Assessing non-response to a mailed health survey including self-collection of biological material

Anneli Uusküla¹, Mart Kals¹, Louise-Anne McNutt²

¹ Department of Public Health, University of Tartu, Tartu, Estonia
² Department of Epidemiology, School of Public Health, State University of New York, Albany, USA

Correspondence: Anneli Uusküla, Department of Public Health, University of Tartu, Ravila 19, 50411 Tartu Estonia, tel: 3727374195, fax: 3727374192, e-mail: anneli.uuskula@ut.ee

Received 6 December 2009, accepted 9 April 2010

Background: Collection of biological material via mailed health surveys is an emerging trend. This study was conducted to assess non-response bias in a study of sexually transmitted infection utilizing self-collected, home-obtained specimens. Methods: Data from a nationwide administrative database on health care utilization together with data from a research study were used. The research study was an outreach screening programme including home-obtained, participant-collected, mail-delivered testing for Chlamydia trachomatis. A random sample of 1690 persons aged 18–35 years from the population registry was selected. Study materials (specimen collection kit, informed consent, questionnaire) were mailed in three waves. Results: The first mailing yielded a response rate of 18.5% (n = 259), the second 10.1% (n = 141) and the third 11.4% (n = 160). Women were more likely to respond than men, and responders were less likely to have had medical care in the past year and more likely to have had a prior sexually transmitted infection than non-responders. Chlamydia trachomatis infection rates tended to be higher in early responders. Late responders appeared more like non-responders in terms of demographic factors, health care utilization patterns and potential disease status. Conclusion: Non-response in a health survey including biological material self-collection warrants research as it may differ from non-response in general health questionnaires.

Keywords: Chlamydia, epidemiologic study, participant collected specimen, participation

Introduction

Research using mailed population surveys has grown increasingly difficult because response rates are diminishing. Several decades ago a non-response rate of ≥20% was considered fatal but modern non-response proportions typically exceed 20%.1 Because the factors that predict research participation may vary over time, continued study of non-response is needed to profile today’s non-responders. Research that quantifies the factors associated with non-response should inform the development of more economical population-based research while optimizing scientifically valid estimates of disease, behaviours and other health-related factors.2,3

In most epidemiological studies based on gathering information through direct contact with study subjects a certain proportion of the target population refuses or is not able to participate. A long-held survey research model to understand the impact of non-response is the ‘continuum of resistance’.4 This model postulates that there is a continuum that ranges from those who will respond quickly to those who will not respond at all, and that the late responders are similar to non-responders. Late responders are subjects who need either repeated contacts or particular strategies of contact before participation. It has been proposed that the characteristics of non-respondents may be extrapolated from those of late respondents.5–7

The collection of biological material by non-invasive patient-collected samples is increasingly used. This technique has been used in studies of markers of hepatitis A⁸ and HIV,⁹,¹⁰ cancer screening,¹¹ premalignant states,¹²,¹³ vaccines,¹⁴ gynecology and antenatal care,¹⁵–¹⁷ and sexually transmitted infections.¹⁸,¹⁹ New research models for sexually transmitted infections (STIs) are now possible. Nucleic acid amplification tests (NAATs) for patient-collected samples of several STIs allow researchers to survey probability samples from the general population rather than samples from patients attending a clinic or other special populations.¹⁹ Another innovation is collection of biological material via mailed health surveys. There is evidence that research utilizing self-collected, home-obtained specimens is an efficient method of reaching people.²⁰–²² However, non-response continues to be a concern with any sampling approach for STI research. Response rates in research utilizing self-collected, home-obtained specimens range from 20% to 50%.²⁰,²¹ Scant evidence is available on whether non-respondents differ from respondents in terms of infection rates, health or risk behaviour profiles.

Like many European countries, Estonia has a nationwide administrative database on health care utilization and population registry data available for research purposes. This data together with data from a mailed health survey including biological material self-collection were used to assess non-response bias in a study of STIs among a random sample of the general population. The study utilized participant-collected, mail-delivered testing for Chlamydia trachomatis together with a self-administered survey. Infection rates, socio-demographic and health care utilization data are presented for respondents (by response waves) and non-respondents.
Methods

Sample
A total of 1690 persons (845 women and 845 men) aged 18–35 years were randomly sampled from the Estonian Population Registry’s list of Tartu county residents (which includes residents of Tartu city and the rural areas around the city). Tartu is Estonia’s second largest city, and Tartu county is typical for Estonia including the proportion of rural residents (~30%).

The Estonian Population Registry is an electronic database containing personal data on all citizens and aliens residing in Estonia. The registry may be extracted by public and private sector organizations according to defined criteria. The data are used in statistics, polling, medical research, direct mailing, etc. The data abstracted for this study included demographic variables (i.e. age, sex, residence, ethnicity/mother tongue).

Surveying and measurements
This outreach screening programme included a questionnaire and home-obtained, participant-collected, mail-delivered testing for C. trachomatis. Each participant was mailed a package containing a specimen collection kit, collection instructions and an informed consent form with a 35-item questionnaire and asked to collect their own specimens (urine in men and vaginal swab in women) and post them to the researchers. Study materials were mailed in three waves:

(1) Wave 1: the first mailing of the questionnaire and specimen collection kit;
(2) Wave 2: 10 days later a reminder was sent to those who had not responded; and
(3) Wave 3: a week later subjects who had still not responded received another study package.

Study mailing was considered ‘undeliverable’ in cases of incorrect address (the indicated house or apartment could not be found), wrong address (the person no longer lived at the stated address) or missing mailbox.

Participants returned specimens to the laboratory in pre-stamped, pre-addressed envelopes. NAAT (artus C. trachomatis) TM PCR Kit (Qiagen, Hamburg, Germany) with ABI PRISM 7900HT Sequence Detection System (Applied Biosystems, Foster City, CA, USA) was used to detect C. trachomatis in self-collected samples.

A letter or an email (depending on the participant’s preference) containing the C. trachomatis test result was sent to participants. In the case of a positive result, respondents were advised to seek treatment from their regular provider or a dermatovenerologist on the study team.

Health care utilization data
The nationwide administrative database of the Health Insurance Fund (HIF) was used to abstract health care utilization data (for respondents and non-respondents): dates of outpatient/inpatient care, number of health care encounters and diagnoses (coded using the international classification of diseases and related health problems, ICD-10). For non-respondents, the data abstracted from the HIF did not contain information allowing personal identification. For respondents, permission to link personal data to the HIF data was obtained through informed consent. Investigators were specifically interested in STI/genital tract inflammation diagnoses (cervicitis, vaginitis, urethritis and inflammatory prostatitis) within the 12 months before the study.

Health insurance in Estonia is funded through a compulsory scheme under which employers are obliged by law to pay social and health insurance taxes for their employees. Self-employed people pay a social tax based on their income. Individuals for whom social/health insurance tax is paid by their employer or who pay it themselves are considered covered by health insurance (‘the insured’) and are members of the HIF. As of 31 December 2006, the number of insured persons registered by the HIF was 1278 016 (or 95.2% of the Estonian population).

Statistical analysis
Differences in characteristics between study groups (respondents, respondents by mailing waves vs. non-respondents) were analysed using the χ²-test, and odds ratios (OR) with 95% confidence intervals (95% CI) together with P-values. Adjusted odds ratios (AOR) were calculated using all variables (age, gender, residency, health care utilization, STI diagnosis with in last 12 months) in the logistic regression models. For the comparison analysis, 36 non-responders (5%; 36/838) were excluded from the analysis due to missing information in the HIF database. Prevalence of genital chlamydial infection was calculated by the number of cases divided by the number of submitted specimens. All analyses were performed with SAS statistical software (version 9.1).

The study protocols followed data protection legislation and ethical regulations of the University of Tartu and the Estonian Data Protection Inspectorate. The study was approved by the researchers’ universities. The study subjects did not receive any financial incentives for participation.

Results

Response rate
Study invitations were received by 83% (1398/1690) of the targeted sample but 292 individuals could not be reached because the packet was undeliverable (e.g. moved, wrong address and post box missing). These people were excluded from the analysis. Overall, 560 (40%) of those contacted participated: 479 returned both questionnaire and specimen (34%), 73 only the questionnaire and eight only the specimen. For this analysis, respondents were defined as those who returned either the questionnaire or the specimen. The first mailing (Wave 1) yielded a response rate of 18.5% (n = 259), the second mailing an additional 10.1% (n = 141) and the third wave yielded an additional 11.4% (n = 160).

Differences between early respondents, late respondents and non-respondents
Demographic profiles on early respondents (Wave 1), late respondents (Wave 3) and non-respondents are presented in table 1. There were significant differences between respondents and non-respondents. In terms of socio-demographic characteristics measured, respondents included a higher proportion of women and rural residents. Early respondents (Wave 1) differed significantly from non-respondents by gender, and late respondents did not differ from non-respondents. Health care utilization patterns differed significantly between respondents and non-respondents, and between early and late responders and non-responders. In all the comparisons, a higher proportion of non-respondents had utilized outpatient health care within the past 12 months and higher proportions of respondents (early, late, all) had a STI and/or
The 'continuum of resistance' model suggests that late responders are similar to non-responders. Still, while some studies have failed to support the assumption that late responders can be used as a proxy for non-responders, others have suggested that non-response bias can be estimated by analysing late respondents.

In our study, some patterns are consistent with the continuum of resistance model. Late responders appeared more like non-respondents in terms of demographic factors, health care utilization patterns and potentially disease status. Some of these patterns may be associated with characteristics different exposure- and disease-related characteristics from respondents. Differences in respondent characteristics may lead to bias in the measure of disease or exposure (as in this study), and in the estimate of the measure of association between exposure and disease. When participation is associated only with either exposure or outcome the estimate of the OR is not biased. Combined response bias (response depending on both exposure prevalence and disease incidence) can alter both the relative risk and the OD, and is probably the source of any significant distortion in these measures of association. Low participation rates in prevalence studies do not necessarily cause fatal selection bias; however, they do underscore the need to estimate potential bias. Ideally, a separate investigation of non-response is conducted either for all non-responders or for a representative subsample. Alternatively, non-response bias may be indicated from the population characteristics, by outcome differences for early and late respondents, or between respondents and non-respondents. Our analysis of non-response relied on comparisons of the socio-demographic and health care utilization characteristics of both respondents and non-respondents, as well as early and late respondents. Methodologically, our analysis took advantage of the study design (multiple recruitment waves) and the opportunity to obtain complete administrative data on respondents and non-respondents.
not directly related to the disease studied, while others might be due to the clinical incentive of this particular study. Women generally tend to be more willing to participate in health surveys,\textsuperscript{36,37} thus their earlier response here was not surprising. The clinical benefit is worthy of particular notice and discussion. In most cases, for population research involving diagnostic tests (including those for STI), ethical considerations require that results are shared with participants. Thus, it is reasonable to expect selection bias arising from this clinical/health care incentive. Therefore the clinical incentive effect on the response pattern in a health survey including biological material collection/testing might differ from that observed in general health questionnaires. Selective participation may arise among people who have limited access to health care and who might misunderstand or use the study as an opportunity to receive medical services. Collecting data to assess selection bias, whether through multiple waves of sampling, comparison with a population database, or through sub-studies will be routinely needed if changes in health care delivery occur. In addition, in the absence of face-to-face contact with the research team, selective participation may arise from distrust or fear of a possibility of the submitted biological material being misused (e.g. for illicit drug use detection, paternity testing, DNA testing/storage or for other purposes for which participants have not given consent).

Some limitations in our study have to be acknowledged. We did not conduct follow-up surveys with non-respondents. Frequently non-response follow-up studies suffer from low response rates.\textsuperscript{31,38,59} Instead of this traditional approach, we complemented the analysis with data from the HIF. It should be noted that 5\% ($n = 36$) of non-respondents who had no records in the HIF database were excluded from the analysis. This differential exclusion probably leads to bias in measuring the association between response status and (out-patient) health care utilization. However, even when the excluded 36 non-respondents are included into the category of non-respondents with zero out-patient visits the same pattern (individuals who responded were less likely to have had medical care in the past year) of significant difference remains between study groups: proportion with zero out-patient visits: Wave 1: 18\%, Wave 3: 21\%, respondents: 18\%, non-respondents (non-R): 6\%; Wave 1 vs. non-R: $P < 0.001$, Wave 3 vs. non-R $P < 0.001$, All respondents vs. non-R: $P < 0.001$. Further research is also needed to assess the potential for bias in research arising from the non-response related to the subset of population who cannot be reached located via mail.

Research that quantifies factors associated with, and the impact of, non-response informs the development of public health initiatives. Conversely, as home sampling programs are increasingly promoted for public health,\textsuperscript{40} it may be prudent to consider the voluntary collection of population and health characteristics that would be useful in estimating population disease levels. Integrating public health initiatives with population research components is feasible and may be particularly beneficial to ensure an economical public health approach.

**Funding**

National Institutes of Health (NIH), USA grants 2D43TW000233 and TW006990 (to A.U. and L.A.M.); Norwegian Financial Mechanism/EEA grant EE0016 and grant No SF0180060s09 from the Estonian Ministry of Education and Research (to A.U.).

**Conflicts of interest:** None declared.

**Key points**

- This study utilizing self-collected, home-obtained specimen collection for STI testing found that individuals who responded were less likely to have had medical care and more likely to have had a prior STI than non-responders.
- Late respondents appeared more like non-respondents on demographic factors, health care utilization patterns and potential disease status.
- Non-response in a health survey including biological material self-collection warrants research as it may differ from non-response in general health questionnaires.
- Integrating public health initiatives with population research components is feasible and may be particularly beneficial to ensure an economical public health approach.

**References**


Carling C. International questionnaire postal response rate: an experiment comparing no return postage to provision of International Postage Vouchers—‘‘Coupon-Re´ponse International’’. *BMC Health Serv Res* 2004;4:16.
