

# The problem of nonresponse in population surveys on the topic of HIV and sexuality: A comparative study

---

Lisa Wallander<sup>1</sup>, Ronny H Tikkanen<sup>2</sup>, Louise Nilunger Mannheimer<sup>3</sup>, Per-Olof Östergren<sup>4</sup>, Lars Plantin<sup>5</sup>

<sup>1</sup> PhD, Department of Social Work, Malmö University, Malmö, Sweden

<sup>2</sup> PhD, Department of Social Work, University of Gothenburg, Göteborg, Sweden

<sup>3</sup> PhD, Department of Learning, Informatics, Management and Ethics, Karolinska Institutet, Stockholm, Sweden

<sup>4</sup> Professor, Division of Social Medicine and Global Health, Malmö University Hospital, Lund University, Malmö, Sweden

<sup>5</sup> Professor, Department of Social Work, Malmö University, Malmö, Sweden

Correspondence: Lisa Wallander, Gemensamt verksamhetsstöd (40), Malmö högskola, 205 06 Malmö, Sweden. Tel: + 46 40 665 75 88, e-mail: [lisa.wallander@mah.se](mailto:lisa.wallander@mah.se).

Funding: This project was supported by the Swedish Institute for Communicable Disease Control (grant codes 359/2010 and 01-12380/2008), Stockholm, Sweden.

## **Abstract**

**Background:** Throughout the western world, response rates are falling in population surveys. The aim of the study was to explore for the presence of nonresponse bias in two Swedish research projects on HIV (“HIV in Sweden” [HIS11]) and sexuality (“Sex and health” [UngKAB09]). **Methods:** The study used four data sets (two from each project), each of which had been generated using different methods. By comparing means and percentages on 15 items across the data sets, we explore the potential presence of nonresponse bias due to interest in the survey topic (topic salience), and discuss the suitability of two increasingly common methods for recruiting respondents: the pre-recruited probability web panel and the self-selected web survey. **Results:** While a higher proportion of the respondents in the HIS11 sub-studies had been tested for HIV and were inclined to perceive themselves as being at high risk of HIV infection, the respondents in the UngKAB09 sub-studies were on average more sexually active. Further, while there was little variation in the results between the HIS11 sub-studies (postal/web survey and pre-recruited web panel), there were some fairly large differences in the results between the UngKAB09 sub-studies (web surveys, one of which was based on a self-selected sample). **Conclusion:** The study concludes that 1) there are signs of nonresponse bias that may be due to topic salience, 2) while care must be taken when using self-selection sampling methods, the pre-recruited probability web panel might provide a cost- and time-effective alternative for recruiting respondents in future population surveys.

## **Keywords**

HIV, population survey, nonresponse

## **Introduction**

As part of the measures introduced to tackle the emerging AIDS epidemic in Sweden in the mid-1980s, a study was initiated in 1986 on the public's knowledge, attitudes and behaviour in relation to HIV ("HIV in Sweden" [HIS]).<sup>1</sup> This population survey was based on a random national sample of 3,900 individuals living in Sweden. The study has thereafter been repeated every third or fourth year, providing data that can be used as a basis for political decision-making in the area of HIV prevention, and for analysing possible changes in HIV-related attitudes and behaviour over time.

The HIS surveys conducted in the 1980s had response rates of around 70 % – rates which are quite acceptable. Over subsequent decades, however, response rates declined, and the 2007 study included only 50 % of the original sample of 6,000 respondents.<sup>2</sup> Four years later, the response rate was as low as 25 %.<sup>3</sup> A similar trend has been experienced by other repeated sexuality surveys in the Nordic countries.<sup>4,5,6</sup> The problem of low response rates is not unique to sexuality studies however. The recent steep decline in response rates has been observed by researchers from many different disciplines and countries,<sup>7,8,9</sup> and has resulted in a range of studies focused on specifying the nature of, and identifying possible ways of countering, this new "nonresponse challenge".<sup>10, 11</sup>

Although most researchers would agree that low response rates pose new challenges to survey research, the magnitude of the problem is a matter of dispute. While some would argue that we have come to "the end of an era" for population surveys, others maintain that no superior alternative to the population survey is currently available, and that we must learn to live with low response rates and to deal with the complex issues associated with them.<sup>12</sup> One such issue is that there is no necessary relationship between the rate of nonresponse and the degree of nonresponse bias.<sup>13</sup> In fact, the potential bias resulting from a certain rate of nonresponse is likely to vary between studies. Secondly, because the degree of bias is

dependent on the relationship between the likelihood of response and the variables that are analysed, there may be more variation in bias between the items within a survey than between surveys.<sup>14</sup> Thirdly, although nonresponse can lead to bias in many important statistics,<sup>15</sup> the largest threat is that of bias in the estimated means, percentages and totals – statistics that are pivotal in surveys focused on public health and prevention.

There are currently two well-established approaches to dealing with low response rates. The first involves limiting nonresponse, and includes various techniques used to increase respondent participation during the data collection (e.g. advance information letters, incentives, multimodal data collection methods, reminders).<sup>16,17</sup> The second approach considers the employment of auxiliary information about the non-respondents (e.g. administrative records or paradata) to assess the degree of nonresponse bias and to reduce potential bias of this kind by using response propensity weights in the statistical analyses.<sup>18,19</sup>

The present article is not directly linked with either of these two approaches, but presents a somewhat different contribution to the discussion of nonresponse and its implications, by systematically comparing a number of estimates of the respondents' knowledge, attitudes and behaviour in relation to HIV and sexuality from four data sets that were collected and investigated using different sampling and data collection methods. The study proceeds from the most recent (2011) "HIV in Sweden" survey [HIS11], and compares some of the results from this project with those of another research project on sexuality and health [UngKAB09],<sup>20</sup> which was conducted using samples from largely the same population less than two years earlier. Although response rates were low for both research projects (around 25 %), the projects share the strength of being based on two samples from the same population that were recruited and investigated using different methods. While the first sub-study in both projects was based on a traditional postal and/or web survey of a probability sample from the general population, the second HIS11 sub-study was based on a probability

sample from a pre-recruited web panel register, which aimed to be representative, while UngKAB09 employed a web survey administered at online communities, thus involving a self-selected non-representative sample of respondents. For both projects, the principal reasons for using multiple samples and methods were to provide the opportunity to cross-validate findings using two data sets, and to explore new and alternative approaches to data collection on the subject of HIV and sexuality.

The study presented in this article aims to explore for the presence of nonresponse bias in the data generated in the HIS11 and UngKAB09 research projects by comparing estimates of the respondents' knowledge, attitudes and behaviour concerning HIV and sexuality. Since none of the data sets can be used as a “gold standard”, against which the remaining sets of data might be assessed, we employ other rules of comparison. First, by contrasting the results from the two research projects (each represented by two sub-studies), we investigate the potential presence of one of the most common forms of nonresponse bias, i.e. bias due to certain individuals being more inclined to respond to surveys that cover topics in which they have a particular interest (topic salience) – in this case HIV (HIS11) and sexuality and health (UngKAB09).<sup>21</sup> Second, by comparing the results from the two sub-studies within each project, we are able to discuss, from the perspective of nonresponse bias, the suitability of two respondent-recruitment methods that are increasingly being used in population surveys, namely the pre-recruited probability web panel and the self-selected web survey.

## **Methods**

The article is based on data from two large research projects that were conducted in Sweden at around the same time.<sup>3,20</sup> As mentioned above, the projects employed two samples each, recruited from the same population. The HIS11 population survey included all individuals aged 16 to 49 living in Sweden in 2011, while the UngKAB09 population included all youths

and young adults aged 15 to 29 living in Sweden in 2009. For the sake of comparison, the analyses presented below are based on reduced data sets that include all respondents aged between 16 and 29.

The HIS11 survey was based on two sub-studies: (1) a multimodal post- and web survey of an age-stratified random national sample selected via the SPAR register (a national register of the Swedish population) and, (2) a pre-recruited web panel survey based on a stratified random national sample drawn from an existing web panel register owned by the private data collection company Norstat.<sup>22</sup> In the first of these sub-studies, only 25 % of the original sample of 6,000 individuals responded to the survey. The respondent sample for the second sub-study consisted of 1,500 web panellists from a large register comprising a random sample of the Swedish population. The panellists in the register were initially recruited by telephone interviews (with a response rate of 69 %). They receive incentives for study participation (e.g. lottery tickets) and stay in the register for an average time of 18 months. The panel for this study was drawn by means of age-stratified random sampling, and also reflected the proportions of individuals in the Swedish population on the variables sex and region. In this second HIS11 sub-study, the response rate was 23 %.

The two sub-studies included in the UngKAB09 survey were: (3) a web survey of an age-stratified random national sample recruited via the SPAR register and, (4) a self-selected web survey. The first sub-study included a large sample of 24,000 individuals, of which 24 % responded to the survey. Respondents for the second sub-study were recruited via various types of internet communities. These websites were selected by a panel of young people who identified a number of internet meeting places that they judged to be “representative sites” for youths and young adults. Before releasing the study on the sites, it was carefully marketed. Following a data collection period of two months, approximately 10,000 young people had responded to the questionnaire.

For an overview of the basic characteristics of the two projects, with their respective sub-studies, see Table 1.

--- Table 1 about here ---

In the results section below, we present results from univariate analyses of the items that were identical across the two questionnaires. Although we present statistics on a number of socio-demographic respondent characteristics, our main focus is directed at the 15 items associated with HIV and sexuality. The results section does not include any significance tests (since our aggregated sample of 16,410 individuals would mean that even the smallest difference would be statistically significant). Instead we focus on the broader patterns of similarities and differences in percentages and means across the sub-studies.

## **Results**

Below are three tables presenting results from univariate analyses of the respondents' socio-demographic characteristics and their knowledge, attitudes and behaviour related to HIV and sexuality.

### *Comparison of socio-demographic respondent characteristics*

Table 2 presents socio-demographic respondent characteristics for the four sub-studies.

--- Table 2 about here ---

Comparing the socio-demographic composition of the four samples, we find differences in the age distribution across the data sets, and the two HIS11 sub-studies include a smaller

proportion of male respondents than the UngKAB09 sub-studies (35 % compared to 44 – 51 %). Since the variation in response rates across the age groups within each study was small (see Table 1), the differences between the sub-studies in the age distribution is almost fully explained by the fact that the probability samples were stratified according to age, and that the HIS11 and the UngKAB09 studies used different stratification principles (see Table 1). Due to the variation in the age distribution among the respondents in the four samples, all the figures presented in subsequent tables have been weighted in accordance with age.

*Comparison of respondents' knowledge and attitudes related to HIV and sexuality*

Table 3 presents the respondents' knowledge and attitudes related to HIV and sexuality.

--- Table 3 about here ---

As regards the knowledge items, the numbers are similar across the four sub-studies. If we look at the two items on which there is some small variation, the results are slightly contradictory. While a larger proportion of the UngKAB09 respondents correctly recognize the benefits of using a condom to prevent HIV transmission (94 – 96 % compared to 90 – 92 %), a larger proportion of the UngKAB09 respondents also believe that HIV can be transmitted by kissing (19 % compared to 15 – 16 %). Looking to the attitude items, two results are particularly noteworthy. First, by comparison with the UngKAB09 respondents, the HIS11 participants agree somewhat more often with the statement that sexual intercourse should only occur in stable relationships (mean = 3.9 – 4.0 compared to 4.5 – 4.6). Second, a larger proportion of the HIS11 respondents perceive themselves to be at risk of becoming infected with HIV (3 % compared to 1 %) or chlamydia (8 – 11 % compared to 4 – 6 %).

### *Comparison of respondents' behaviour related to HIV and sexuality*

Table 4 presents the respondents' HIV- and sexuality-related behaviour.

--- Table 4 about here ---

Table 4 shows that, on the whole, the respondents in the UngKAB09 sub-studies are more sexually active than the HIS11 respondents. Among the former, we observe a higher average number of sex partners during the past year (mean = 2.2 – 3.1 compared to 1.3 – 1.6), a larger proportion of individuals with more than 5 sexual partners during the same period (7 – 14 % compared to 3 – 4 %) and a larger proportion of individuals who experienced sex on their first date (26 – 41 % compared to 17 – 20 %). Further, if we compare the results from the two UngKAB09 sub-studies, we see that the self-selected respondents are far more sexually active than those selected by probability sampling. Interestingly, although the UngKAB09 respondents are more sexually active than the HIS11 respondents overall, they are on average equally inclined to use a condom during sexual intercourse (23 – 25 %). As regards the questions about using the internet for flirting or to search for love, the prevalence of these activities was much higher among the self-selected respondents who had accessed the questionnaire at online communities than it was among the respondents in the three sub-studies based on probability sampling (64 % compared to 36 – 39 %; 47 % compared to 22 %).

### **Discussion**

The study's aim was to explore for the presence of nonresponse bias in the data from two Swedish research projects on the topics of HIV and sexuality (HIS11 and UngKAB09). While all the results are based on a descriptive comparison of estimates from four data sets, none of which can be used as a "gold standard", a number of results nonetheless support the

contention that salience bias, i.e. bias due to some individuals being more inclined to respond to surveys on topics in which they have a particular interest, may be present in the data to some extent. For example, while a larger proportion of the respondents in the “HIV in Sweden” surveys (HIS11) had been tested for HIV during the past 12 months, and were inclined to perceive themselves as being at high risk of becoming infected with either HIV or chlamydia, the respondents in the “Sex and health” surveys (UngKAB09) were more sexually active, in terms of both the average number of partners during the past 12 months and having experienced sex on the first date. In addition, the UngKAB09 respondents disagreed more often with the statement that sexual intercourse should only occur in stable relationships.

Further support for the view that topic salience plays a role in explaining survey participation is provided by results from the HIS11 nonresponse study, where 20 % of a random sample of 450 non-respondents reported having declined to participate due to a lack of interest in the topic (the nonresponse study was conducted via telephone interviews and included questions about socio-demographic characteristics and reasons for not having participated in the survey). However, it is also worth noting that for the remaining non-respondents, factors not associated with the survey topic – such as lack of time or energy – constituted the reasons given for not responding. Nonetheless, if both the HIS11 and the UngKAB09 projects suffer to some extent from a systematic non-participation of individuals who are less interested in the survey topics, another tentative conclusion from the current study would be that careful assessment is needed of which items may and may not be affected by nonresponse bias as a consequence of this. For example, although we need to be careful in drawing overly far-reaching conclusions regarding the validity of the estimates that were similar across the sub-studies, it is a fact that several of the estimates for the items relating to the respondents’ knowledge about the transmission of HIV were almost identical across the four data sets.

One potential future strategy for avoiding nonresponse due to a lack of interest in the topic of HIV might be to start integrating this specialized survey into more general public health surveys. Such a strategy might also counteract the potentially refined effects of salience bias, which are linked to the actual meanings and implications of topic relevance or interest. In fact, it has been shown that while topics that are associated with positive feelings and memories may trigger survey participation, those that are equally relevant to a particular respondent, but which are associated with negative feelings and memories, may decrease the likelihood of participation.<sup>23</sup>

The comparison of estimates from the two HIS11 sub-studies showed that the results from the pre-recruited web panel survey were largely similar to those of the traditional postal/web survey, with only a few small variations being found between the sub-studies. Although the observed similarities across the two sub-studies cannot be used as evidence for the contention that the results constitute valid estimates of the population parameters (since both sub-studies may suffer from the same forms of nonresponse bias), one conclusion from the current study is that the pre-recruited web panel based on probability sampling may constitute a cost- and time-effective alternative to traditional respondent-sampling methods. One advantage of using a sample from a web-based register is that it might be possible to acquire additional relevant information about the non-respondents – information that could subsequently be used for response propensity weighting in the statistical analyses.<sup>24</sup>

As regards the comparison of estimates from the two UngKAB09 sub-studies, some quite substantial variation in estimates was noted between the sub-study based on a probability sample and the sub-study based on a self-selected sample – the respondents in the latter comprised a much more sexually active group of young people, and a much larger proportion of these respondents had experience of using the internet for flirting or to search for love. One conclusion based on this finding, and which corroborates those of numerous

previous studies, is that although self-selection sampling may provide an easy and inexpensive procedure for recruiting many respondents in a short time, such methods are likely to produce biased samples and must therefore be used with care in population surveys.<sup>25, 26, 27</sup> Taking this into account, it is worth noting that web surveys based on self-selected samples have proved to be particularly suitable for studies that examine so-called “hidden populations”, such as men who have sex with men.<sup>28</sup> At the same time, if the aim is to sample hard-to-reach populations, it might be preferable to use respondent-driven sampling – a method which resembles the non-probability snowball sampling method, but which makes use of network theory, physics, statistics and mathematics to reduce the biases associated with snowball sampling.<sup>29</sup>

Two limitations of the present study are that it is explorative rather than experimental in design, and that the conclusions are based on observations of similarities and differences across four data sets, which might all be biased due to nonresponse and the use of self-selection sampling. Despite this, our results and conclusions regarding the potential presence of nonresponse bias in two large, recent research projects on the topic of HIV and sexuality provide important insights for researchers who are willing to work hard in order to meet the nonresponse challenge.

### **Acknowledgements**

The authors would like to thank Margareta Forsberg, PhD, and Jonna Abellsson, BSc, for the development of the UngKAB09 questionnaire and for the collection of the UngKAB09 data.

### **Conflicts of interest**

None declared

## **Key-points**

- As a consequence of falling response rates in population surveys across countries and scientific disciplines, there has been an increase in studies focused on a search for new ways to explore and counter potential nonresponse bias.
- The study explores for the presence of nonresponse bias in the data from two large Swedish research projects on the topics of HIV and sexuality, by comparing a number of estimates of the respondents' knowledge, attitudes and behaviour in relation to HIV and sexuality across four data sets that were generated using different sampling and data collection methods.
- The results suggest that researchers working with population surveys in the area of HIV and sexuality must take into account the risk for nonresponse bias due to topic salience. However, it is important to be careful in assessing which survey items may or may not be affected by this form of nonresponse bias.
- The results from the study also indicate that the pre-recruited probability web panel may indeed provide a cost- and time-effective way of sampling respondents in future population surveys, but that researchers must be careful in drawing conclusions from population surveys based on self-selection sampling methods.

## **References**

1. Herlitz C, Brorsson B. AIDS in the minds of Swedish people: 1986-1989, *AIDS* 1990;4:1011-8.
2. Socialstyrelsen (The National Board of Health and Welfare). HIV och Aids i Sverige: Kunskaper, attityder och beteenden hos allmänheten 1987-2007 (HIV and AIDS in Sweden: Knowledge, attitudes and behaviours in the general population 1987-2007). Stockholm: Socialstyrelsen; 2008 January.

3. Smittskyddsinstitutet (Swedish Institute for Communicable Disease Control). HIV i Sverige: Kunskaper, attityder och beteenden hos allmänheten 1987-2011 (HIV in Sweden: Knowledge, attitudes and behaviours in the general population 1987-2011). Solna: Smittskyddsinstitutet; 2013 April.
4. The Population Research Institute. Sexual Trends in the Baltic Sea Area. Helsinki: The Population Research Institute; 2003 Report NO.: D 41/2003.
5. Nasjonalt folkehelseinstitutt (Norwegian Institute of Public Health). Rapport fra seksualvaneundersøkelsene i 1987, 1992, 1997 og 2002 (A report on the sexual behaviour projects from 1987, 1992, 1997 and 2002). Oslo: Nasjonalt folkehelseinstitutt; 2003.
6. Sundhedsstyrelsen (Danish Health and Medicines Authority). UNG2006. 15-24-åriges seksualitet: Viden, holdninger og adfærd (UNG2006. 15-24-year-olds' sexuality: Knowledge, attitudes and behaviour). København: Sundhedsstyrelsen; 2007 January.
7. Special issue: Survey nonresponse. *Journal of Official Statistics* 1999;15.
8. Special issue: Nonresponse bias in household surveys. *Public Opinion Quarterly* 2006;70.
9. Special issue: The Nonresponse Challenge to Surveys and Statistics. *The ANNALS of the American Academy of Political and Social Science* 2013;645.
10. Kreuter F. Facing the nonresponse challenge. *The ANNALS of the American Academy of Political and Social Science* 2013;645:23-35.
11. Tourangeau R, Plewes TJ (Eds.). Nonresponse in social science surveys: A research agenda. Washington: The National Academies Press; 2013.
12. Massey DS, Tourangeau R. Introduction: New Challenges to Social Measurement. *The ANNALS of the American Academy of Political and Social Science* 2013;645:6-22.
13. Bethlehem J. 2002. Weighting nonresponse adjustments based on auxiliary information. In Groves RM, Dillman DA, Eltinge JL, Little RJA, editors. *Survey nonresponse*. New York: Wiley, 2002:265-87.

14. Groves RM, Peytcheva E. The impact of nonresponse rates on nonresponse bias: A meta-analysis. *Public Opinion Quarterly* 2008;72:167-89.
15. Peytchev A. Consequences of survey nonresponse. *The ANNALS of the American Academy of Political and Social Science* 2013;645:88-111.
16. Massey DS, Tourangeau R. Where do we go from here? Nonresponse and social measurement. *The ANNALS of the American Academy of Political and Social Science* 2013;645:222-36.
17. Singer E, Ye C. The use and effects of incentives in surveys. *The ANNALS of the American Academy of Political and Social Science* 2013;645:112-41.
18. Czajka JL. Can administrative records be used to reduce nonresponse bias? *The ANNALS of the American Academy of Political and Social Science* 2013;645:171-84.
19. Olson K. Paradata for nonresponse adjustment. *The ANNALS of the American Academy of Political and Social Science* 2013;645:142-70.
20. Tikkanen RH, Abellson J, Forsberg M. UngKAB09: Kunskap, attityder och sexuella handlingar bland unga (UngKAB09: Knowledge, attitudes and sexual behaviour among youths). Göteborgs universitet: Institutionen för socialt arbete; 2011 Report No.: 1.
21. Groves RM, Presser S, Dipko S. The role of topic interest in survey participation decisions. *Public Opinion Quarterly* 2004;68:2-31.
22. Norstat, Sweden. <http://www.norstat.se/> Accessed [19<sup>th</sup> of April 2014]
23. Groves RM, Couper MP, Presser S, Singer E, Tourangeau R, Piani Acosta G, Nelson L. Experiments in producing nonresponse bias. *Public Opinion Quarterly* 2006;70:720-36.
24. Couper, MP. Web surveys: A review of issues and approaches. *Public Opinion Quarterly* 2000;64:464-94.

25. Chang L, Krosnik JA. National surveys via RDD telephone interviewing versus the internet: Comparing sample representativeness and response quality. *Public Opinion Quarterly* 2009;73:641-78.
26. Yeager DS, Krosnick, JA, Chang L, Javitz HS, Levendusky MS, Simpser A, Wang R. Comparing the accuracy of RDD telephone surveys and internet surveys conducted with probability and non-probability samples. *Public Opinion Quarterly* 2011;75:709-47.
27. Baker R, Blumberg SJ, Brick JM, Couper MP, Courtright M, Dennis JM, Dillman D, Frankel MR, Garland P, Groves RM, Kennedy C, Krosnick J, Lavrakas PJ. Research synthesis: AAPOR report on online panels. *Public Opinion Quarterly* 2010;74:711-81.
28. Ross MW, Tikkanen RH, Månsson S-A. Differences between internet samples and conventional samples of men who have sex with men: Implications for research and HIV interventions. *Social Science & Medicine* 2000;51:749-58.
29. Johnston LG, Sabin, K. Sampling hard-to-reach populations with respondent driven sampling. *Methodological Innovations Online* 2010;5:38-48.

Table 1. Characteristics of the HIS11 and the UngKAB09 data sets.

	HIS11 (Data set 1)	HIS11 (Data set 2)	UngKAB09 (Data set 3)	UngKAB09 (Data set 4)
Target population	People in Sweden (16-49 years)	People in Sweden (16-49 years)	Youths and young adults in Sweden (15-29 years)	Youths and young adults in Sweden (15-29 years)
Sampling procedure	Random sample stratified by age	Random sample stratified by age, sex and region	Random sample stratified by age	Self-selected sample
Mode of age stratification	20 % of the sample in each of the following age groups: 16-17, 18-19, 20-24, 25- 34, 35-49	20 % of the sample in each of the following age groups: 16-17, 18-19, 20-24, 25- 34, 35-49	25 % of the sample in each of the following age groups: 16, 20, 24, 28	--
Mode of data collection	Multimode Post/Web survey	Pre-recruited web panel survey	Web survey	Web survey
Response rate	25 %	23 %	24 %	--
Response rate per age group (original classification)	16-17: 21 % 18-19: 23 % 20-24: 24 % 25-34: 26 %	16-17: 25 % 18-19: 26 % 20-24: 24 % 25-34: 32 %	16: 28 % 20: 22 % 24: 23 % 28: 21 %	--
Sample size of reduced data set (16-29 years)	973	889	5602	8946

Table 2. Socio-demographic characteristics of the respondents. Percentages.

	HIS11 Post/web survey <sup>a</sup> (Data set 1)	HIS11 Pre-recruited web panel survey <sup>a</sup> (Data set 2)	UngKAB09 Web survey <sup>a</sup> (Data set 3)	UngKAB09 Web survey <sup>b</sup> (Data set 4)
<b>Age</b>				
16-19 years	54	45	30	33
20-24 years	29	41	47	43
25-29 years	17	14	23	24
<b>Sex</b>				
Male	35	35	44	51
Female	65	65	56	49
<b>Ethnicity</b>				
Swedish	77	84	80	81
2 <sup>nd</sup> generation immigrant	12	12	13	13
1 <sup>st</sup> generation immigrant	11	4	8	6
<b>Completed education</b>				
Nine-year compulsory school	35	31	32	27
Upper secondary school	47	51	45	53
University	15	17	14	10
<b>Current occupation</b>				
Employment	25	26	33	26
Studies	60	61	53	55
Unemployed	8	7	7	12
Long-term sick leave	0	1	1	3
<b>Current partner</b>				
Of opposite sex	96	95	97	96
Of same sex	4	5	3	4

<sup>a</sup> Sub-study is based on probability sampling

<sup>b</sup> Sub-study is based on self-selection sampling

Table 3. Knowledge and attitudes related to HIV and sexuality. Percentages and Means (Standard deviations).<sup>a</sup>

	HIS11 Post/web survey <sup>b</sup> (Data set 1)	HIS11 Pre-recruited web panel survey <sup>b</sup> (Data set 2)	UngKAB09 Web survey <sup>b</sup> (Data set 3)	UngKAB09 Web survey <sup>c</sup> (Data set 4)
<b>Knowledge</b>				
Agrees that HIV can be transmitted by kissing	16	15	19	19
Agrees that HIV can be transmitted by drinking from the same glass/bottle	11	10	12	11
Agrees that the risk of HIV is reduced if condom is used every time one has sex	92	90	94	96
Agrees that a person who looks healthy can be infected with HIV	96	93	95	96
<b>Attitudes</b>				
Agrees that sexual intercourse should only occur in stable relationships (lower means = agrees)	3.9 (1.4)	4.0 (1.3)	4.5 (1.1)	4.6 (0.9)
Agrees that sexual relations between two persons of the same sex are OK (lower means = agrees)	1.8 (1.2)	1.7 (1.0)	1.8 (1.3)	1.6 (1.2)
Perceives oneself at (very or rather) high risk of becoming infected with HIV	3	3	1	1
Perceives oneself at (very or rather) high risk of becoming infected with chlamydia	11	8	4	6

<sup>a</sup> Weights adjusting for age

<sup>b</sup> Sub-study is based on probability sampling

<sup>c</sup> Sub-study is based on self-selection sampling

Table 4. Behaviour related to HIV and sexuality. Percentages and Means (Standard deviations).<sup>a</sup>

	HIS11 Post/web survey <sup>b</sup> (Data set 1)	HIS11 Pre-recruited web panel survey <sup>b</sup> (Data set 2)	UngKAB09 Web survey <sup>b</sup> (Data set 3)	UngKAB09 Web survey <sup>c</sup> (Data set 4)
Total number of sex partners during past 12 months	1.6 (2.6)	1.3 (2.1)	2.2 (4.3)	3.1 (4.8)
> 5 sex partners during past 12 months	4	3	7	14
Use of condom at last vaginal or anal intercourse	23	24	23	25
Sex on first date/evening during past 12 months	20	17	26	41
HIV-tested during past 12 months	19	16	10	14
Flirted on the internet during past 12 months	36	39	37	64
Searched for sex or love on the internet during past 12 months	22	22	22	47

<sup>a</sup> Weights adjusting for age

<sup>b</sup> Sub-study is based on probability sampling

<sup>c</sup> Sub-study is based on self-selection sampling