the last year were also more likely to report sex with a CSW. In 2008, 74% of HIV positive men, versus 62% of HIV negative men (p-value=0.0001), had had

---

Lopman B et al. (2008). HIV incidence in 3 years of follow-up of a Zimbabwe cohort—1998–2000 to 2001–03: contributions of proximate and underlying determinants to transmission. Intl J Epidem, 37(1):88-105 sex in the last year and 68% of HIV positive and 57% of HIV negative had had sex in the last month (p-value 0.03). These differences are explained by the difference in the age distribution of HIV positive and HIV negative men, and disappear once age is controlled for. However, this reinforces the point that HIV positive men are in the age group that reports the most sexual activity, and HIV infection does not appear to limit their activity. This is undoubtedly because most men are not aware of their status and because early infection is rarely symptomatic.

Having had an HIV test was negatively associated with having sex with a sex worker. This could be because men who are concerned about HIV infection avoid CSW partners and take HIV tests, or because the experience of HIV testing discourages men from visiting sex workers. There is not way to date the HIV test relative to the CSW partner, or to ascertain whether men who did not report CSW partners but did report HIV tests had ever visited CSW at some point in their lives.

Although, the number of men reporting sex with a CSW was small in 2008, and the proportion of the population equally small, the population impact could still be noticable.

In addition, the reversal of the proportions reporting non-regular partners and CSW partners between 2005 and 2008 suggests that there may have been some social desirability bias operating. If the lay person's definition of a CSW is quite fluid, or if it is undesirable to admit to having such a partner, perhaps men in 2008 were more inclined to report such partners as non-regular than the respondents in 2005. If so, the proportion of men with this behaviour might be higher than the 2008 survey suggests.

#### HIV trend over time

The logistic regression models for HIV status at survey, based on the data from the 2002, 2005 and 2008 HSRC surveys, show that the odds of being HIV positive among survey respondents have not changed over the surveys. This is also reflected in the age-specific prevalence graph (see section XXX main report) which shows an increase in the age of those infected over time, but also wide confidence intervals which show that any changes within individual age groups could be due to chance.

#### Multiple partners trend

Comparison of the results from the individual surveys suggests that the proportion of men reporting more than one partner in the year before the survey had been increasing between 2002 and 2009. However, results from logistic regression did not support this observation; no trend was evident over time when geographic variables (province, geotype & race) and age and marital status were controlled for. The 2006 NCS survey was different to the others; there is no obvious reason why this should be so.

In the report on the 2008 HSRC survey there was a trend apparent between 2002 and 2008 in the proportion of both men and women who reported more than one partner in the last year. There were some differences between that analysis and this; the analysis in the HSRC report was based only on the HSRC data whereas NCS data were included in this analysis. However, there was no trend evident when this analysis was restricted to HSRC surveys only. In the HSRC report, the analysis included only those aged 15-49 who had had sex in the year before the survey. They also included respondents who were not tested for HIV. The analysis presented here included all respondents aged 16-55, and, in the HSRC surveys, tested for HIV (with appropriate weights).

The increase over time, in the proportion of men reporting multiple partners, and the apparent sensitivity of the results to the analytical approach, suggest that no evidence for a trend should not be taken as evidence of no trend. The differences, and some deficiencies, in the behavioural data from the different surveys make rigorous analysis difficult. The only survey with a partner history is the NCS 2009. Partner histories permit a more detailed analysis of behaviour because, unlike direct questions on a behaviour of interest, they can be used to build up a profile of a respondent's behaviour during the period before the survey.

## Section 1E. Recommendations

Comparison of behaviour between subsequent surveys and the existing data would be facilitated by the inclusion of a partner history. This would permit backwards compatible analysis since the measures used to date could be derived from the partner history but it would also be possible to describe individual behaviour more fully.

Multi-level modelling of HIV status might better describe the processes operating at geographic and social levels: their relative importance and their influence on individual level factors.

To do this would require detailed information on the survey designs in order to disaggregate the survey weights and assign each component to the correct level in the model.

The information on partner acquisition and loss in the 2009 NCS could be analysed further. Further differentiating partner status by cohabitation would help explain the associations seen in this analysis. Loss rates, and concurrency start rates (partner acquisition by those who already have a partner) could be investigated. A multi-level model would permit comparison of the factors associated with acquiring a partner and those associated with acquiring an additional partner. It is important to understand whether these two processes are different in order to target HIV prevention efforts and to inform the debate on the importance of concurrent partnership in HIV transmission.

Macro level variables, such as race and location of residence, were more important than individual characteristics in describing HIV status of the survey respondents. There is some evidence that HIV positive men may be slightly more likely to report recent risky sexual behaviour than HIV negative men.

There was little statistical evidence for trends over time in either HIV status or multiple partnerships. However, the results of both analyses are sensitive to the analytical approach used. Changes are apparent in the crude results, and may become statistically significant as more data are collected. Comparisons over time were undermined by different survey methods in each year.

The importance of macro level variables such as province and race suggests that multi-level analysis of these data could elucidate some of the associations seen in this analysis.