

A Comparison of Item Nonresponse in Web and Pen-and-Paper Surveys of Sexual Behaviour

Alison R Evans, Jonathan Elford, Graham Bolding, and Richard D Wiggins.

Introduction

A number of factors contribute to item non-response in questionnaire surveys. A question may not be answered because the respondent does not know the answer or is unable to give one; by mistake or on purpose; because the appropriate answer is not included among the response categories; or because the response given cannot be read or coded (de Leeuw 2001).

Excluding respondents who fail to answer a question will reduce effective sample size. This weakens statistical power and creates problems for the creation of summative indicators, the analysis of multi-item scales and the application of multivariate techniques (Dickinson & Kirzner 1985). When item nonresponders differ systematically from responders on the variables of interest, item nonresponse will also lead to biased population estimates (Kupek 1999). One of the major goals of any survey design is therefore to minimise item nonresponse.

This paper is concerned with item nonresponse in surveys of sexual behaviour. Such surveys are sensitive to mode effect. Interviewer-administered surveys typically report lower item nonresponse than self-completion surveys (Dillman 1978). The advantage of the self-completion mode, however, is that it encourages more candid responses to questions of a sensitive nature (Fenton *et al.* 2001).

In self-completion surveys of sexual behaviour, use of computer-assisted self-interviews (CASI) is associated with lower item nonresponse than self-completed pen-and-paper questionnaires (Copas *et al.* 2002). For example, Johnson *et al.* (2001) found that the use of CASI in the National Survey of Sexual Attitudes and Lifestyles (Natsal) 2000 survey was associated with lower item nonresponse than the 1990 pen-and-paper version which may be partially attributed to the automatic routing of questions using CASI (Copas *et al.* 2002).

Web surveys (ie computerised questionnaires accessed through the Internet and completed online) share some of the features of other computer-assisted, self-administered surveys as well as having properties that are specific to the Internet itself. Although item nonresponse is lower for computerised questionnaires than for pen-and-paper self-completion surveys, the use of the Internet may bring other influences to bear. For example, web surveys are likely to be completed in a less controlled environment than CASI surveys which are almost always completed under supervision of an interviewer. Use of the Internet has been associated with multi-tasking which may distract some respondents and increase the incidence of item nonresponse (Lozar Manfreda & Vehovar 2002). People also tend to scan through websites with their fingers on the mouse ready to proceed (Bauman *et al.* 2000) which may also impact on the likelihood of responding to questions in the Internet mode. Use of the Internet has also been associated with disinhibited behaviour (Joinson 1998; Kiesler *et al.* 1984) which is believed to increase honesty but may have other consequences for the type of response. For example, respondents may feel less constrained by a social norm which suggests that they are expected answer all the questions.

This study set out to investigate whether the lower item nonresponse reported when comparing CASI to pen-and-paper in a survey of sexual behaviour (Copas *et al.* 2002; Johnson *et al.* 2001) would also be seen when comparing a *web survey* of sexual behaviour

with an equivalent *pen-and-paper survey*. And if so, was this effect heightened for questions of a sensitive nature?

Methods

This investigation examines data from the Internet and HIV study, a cross-sectional observational study of sexual behaviour among gay and bisexual men conducted in 2003. The methods of the Internet and HIV study are described in detail elsewhere (Elford *et al.* 2004).

Web survey

Web survey respondents were recruited via gaydar (www.gaydar.co.uk/) and gay.com (uk.gay.com/), two of the most popular websites for gay men in the UK. Over a five-week period in May and June 2003 pop-up and banner advertisements appeared in chatrooms and profile pages asking men to participate in the survey. Clicking on a pop-up or banner took respondents to the web survey which they could complete and submit online. Only respondents who said they were at least 18 years old were permitted to answer the questionnaire. Although the pop-ups and banners were restricted to UK chatrooms and profiles, people from anywhere in the world using these sites had the opportunity to participate (Elford *et al.* 2004).

Pen-and-paper survey

In 2003, all men who used one of seven central London gyms during a one-week period between January and March were invited to participate in the pen-and-paper survey. All seven gyms had a substantial gay membership (40-100% of clients); only men who identified as gay or bisexual were asked to answer the questions on sexual behaviour. Completed pen-and-paper questionnaires were returned to collection boxes in the gym or posted back to the research team. Gay and bisexual men who use central London gyms

are broadly representative of London gay men 'on the scene', who go to gay bars, clubs and so on (Elford 2002). Gym-based surveys are therefore an effective method of gathering detailed data on sexual behaviour from this group of men (Elford *et al.* 2004).

Questionnaire comparability

Although the two questionnaires were designed to gather commensurate data as part of the same study, the questions that they contained were not all identical. The web survey contained a total of 160 questions and the pen-and-paper survey contained a total of 124 questions. The content of the majority of the questions was the same and questions followed the same order but the web survey included some more detailed questions, particularly about use of the Internet and sexual partners sought on the Internet. Both questionnaires contained complex routing patterns. In order to programme these skip and branch patterns in the web survey, it was necessary to include some compulsory questions in order to ensure correct routing. Account was taken of this in an analysis of 78 questions that were not compulsory in the web survey, as described below.

Item nonresponse

Item nonresponse is defined here as occurring when respondents failed to provide an answer to questions to which they were directed. Although nonsubstantive responses such as 'don't know' or 'not sure' are often coded as missing data in studies investigating item nonresponse, only cases where the respondent left the response blank are counted here. In the majority of these questions, the 'don't know' type response conveyed information (eg 'Do you know your boyfriend's HIV status?') and was included in both the web and pen-and-paper surveys.

Item nonresponse for each question was calculated as the total number of respondents who missed the question divided by the total number of respondents who were presented with the question. Item nonresponse for each question is expressed as a percentage.

The content of the Internet and HIV surveys required complex routing and, for the web survey, the programming of the skip and branch patterns necessitated inclusion of compulsory questions to ensure correct routing. Thus, 22.5% of the questions in the web survey were compulsory and *could not* be missed. Where the same questions occurred in the pen-and-paper survey, they *could* be missed. In order to compare item nonresponse for individual questions, 78 questions that were not compulsory in the web survey (and could therefore be missed in either survey) were identified as being identical or near-identical in the two modes. Examples of 'near-identical' questions include those where the web survey used slightly different questions to those in the pen-and-paper questions because they were presented on separate web pages or where words were contracted in one mode but not the other.

In order to examine the effect of mode on questions of a sensitive nature, a subset of matching questions on unprotected anal intercourse (UAI) was identified (n = 8). These questions were selected because UAI with a non-primary partner is likely to be a sensitive topic among gay and bisexual men (Acree *et al.* 1999).

Sample comparability

Respondents were not randomly assigned to survey mode. To ensure greatest comparability between the two samples, only men who finished each survey (those who reached the last question in the web survey and those who returned a completed pen-and-paper survey), who said that they were living in London and reported sex with a man in the last year were included in the analysis.

Statistical analysis

Item nonresponse for all questions was compared using the Mann-Whitney U test and the Wilcoxon matched pairs test was used for the matched questions. Nonparametric tests were adopted because violation of normality precluded use of t tests (Siegel & Castellan 1988). The linear relationship between matching questions was examined using Pearson's correlation.

Findings

All questions

Overall mean item nonresponse was 1.6% (s.d. 3.8%) for the web survey and 6.7% (s.d. 8.1%) for the pen-and-paper survey ($U = 3,801.5$, $p \leq 0.01$); median item nonresponse was 0.5% and 3.6% respectively; maximum item nonresponse was 23.9% in the web survey for a question about partner HIV status and 40.4% in the pen-and-paper survey for a question about job description.

Some of the questions in the web survey could not be missed which may have contributed to the lower item nonresponse in the web mode. The remaining analysis therefore excludes any questions that were compulsory in the web survey and includes only questions which could be matched across modes. The questions included are those where the wording and format were identical or nearly so, resulting in a total of 78 matched questions.

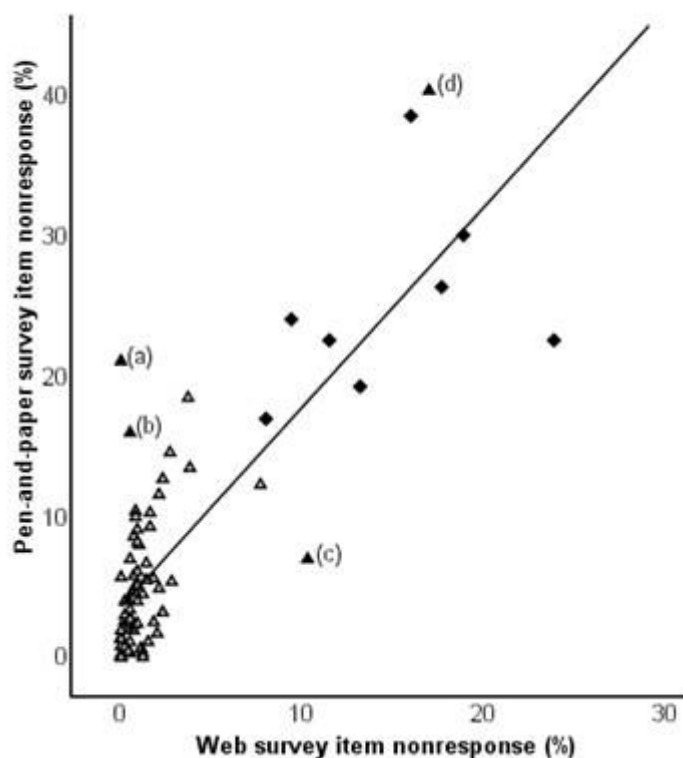
Matched questions

After excluding compulsory questions in the web survey, item nonresponse was still lower than in the pen-and-paper survey. Mean item nonresponse for the 78 matched questions was 2.8% (s.d. 5.0%) for the web survey and 7.5% (s.d. 8.7%) for the pen-and-paper survey

($z = -6.93, p \leq 0.01$). Median item nonresponse for the 78 matched questions was 0.9% and 4.5% respectively; maximum item nonresponse remained unchanged.

The correlation coefficient for item nonresponse for the 78 matched questions was 0.82 ($p \leq 0.01$), indicating that the same questions were likely to be missed in each survey (Figure 1).

Figure 1: Item nonresponse for the 78 matched questions



There were three questions [(a), (b) and (c) in Figure 1] which were exceptions to this pattern. Questions (a) and (b) had extremely low item nonresponse in the web survey and disproportionately high item nonresponse in the pen-and-paper survey. Question (a) asked respondents whether they personally knew gay or bisexual men who were HIV positive. It resulted in 0.0% nonresponse in the web survey and 21.1% in the pen-and-paper survey. It is likely that the different presentation of this question in the two surveys accounted for this

difference. Figure 2 shows the format of the question in the pen-and-paper survey and illustrates how it is likely that the majority of men who missed this question thought that they were answering it but ticked one of the boxes corresponding to the question that followed (about whether these men were diagnosed in the last 12 months). Inspection of the data revealed that 88.3% of the pen-and-paper respondents who missed this question went on to answer the question about the time of diagnosis.

Figure 2: Pen-and-paper survey format for question about knowing HIV positive men

Do you personally know any gay/bisexual men with HIV infection?

YES NO

If YES, were any of these men diagnosed with HIV in the last 12 months? YES NO

Figure 3 shows how the two questions were presented separately in the web survey, with respondents automatically routed to the appropriate follow-up question. It appears that the higher item nonresponse for this question in the pen-and-paper survey is related to the question format rather than content.

Figure 3: Web survey format for question about knowing HIV positive men



Question (b) asked respondents how frequently they had used amyl nitrate or poppers in the past 12 months. It resulted in 0.5% nonresponse in the web survey and 16.0% nonresponse in the pen-and-paper survey. Amyl nitrate is not a classified drug and yet this item was more likely to be missed in the pen-and-paper survey than any of the other recreational drugs listed (marijuana, ecstasy, amphetamine, cocaine, viagra, ketamine,

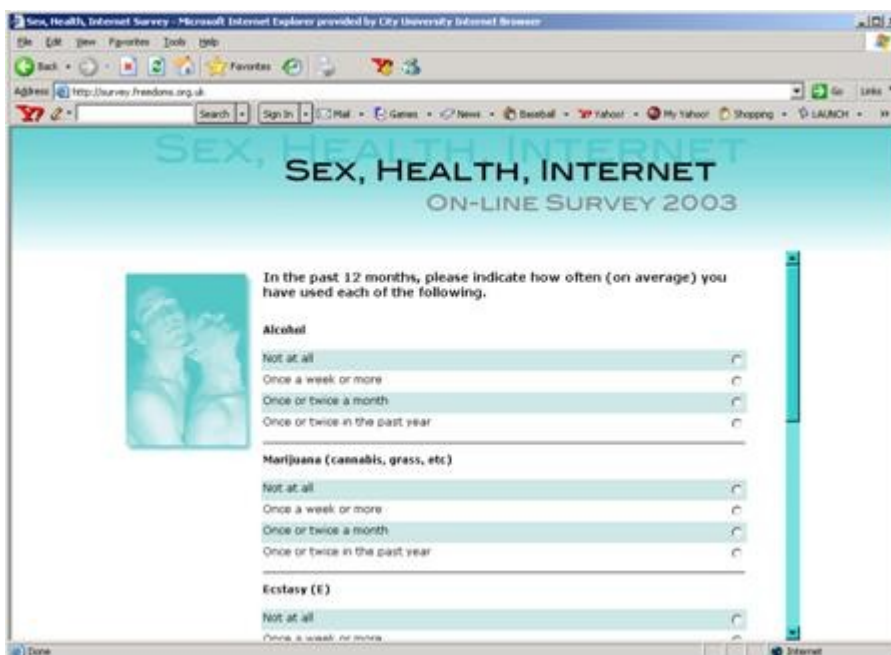
crystal methamphetamine). Inspection of the original questionnaires suggests that the grid-like format of the pen-and-paper survey (Figure 4) contributed to this effect as respondents may have become confused further down the list about which tick box referred to which substance. Items in the web survey were presented in the same order as those in the pen-and-paper survey but with a much higher degree of separation, as illustrated in Figure 5.

Figure 4: Pen-and-paper survey format for questions about alcohol and drug use

IN THE PAST 12 MONTHS please indicate how often (on average) you have used each of the following:

	About once a week or more	About once or twice a month	Only once or twice in the past year	Not at all
Alcohol	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Marijuana (cannabis, grass, etc)	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Ecstasy ('E')	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Speed (amphetamine)	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Cocaine ('coke')	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Viagra	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Ketamine ('K')	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Crystal methamphetamine	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Poppers (amyl nitrate)	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>

Figure 5: Web survey format for questions about alcohol and drug use



The fact that the response categories were presented in different orders in the two surveys, may also have contributed to gym respondents being more likely to miss the question than tick the 'not at all' box. These differences in format between the two modes appear to have had an important impact on item nonresponse. In this way, the question format again appears to be implicated in the disproportionately high levels of item nonresponse for this question in the pen-and-paper survey.

Question (c), on the other hand, had higher item nonresponse in the web survey. It asked respondents whether they supervised any employees. Interestingly this was one of the few occasions where two related questions were presented on the same web page although there are no other obvious reasons as to why this question might result in higher item nonresponse in the web survey.

High item nonresponse

Question (d) had high item nonresponse in both surveys: 17.0% in the web survey and 40.4% in the pen-and-paper survey. It was an open question which asked respondents to provide a description of their job which may have been difficult to articulate following the question which asked them to provide their job title.

The set of questions that are denoted by a diamond (♦) in Figure 1 also had high nonresponse in both surveys. Six of these eight questions were about disclosing HIV status to sexual partners with whom respondents reported having unprotected anal intercourse (UAI). Questions about UAI are likely to be a sensitive among this population of men (Acree *et al.* 1999) and the issue of disclosing HIV status to these partners is particularly delicate. The sensitivity of this topic is likely to have contributed to the high level of item nonresponse in both surveys. All eight of the questions followed the same format, as illustrated in Figure 6. While this format required respondents to provide a response to

each item, it is possible that they left a blank instead of inserting a '0' which is a recognised cause of item nonresponse (de Leeuw 2001).

Figure 6: Format for questions about anal sex partners



Of these men (ie those you had anal sex without a condom with) how many...

- ... told you their HIV status? _____men
- ... did you tell your HIV status? _____men
- ... did you KNOW had the same HIV status as you? _____men
- ... lived in London? _____men

Sensitive questions

If it is the case that web surveys result in more complete reporting of *sensitive information* than pen-and-paper surveys, then item nonresponse should be disproportionately lower for those questions in the web survey. That is to say, there should be a reduction in item non-response over and above the general mode effect. This was not found to be the case for

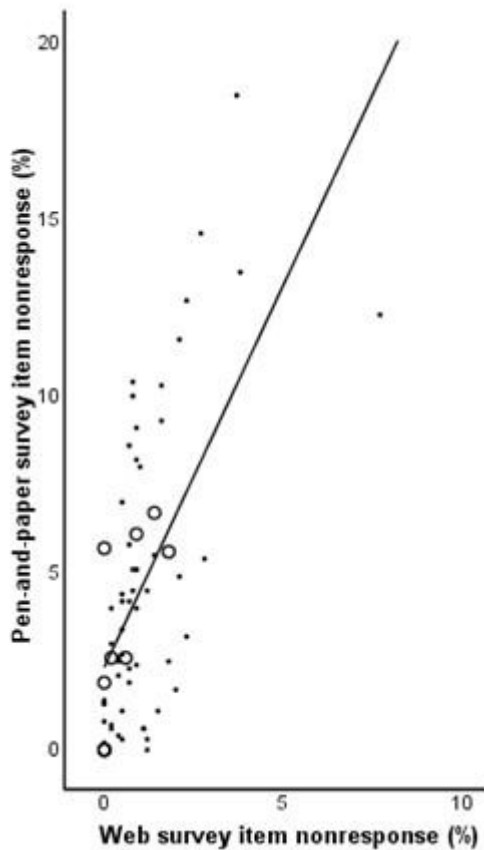
questions where there was high item nonresponse, as described in the previous section. In this section, we examine questions which had lower levels of item nonresponse by excluding the 12 high-item-non-response questions and those with disproportionate item nonresponse described in the last section [questions (a), (b), (c), (d) and the questions indicated by a diamond (◆) in Figure 1]. Item non-response was then examined in the remaining 66 matched questions, which included 8 questions on UAI.

Mean item nonresponse for the 66 matched questions was 1.0% (s.d. 1.2%) for the web survey and 4.6% (s.d. 4.1%) for the pen-and-paper survey ($z = -6.50, p \leq 0.01$) (Table 1). The correlation coefficient for item nonresponse for the 66 questions was 0.63 ($p \leq 0.01$). Mean item nonresponse for the 8 UAI questions was 0.6% (s.d. 0.7%) for the web survey and 3.9% (s.d. 2.4%) for the pen-and-paper survey. The correlation coefficient for item nonresponse for the 66 matched questions was 0.67.

In figure 7, questions about UAI are highlighted by circles (o). The figure shows that the UAI questions conform reasonably well to the linear relationship between modes. They are clustered relatively close to the regression line and fall on both sides of it, in a manner indicating that their item non-response is not systematically different from that of the other matched questions.

Figure 7:

Item nonresponse for UAI questions in the context of 66 matched questions with low item non-response



Notes for Figure 7

Questions on UAI are indicated by a circle (o)

Sample effects

Respondents were not randomly allocated to survey mode and it is possible that the differences in item nonresponse were due to sample as well as mode effects. Comparisons made elsewhere between the men recruited on the Internet and men recruited in the gyms found that the Internet respondents were younger, less educated, less likely to have sex only with men, be in a relationship or have tested for HIV (Elford *et al.* 2004). However, there were no significant differences in item nonresponse for any of these variables in either sample.

Discussion

Our study has highlighted a strong mode effect whereby the web survey consistently had lower item nonresponse than the pen-and-paper survey. There was no evidence, however, that the mode effect was heightened for questions of a sensitive nature.

This new evidence on web surveys of sexual behaviour corroborates findings from previous surveys of sexual behaviour where *computerisation* reduced item nonresponse (Copas *et al.* 2002; Johnson *et al.* 2001) and from web surveys on a range of other topics where item nonresponse was lower in the web mode (Kwak & Radler 2002; McMahon *et al.* 2003; Stanton 1998).

The Internet and HIV surveys (web and pen-and-paper) contained complex routing patterns. This is likely to have contributed to the lower item nonresponse in the web survey where the routing was automatically programmed. Web survey respondents who tried to miss compulsory questions were told that they needed to provide an answer. However, the lower item nonresponse in the web survey could not be explained solely by the inclusion of compulsory questions since item non-response was also lower for the non-compulsory questions. It would have been interesting to explore the reasons for the mode effect using follow-up qualitative interviews with respondents. However, there was no provision for this in the research design.

There was a high degree of correlation between item nonresponse in the two surveys indicating that the question itself was an important source of item nonresponse. In general, questions that were more likely to be missed in the web survey were also more likely to be missed in the pen-and-paper survey, and vice-versa. However, there was no evidence that the web survey offered any advantage in the reporting of sensitive behaviours over and

above the general mode effect, consistent with other research into computerised questionnaires (Johnson *et al.* 2001; Richman *et al.* 1999)

High item nonresponse in both surveys may be attributed to a combination of question format and question sensitivity which increased item nonresponse in both modes. Other studies have shown that questions requiring more thought produce higher item nonresponse (Ferber 1966) and that item nonresponse is associated with recall difficulties or complexity of format rather than question sensitivity (Kupek 1998; Wadsworth *et al.* 1993).

There are some limitations to this study which compared item nonresponse in the 2003 Internet and HIV web and pen-and-paper surveys. Since participants were not randomly allocated to each survey mode, there were some differences between the characteristics of the two samples. However, respondent characteristics that vary between the Internet and gym samples (Elford *et al.* 2004) are unlikely to explain our findings. There were no significant differences in item non response according to age, education, gender of sexual partners, being in a relationship or having tested for HIV in either sample. Sample comparability was enhanced by including only men who finished each survey, who said that they were living in London and reported sex with a man in the last year.

The questionnaires were not identical in length or content and while the analysis of matched questions controlled for question content, these 78 questions may not be representative of each of the questionnaires as a whole. On the other hand, the web survey contained *more* questions than the pen-and-paper survey and similar content, but its overall level of item nonresponse remained consistently lower.

It is estimated the web survey respondents constituted less than 1% of men using gaydar and gay.com during the time that the web survey was online whereas the estimated response rate for the gym sample was 50-60% (Elford *et al.* 2004). The low “response rate” among web survey respondents suggests that they may have been committed participants who were strongly motivated to respond to each question. However, pen-and-paper survey respondents were personally handed a questionnaire and invited to participate as they came into or left the gym, which may also have been a motivating factor. A case based analysis indicated that respondent variables were not strong predictors of item nonresponse in either survey (Evans 2006).

To conclude, gay and bisexual men who volunteer for surveys of sexual behaviour are clearly willing to provide information on their sexual behaviour and our findings suggest that, on the whole, those who decide to participate try to provide the information requested. It appears, however, that use of the Internet mode is likely reduce item nonresponse compared with a pen-and-paper questionnaire but there was no evidence that the mode effect was heightened for questions of a sensitive nature. Our study suggests that web surveys provide an effective means of gathering more complete data in studies of sexual behaviour among gay and bisexual men.

References

Acree, M., Ekstrand, M., Coates, T. J., & Stall, R. (1999) 'Mode effects in surveys of gay men: A within-individual comparison of responses by mail and by telephone', *Journal of Sex Research*, Vol. 36, pp. 67 - 75.

Bauman, S., Jobity, N., Airey, J., & Atak, H. (2000) 'Invites, Intros and Incentives: Lessons from a web survey', Paper presented at 55th Annual Conference of the American Association for Public Opinion Research, 17 - 21 May, 2000, Portland, OR. Retrieved April 18, 2005, from: <http://www.wirthlin.webaxxs.com/wintellitech/aapor5.pdf>

Copas, A. J., Wellings, K., Erens, B., Mercer, C. H., McManus, S., Fenton, K. A., Korovessis, C., Macdowall, W., Nanchahal, K., & Johnson, A. M. (2002) 'The accuracy of reported sensitive sexual behaviour in Britain: exploring the extent of change 1990-2000', *Sexually Transmitted Infections*, Vol. 78, pp. 26 - 30.

de Leeuw, E. D. (2001) 'Reducing missing data in surveys: An overview of methods', *Quality & Quantity*, Vol. 35, pp. 147 - 160.

Dickinson, J. R. & Kirzner, E. (1985) 'Questionnaire item omission as a function of within-group position', *Journal of Business Research*, Vol. 13, pp. 71 - 75.

Dillman, D. A. (1978) *Mail and telephone surveys the total design method*. New York, NY: John Wiley.

Elford, J. (2002) 'Behavioural surveillance in London gyms: Trends, comparisons and methodological strengths', Paper presented at Behavioural Surveillance Collaborator's Meeting, 8 November, 2002, London. Retrieved August 30, 2005, from: http://www.hpa.org.uk/infections/topics_az/hiv_and_sti/behavioural/report.htm

Elford, J., Bolding, G., Davis, M., Sherr, L., & Hart, G. (2004a) 'The Internet and HIV study: design and methods', *BMC Public Health*, Vol. 4, No. 39. Retrieved from: <http://www.biomedcentral.com/1471-2458/4/39>

Elford, J., Bolding, G., Davis, M., Sherr, L., & Hart, G. (2004b) 'Web-based behavioral surveillance among men who have sex with men: A comparison of online and offline samples in London, UK', *Journal of Acquired Immune Deficiency Syndromes*, Vol. 35, pp. 421 - 426.

Evans, A.R. (2006) 'Use of the Internet as a data collection tool: a methodological investigation of sampling and mode effect', Thesis (PhD), City University.

Fenton, K. A., Johnson, A. M., McManus, S., & Erens, B. (2001) 'Measuring sexual behaviour: methodological challenges in survey research', *Sexually Transmitted Infections*, Vol. 77, pp. 84 - 92.

Ferber, R. (1966) 'Item Nonresponse in a Consumer Survey', *Public Opinion Quarterly*, Vol. 30, pp. 399 - 415.

Johnson, A. M., Copas, A. J., Erens, B., Mandalia, S., Fenton, K., Korovessis, C., Wellings, K., & Field, J. (2001) 'Effect of computer-assisted self interviews on reporting of sexual HIV risk behaviours in a general population sample: a methodological experiment', *AIDS*, Vol. 15, pp. 111 - 115.

Joinson, A. N. (1998) 'Causes and implications of disinhibition on the Internet', in J. Gackenbach (editor) *Psychology and the Internet*. London: Academic.

Kiesler, S. L., Siegel, J., & McGuire, T. W. (1984) 'Social Psychological Aspects of Computer-Mediated Communication', *American Psychologist*, Vol. 39, pp. 1123 - 1134.

Kupek, E. (1998) 'Determinants of item nonresponse in a large National Sex Survey', *Archives of Sexual Behavior*, Vol. 27, pp. 581 - 594.

Kupek, E. (1999) 'Estimation of the number of sexual partners for the nonrespondents to a large national survey' *Archives of Sexual Behavior*, Vol. 28, pp. 233 - 242.

Kwak, N. & Radler, B. (2002) 'A Comparison Between Mail and Web Surveys: Response Pattern, Respondent Profile, and Data Quality', *Journal of Official Statistics*, Vol. 18, pp. 257 - 273.

Lozar Manfreda, K. & Vehovar, V. (2002) 'Do Mail and Web Surveys Provide Same Results', *Metodološki zvezki*, Vol. 18, pp. 149 - 169.

McMahon, S. R., Iwamoto, M., Massoudi, M. S., Yusuf, H. R., Stevenson, J. M., David, F., Chu, S. Y., & Pickering, L. K. (2003) 'Comparison of e-mail, fax, and postal surveys of pediatricians', *Pediatrics*, Vol. 111, pp. 299 - 303.

Richman, W. L., Kiesler, S., Weisband, S., & Drasgow, F. (1999) 'A meta-analytic study of social desirability distortion in computer-administered questionnaires, traditional questionnaires, and interviews', *Journal of Applied Psychology*, Vol. 84, pp. 754 - 775.

Siegel, S. & Castellan, N. J. (1988) *Nonparametric Statistics for the Behavioral Sciences* (2nd edn). New York, NY: McGraw-Hill.

Stanton, J. M. (1998) 'An empirical assessment of data collection using the Internet', *Personnel Psychology*, Vol. 51, pp. 709 - 725.

Wadsworth, J., Field, J., Johnson, A. M., Bradshaw, S., & Wellings, K. (1993) 'Methodology for the National Survey of Sexual Attitudes and Life-styles', *Journal of the Royal Statistical Society: Series A - Statistics in Society*, Vol. 156, pp. 407 - 421.

Acknowledgements

The authors would like to thank gaydar, gay.com and all the men who participated in the studies.

Funding

This research was funded by the Economic and Social Research Council. The “Internet and HIV” study was funded by the Medical Research Council (grant number GO 100159). Natsal 2000 was supported by a grant from the Medical Research Council with funds from the Department of Health, the Scottish Executive, and the National Assembly for Wales.