Full Title:

Comparing approaches to collecting self-reported data on HIV status in population-based surveys

Running Head:

Self-reported HIV Status in Population Surveys

Authors:

Sarah BRUMFIELD[1], Albert DUBE[2], Georges RENIERS[3], Stephane HELLERINGER[1]

1. Johns Hopkins School of Public Health, Department of Population, Family, and Reproductive Health, Baltimore, MD
2. Malawi Epidemiology and Intervention Research Unit, Lilongwe, Malawi
3. London School of Hygiene & Tropical Medicine, Department of Population Health, London, UK

Correspondence to:
Sarah Brumfield
Department of Population, Family, and Reproductive Health
Johns Hopkins School of Public Health
Baltimore, MD USA
sbrumfi2@Jh.edu

Conflict of Interest and Source of Funding:
The authors have no conflicts of interest to declare.
This project was supported by grant R21AI127186 (PI: Helleringer) from the National Institute of Allergy and Infectious Diseases, and by grant R01HD088516 (PI: Helleringer) from the Eunice Kennedy Shriver National Institute of Child Health and Human Development.

Word count:
1783/1500

Key Words:

HIV status, sub-Saharan Africa, surveys and questionnaires, self-report

Introduction

Self-reported data on HIV status collected from respondents during population-based surveys are useful to gauge progress towards the UNAIDS 90-90-90 goals [1]. In combination with HIV testing [2], such data allow estimating the proportion of persons living with HIV who are aware of their infection (i.e., the first 90). However, self-reported data on HIV status may be subject to missing values and/or misreporting [3]: respondents may refuse to answer questions about their HIV test results or provide inaccurate answers. This might happen if they feel such questions are intrusive, or if they perceive stigma associated with the disclosure of HIV test results to a data collector. Errors in self-reported data on HIV status might particularly affect survey estimates of progress towards the first 90 if they are correlated with respondents' awareness of their HIV status. This might be the case, for example, if persons living with HIV who have been diagnosed are less likely to answer questions about their HIV status.

Surveys have used different strategies for eliciting self-reported HIV status. Some have inquired indirectly [4], by first asking respondents if they were "willing to share the results of [their] last HIV test". If respondents answered "no" to this preliminary question, then they were not asked about the results of the most recent test. Other surveys [5] have asked directly for the results of the respondent's most recent test without including the preliminary question. It is unclear which approach is preferable. The indirect approach explicitly gives respondents an opportunity to refuse to answer, thus possibly leading to additional missing data on HIV status. Since other survey questions rarely include such a preamble, the indirect approach might also signal to respondents that the upcoming question is sensitive. This might negatively affect respondents' willingness to disclose test results accurately. On the other hand, not including the

preliminary question might lead respondents to answer the question about HIV test results even though they would have preferred not to do so (e.g., due to perceived interviewer pressure).

Here, our main goal is to test whether asking indirectly about HIV test results increases the extent of missing data. In a subset of respondents for whom reference data on HIV status are available, we explore potential differences in accuracy between the direct and indirect approaches to eliciting self-reported survey data on HIV status.

Methods

During a study of adult mortality in the Karonga Health and Demographic Surveillance System (KHDSS) in northern Malawi [6, 7], we conducted a nested methodological experiment comparing strategies to collect data on self-reported HIV status. The KHDSS covers a predominantly rural population, where HIV testing and linkages to care and treatment have been greatly expanded since the introduction of antiretroviral treatment in the area in 2005 [8].

We randomly selected a sample of 613 residents aged 15 to 59 years old of the KHDSS. We then randomized selected individuals 1:1 to either a face-to-face (FTF) interview or an audio computer-assisted self-interview (ACASI). ACASI is a more confidential mode of interview, which might reduce the extent of social desirability bias in the collection of survey data on sensitive topics [9, 10]. Randomization of ACASI vs. FTF was stratified by gender of the respondent and composition of her family (i.e., whether one of her siblings had recently died at adult ages). We then randomized half of the sample to the indirect approach to asking about HIV testing results, whereas the other half was assigned to the direct approach. Randomization of the indirect vs. direct approaches was stratified by gender of the respondent and assigned mode of interview (FTF vs. ACASI). All randomized assignments were obtained using computer-

generated random number sequences in Stata 15. They were then pre-loaded into our data collection platform (Open Data Kit). Study interviewers did not become aware of a respondent's assignment to the direct or indirect approach until they reached this section of the questionnaire.

No respondents were switched from the direct to the indirect strategy or vice versa. However, 14 respondents were switched from FTF to ACASI, and 20 were switched from ACASI to FTF due to a computing error and reported vision problems (ACASI requires the ability to see the tablet's screen).

In this analysis, we included every respondent who reported having ever tested for HIV and who stated that they received the results of their most recent test. We created a binary variable taking value 1 if the respondent had provided his/her test results and taking value 0 otherwise. We then tested whether the proportion of respondents with missing data on HIV test results varied by interviewing approach (direct vs. indirect) and interview mode (ACASI vs. FTF). To explore the accuracy of self-reported data on HIV status, we compared HIV test results reported in each study group to reference data on HIV status. These latter data were collected by the KHDSS, which has conducted several serosurveys and where individual records are routinely linked to clinical registers in health facilities [12, 13]. Reference data were however only available for a subset of respondents. In that group, we calculated the sensitivity of survey data on HIV test results. We cross-tabulated the HIV status of respondents according to the KHDSS dataset and the self-reports of HIV test results they made during the survey.

All analyses used an intent-to-treat approach, i.e., all respondents were analyzed according to their assigned study groups. Due to small sample sizes, we used Fisher's exact test to evaluate group differences in the proportion of respondents with missing data.

Results

Five hundred and thirty-five participants completed the survey (535/613, 87.3%). Among them, 35 reported never being tested for HIV and 25 reported not receiving the results of their most recent test. Among the 475 respondents who reported receiving their most recent HIV test results, 222 were assigned to ACASI and 253 to FTF. In the ACASI group, 119 were assigned to indirect approach vs. 103 to the direct approach. In the FTF group, these figures were 121 and 132, respectively. The study enrollment process is described in Supplemental Figure 1, whereas the demographic characteristics of respondents in each study group are presented in Supplemental Figure 2.

Among the respondents interviewed using the indirect approach, 5.8% (14/240) had missing data on their most recent HIV test results (figure 1). Only 1 of the 235 respondents interviewed using the direct approach (0.4%) had missing information on HIV test results ($p = 0.001$). However, the magnitude of these differences varied between FTF and ACASI interviews. In the ACASI group, 13 of the 119 respondents (10.9%) interviewed using the indirect approach had missing data on HIV test results, compared to 0 of 103 (0%) among those interviewed using the direct approach ($p < 0.001$). In the FTF group, there were no differences in the extent of missing data between those interviewed using the direct and indirect approaches ($p = 0.729$).

There were 67 respondents who were HIV-positive according to data collected by the KHDSS. Among those, 64 (95.5%) self-reported a positive test result in our survey. The other 3 respondents self-reported that their most recent HIV test results were negative. All 3 of these false negative reports were collected from respondents assigned to the direct approach.

Discussion

In our study, the indirect and direct approaches to collecting self-reported HIV status data generated similar levels of missingness in FTF interviews. In ACASI however, missing data on HIV test results were much more common when collected using the indirect approach. Missing data might lead to lower precision in estimates of the undiagnosed proportion of persons living with HIV. It might also lead to bias in such estimates if data on HIV test results are not missing completely at random, for example, if respondents with positive test results are more likely to refuse to answer questions about their HIV status when they are interviewed indirectly [11]. Our sample size was however too small to assess the correlates of missing data on HIV status among respondents interviewed using the indirect approach.

It is not clear from these results that the direct approach to collecting self-reported data on HIV test results should be preferred in population-based surveys, for two reasons. First, the direct approach might lead some respondents to answer a question that they would have preferred not to answer. This is suggested by the ACASI arm, where refusals to provide recent test results were more common when respondents were interviewed indirectly. ACASI limits interactions between respondents and interviewers, thus possibly enhancing the agency perceived by respondents in refusing to answer survey questions about sensitive topics such as HIV test results. In FTF interviews, there were no differences in missing data between direct and indirect approaches. This might be because a) interviewers skip the preliminary question in indirect interviews to save time or because they assume that respondents would agree to give their test results, b) interviewers successfully convince initially reluctant respondents to report their test results, and/or c) respondents perceive limited agency in refusing to report their HIV status when asked by an interviewer. Eliciting participants' reasons for sharing or choosing not to share HIV test results would require follow-up in-depth interviews with a subset of respondents who

refused/did not refuse to provide their test results. Such information may be particularly useful in designing future surveys.

Second, the direct approach might also result in less accurate data, e.g., if respondents who reluctantly answer are more likely to misreport their HIV test results. In our study, we identified 3 respondents who reported negative HIV test results in contradiction to reference KHDSS data. These false negatives were all interviewed using the direct approach. Due to limited sample size, however, we were not able to formally test for differences in the likelihood of misreporting HIV status between the direct and indirect approaches.

Another limitation stems from the longstanding presence of the KHDSS in this study population, including as a provider of home-based HIV testing [13]. Refusals to answer questions about HIV status may be less common in this area because respondents might be more willing to share test results with a trusted organization. Our results may thus not be generalizable to other settings, where HIV testing is more commonly accessed through health facilities and other organizations.

Because of the limited sample size and ability to assess the accuracy of self-reported data, we are not able to identify which interview method is more appropriate overall. However, our study indicates that approach to collecting self-reported information on HIV status may affect the quality of data elicited from participants. Therefore, additional studies are needed to identify the optimal interviewing approach, which minimizes missing data and misclassifications of HIV status while also providing respondents opportunities to decline to answer possibly intrusive questions. Such studies should be conducted in settings characterized by varying levels of HIV testing and HIV-related stigma in the target population.

Acknowledgements

S.H., A.D., and S.B. conceived of the study and contributed to the design of the data collection. A.D. and oversaw data collection. S.B. performed the data analysis, with input from S.H. and G.R. The manuscript was written by S.B. All authors have read and approved the submitted version of the manuscript.

References

1. Johnston LG, Sabin ML, Prybylski D, Sabin K, McFarland W, Baral S, et al. **The importance of assessing self-reported HIV status in bio-behavioural surveys**. *Bull World Health Organ* 2016; 94(8):605-612.

2. Mishra V, Barrere B, Hong R, Khan S. **Evaluation of bias in HIV seroprevalence estimates from national household surveys**. *Sex Transm Infect* 2008; 84 Suppl 1:i63-i70.

3. Fishel JD, Barrère B, Kishor S. **Validity of data on self-reported HIV status and implications for measurement of ARV coverage in Malawi**. 2012.

4. De Cock KM, Rutherford GW, Akhwale W. **Kenya AIDS indicator survey 2012**. *JAIDS Journal of Acquired Immune Deficiency Syndromes* 2014; 66:S1-S2.

5. Justman JE, Mugurungi O, El-Sadr WM. **HIV Population Surveys - Bringing Precision to the Global Response**. *N Engl J Med* 2018; 378(20):1859-1861.

6. Reniers G, Wamukoya M, Urassa M, Nyaguara A, Nakiyingi-Miiro J, Lutalo T, et al. **Data Resource Profile: Network for Analysing Longitudinal Population-based HIV/AIDS data on Africa (ALPHA Network)**. *Int J Epidemiol* 2016; 45(1):83-93.

7. Crampin AC, Dube A, Mboma S, Price A, Chihana M, Jahn A, et al. **Profile: the Karonga Health and Demographic Surveillance System**. *Int J Epidemiol* 2012; 41(3):676-685.

8. Jahn A, Floyd S, Crampin AC, Mwaungulu F, Mvula H, Munthali F, et al. **Population-level effect of HIV on adult mortality and early evidence of reversal after introduction of antiretroviral therapy in Malawi**. *Lancet* 2008; 371(9624):1603-1611.

9. Phillips AE, Gomez GB, Boily MC, Garnett GP. **A systematic review and meta-analysis of quantitative interviewing tools to investigate self-reported HIV and STI associated behaviours in low- and middle-income countries**. *Int J Epidemiol* 2010; 39(6):1541-1555.

10. Mensch BS, Hewett PC, Gregory R, Helleringer S. **Sexual behavior and STI/HIV status among adolescents in rural Malawi: an evaluation of the effect of interview mode on reporting**. *Stud Fam Plann* 2008; 39(4):321-334.

11. Altman DG, Bland JM. **Missing data**. *Bmj* 2007; 334(7590):424-424.

12. Koole O, Houben RM, Mzembe T, Van Boeckel TP, Kayange M, Jahn A, et al. **Improved retention of patients starting antiretroviral treatment in Karonga District, northern**

Malawi, 2005-2012. *J Acquir Immune Defic Syndr* 2014; 67(1):e27-e33.

13. Molesworth AM, Ndhlovu R, Banda E, Saul J, Ngwira B, Glynn JR, et al. **High accuracy of home-based community rapid HIV testing in rural Malawi**. *Journal of acquired immune deficiency syndromes (1999)* 2010; 55(5):625.

Figure Captions

Figure 1. Proportion with missing data on self-reported HIV test results, by interviewing strategy and mode of interview.
* indicates a statistically significant difference by Fisher's exact test.

Figure 1

Click here to access/download;Figure;figure1.png