Preventive Medicine 53 (2011) 274–277

Contents lists available at ScienceDirect

Preventive Medicine

journal homepage: www.elsevier.com/locate/ypmed

# Non-response bias in a web-based health behaviour survey of New Zealand tertiary students☆

Kypros Kypri [a,b,*], Ari Samaranayaka [b], Jennie Connor [b], John D. Langley [b], Brett Maclennan [b]

[a] Centre for Clinical Epidemiology & Biostatistics, School of Medicine and Public Health, University of Newcastle, Callghan, NSW 2308, Australia
[b] Injury Prevention Research Unit, Department of Preventive & Social Medicine, University of Otago, Dunedin, New Zealand

ARTICLE INFO

Available online 30 July 2011

*Keywords:*
Internet
Web
Survey
Health
Risk behaviour
Non-response
Bias

ABSTRACT

*Objective:* There has been little investigation of non-response bias in web-based health surveys. We hypothesised that non-respondents have a higher prevalence of risk behaviours than respondents.
*Method:* In 2005, random samples of students aged 17–25 years from 12 New Zealand tertiary institutions (n = 7130) were invited to complete a web-based health behaviour survey, with three e-mail reminders. Early respondents (before 2nd reminder) were compared with late respondents (after 2nd reminder). Late respondents served as a proxy for non-respondents.
*Results:* 2607 students (37%) responded early, 676 (9%) responded late, and 3847 (54%) did not respond. There were differences between early and late respondents in high school binge drinking (38% vs 47%, p = 0.002) and non-compliance with physical activity guidelines (12% vs 18%, p = 0.004). Differences in overweight/obesity (26% vs 31%, p = 0.058), smoking (18% vs 22%, p = 0.091) and non-compliance with dietary guidelines (76% vs 77%, p = 0.651) were non-significant but point estimates were in the expected direction. Estimated bias in prevalence of risk behaviours was an absolute difference of 1–4% and a relative difference of 0–21%.
*Conclusion:* Respondents whose participation was hardest to elicit reported more risk behaviour. Assuming non-respondents' behaviour is similar or more extreme than that of late respondents, prevalence will have been substantially underestimated.



## Introduction

Falling survey response rates present a significant challenge for health research, primarily because of the increasing effects of selective non-response on estimates of the prevalence of health problems and risk behaviour. A typical approach to studying non-response bias is to undertake intensive follow-up of non-respondents and to compare estimates with those obtained using standard survey procedures (Wild et al., 2001).

An alternative is to compare respondents and non-respondents in surveys imbedded within larger studies (Van Loon et al., 2003). In one such study, involving a postal survey of cancer risk factors of individuals participating in a larger study of behavioural risk factors for chronic disease, smoking, physical inactivity, obesity, and poorer self-rated health were found to be more prevalent among non-respondents (Van Loon et al., 2003).

In a third paradigm, utilising archival records, mortality subsequent to postal and telephone health surveys has been found to be higher among non-respondents (Barchielli and Balzi, 2002; Cohen and Duffy, 2002), as have sickness absence rates (Martikainen et al., 2007) and hospital utilisation (Gundgaard et al., 2008; Kjoller and Thoning, 2005). These findings suggest that people with poorer health tend to avoid participating in health surveys.

There are, however, contrary findings which suggest context specific effects. For example, studies of respiratory health find that respondents have worse respiratory health than non-respondents (Hardie et al., 2003; Kotaniemi et al., 2001; Verlato et al., 2010). Perhaps in some contexts, less healthy people perceive a greater benefit in responding than healthier people.

Differences between respondents and non-respondents have been observed across postal, telephone, and face-to-face surveys. There has been a rapid increase in the use of web-based surveys but little is known about non-response bias in this modality.

A theoretical framework for studying respondent behaviour is the *continuum of resistance model,* which posits that willingness of individuals to participate can be inferred from the effort required to elicit participation (Lin and Schaeffer, 1995). Two methods are used to test the model. In the more commonly used approach, the sampling frame is used to compare the demographic characteristics of those who respond

☆ Sources of support: Health Research Council of New Zealand and Alcohol Advisory Council of New Zealand.
* Corresponding author. Fax: +61 2 4913 8148.
*E-mail address:* kypros.kypri@newcastle.edu.au (K. Kypri).

0091-7435 
doi:10.1016/j.ypmed.2011.07.017

**Table 1**
Respondent risk behaviour and status by response latency, New Zealand tertiary education institutions, 2005.

| | | All respondents % | Early respondents[+] % | Late respondents[++] % | 95% CI for difference (Early–Late) % |
|---|---|---|---|---|---|
| Binge drinker[*] in high school | Women | 39.3% (n = 1993) | 38.2% (n = 1602) | 44.5% (n = 391) | (−13.6, 1.2) p = 0.100 |
| | Men | 40.9% (n = 1281) | 38.7% (n = 999) | 51.0% (n = 282) | (−21.2, −3.4) p = 0.007 |
| | All | 39.9% (n = 3276) | 38.3% (n = 2602) | 47.1% (n = 674) | (−14.3, −3.1) p = 0.002 |
| Current smoker[**] | Women | 19.1% (n = 1952) | 17.5% (n = 1581) | 26.7% (n = 371) | (−15.6, −2.7) p = 0.005 |
| | Men | 17.9% (n = 1225) | 18.7% (n = 969) | 14.2% (n = 256) | (−1.4, 10.4) p = 0.133 |
| | All | 18.6% (n = 3179) | 17.9% (n = 2552) | 21.9% (n = 627) | (−8.5, 0.6) p = 0.091 |
| Not following fruit and vegetable consumption guidelines[#] | Women | 71.5% (n = 1982) | 71.1% (n = 1597) | 73.6% (n = 385) | (−9.3, 4.4) p = 0.482 |
| | Men | 83.6% (n = 1268) | 83.8% (n = 990) | 82.5% (n = 278) | (−5.7, 8.3) p = 0.710 |
| | All | 76.3% (n = 3252) | 76.0% (n = 2588) | 77.2% (n = 664) | (−6.1, 3.8) p = 0.651 |
| Not following physical activity guidelines[##] | Women | 12.9% (n = 1937) | 11.7% (n = 1571) | 18.7% (n = 366) | (−12.2, −1.8) p = 0.009 |
| | Men | 13.7% (n = 1220) | 13.0% (n = 962) | 17.2% (n = 258) | (−10.8, −2.4) p = 0.211 |
| | All | 13.2% (n = 3159) | 12.2% (n = 2535) | 18.1% (n = 624) | (−9.9, −1.9) p = 0.004 |
| Overweight or obese (Body Mass Index of 25 or more) | Women | 23.7% (n = 1862) | 23.1% (n = 1519) | 27.1% (n = 343) | (−10.9, 2.8) p = 0.248 |
| | Men | 31.1% (n = 1191) | 29.9% (n = 946) | 36.6% (n = 245) | (−15.3, 2.0) p = 0.134 |
| | All | 26.6% (n = 3055) | 25.8% (n = 2467) | 30.9% (n = 588) | (−10.4, 1.8) p = 0.058 |

[*] Men who drank more than 6 standard drinks (60 g ethanol)/women who drank more than 4 standard drinks (40 g ethanol) on a single occasion at least fortnightly in their last year of high school.
[**] Smokes occasionally or daily.
[+] Completed or partially completed the survey before the date of the second reminder email.
[++] Completed or partially completed the survey on or after the date of second reminder email.
[#] At least two servings of fruit and three servings of vegetables per day (Ministry of Health, 1998).
[##] At least 30 min per day (Ministry of Health, 1998).

versus those who do not respond. The method is useful to the extent that the usually minimal information in the sampling frame (often only demographic data) is correlated with the attributes of interest.

In the second approach, persons who respond only after considerable effort from the survey administrators – *late respondents* – are compared with early respondents. Differences in prevalence between early and late respondents serve as the basis for inferences about non-respondents, on the assumption that non-respondents lie beyond the late respondents on the continuum of resistance. The method requires accurate documentation of efforts to elicit, and the timing of, the survey response.

In one such study, a web-based survey of alcohol use at a New Zealand university, with 82% response (Kypri et al., 2004a), utilising several evidence-based methods (Edwards et al., 2002), late respondents drank more, had a higher prevalence of heavy drinking, and more alcohol-related problems than early respondents (Kypri et al., 2004b). On the basis of these studies, we hypothesised that people who do not comply with health guidelines on drinking, smoking, diet and physical activity, and have greater body mass, would be less inclined to participate in a health behaviour survey.

## Methods

### *Sample*

New Zealand has eight universities and 19 polytechnic colleges which provide vocational training and some degree courses. All eight universities were invited to participate in a web-based study, and five accepted,

**Table 2**
Prevalence rates adjusted for estimated non-response bias, New Zealand tertiary education institutions, 2005.

| | Prevalence estimate assuming non-respondents report the same prevalence as late respondents | Absolute difference in prevalence[a] (%) | Relative difference in prevalence[b] (%) |
|---|---|---|---|
| Binge drinker in high school | 44.3% (n = 7123) | 4.4 | 11 |
| Current smoker | 21.8% (n = 7026) | 3.2 | 17 |
| Not following fruit and vegetable consumption guidelines | 77.2% (n = 7099) | 0.9 | 0.1 |
| Not following physical activity guidelines | 16.0% (n = 7006) | 2.8 | 21 |
| Overweight or obese (Body Mass Index of 25 or more) | 31.0% (n = 6902) | 4.4 | 17 |

[a] Formula: Absolute difference = adjusted prevalence estimate − raw prevalence estimate.
[b] Formula: Relative difference = (adjusted prevalence estimate/raw prevalence estimate) − 1.

representing six campuses (one of them providing data from two campuses in different cities). Ten of the polytechnic colleges were invited to participate in order to maximise geographic coverage of the country for a study aimed at examining environmental determinants of various health behaviours (i.e., polytechnics in the same cities as universities were not invited). Six of the invited polytechnics accepted, bringing the total number of tertiary education institutions involved in the study to 12.

Māori (the indigenous people of New Zealand) comprise 15% of the New Zealand population, 10% of university students and 18% of polytechnic students (Ministry of Education, 2011). We sought to invite random samples of 430 Māori and 430 non-Māori students aged 17–25 years from each campus in order to maximise the explanatory power of the study for Māori, who have traditionally been poorly served by population surveys despite bearing a considerably greater disease burden (Wellington School of Medicine and Health Sciences, 2002). There was no stratification of the samples by age and sex. All members of the study population had an institution assigned e-mail address which we used to issue the invitation to participate. The questionnaire was offered in Māori and English and users could switch between languages at any stage by clicking a button.

*Procedure*

Students were invited by personalised letter to complete a web survey of their alcohol use, using a procedure described in detail elsewhere (Kypri et al., 2004a, 2009). Sample weighting was used to account for the proportions of Māori and non-Māori at each campus.

*Alcohol use*

Respondents indicated how often they engaged in *binge* drinking (women: >4 drinks, men: >6 drinks) during their last year of high school. Standard drink definitions (10 g ethanol) were provided with pictures (e.g., a glass of beer) and the number of drinks in typical containers.

*Smoking*

Respondents selected a descriptor for their cigarettes use: "Never smoked or never smoked regularly", "Do not smoke now but used to smoke", "Occasionally smoke (on average, <1/day)", "Currently smoke cigarettes regularly (≥1/day)".

*Fruit and vegetable intake*

Respondents indicated how many servings of fruit (fresh, frozen, canned or stewed) and how many servings of vegetables (fresh frozen, canned) they ate per day. Examples were given to illustrate serving sizes.

*Physical activity*

Respondents indicated separately for weekdays and weekends how much time they were physically active, including walking to campus or shops, housework, shopping, sport, and exercise.

*Body mass index*

Respondents indicated their height in metres or feet and inches and their weight in kilograms or pounds.

There were a total of 78 questions in the questionnaire though it should be noted that with branching and skip patterns most participants (e.g., non-drinkers) will not have been presented with all of the questions.

## Results

Of 7130 students invited, 3283 (46%) participated. University response rates ranged from 53% to 72% (63% overall) while polytechnic response rates ranged from 15% to 36% (24% overall). Response did not vary by age and gender, but Māori were less likely to participate (42%) than non-Māori (48%; $p<0.001$).

Table 1 summarises risk behaviour and overweight/obesity prevalence, by gender, as a function of latency to response. Late respondents were significantly more likely to be binge drinkers in high school and to be physical inactive. The differences for being overweight/obese, smoking, and diet were in the expected direction but non-significant. We conducted the analyses separately for the polytechnic colleges versus universities finding results that were consistent for all five parameters so we have reported only the combined results.

Table 2 shows prevalence estimates adjusted under the assumption that non-respondents have the same prevalence of these behaviours as late respondents, and the extent of non-response bias in absolute and relative terms.

## Discussion

Late respondents had a higher prevalence of binge drinking and non-compliance with physical activity guidelines. Differences in the prevalence of non-compliance with dietary guidelines, smoking and overweight/obesity were non-significant but in the expected direction. The apparent non-response bias for binge drinking was mainly driven by differences among men. For physical activity, the effects were mainly driven by differences among women. Notably, smokers were significantly over-represented among female late respondents even though the overall result was non-significant. Assuming that non-respondents have similar behavioural characteristics and body mass to late respondents, the true prevalence of these characteristics will have been under-estimated. The estimated bias in terms of absolute difference in prevalence was 1–4% and 0–21% in relative terms.

Limitations include the self-report of behaviour and height/weight. It is possible that misreporting is correlated with latency to respond. For such a pattern to bias the findings toward the study hypothesis, late respondents would have to have been less likely than early respondents to understate their drinking and compliance with physical activity guidelines, which seems unlikely. It is also possible that the findings from this young population group do not generalise to the wider population.

The response rates were markedly lower for the polytechnic colleges than the universities. While all students ostensibly had access to e-mail and the Internet, it is possible that in 2005 students at polytechnic colleges, which offer vocational training (e.g., forest management) as well as degree courses (e.g., nursing), used their e-mail and the Internet less than university students and were therefore less used to interacting via this medium.

The results are consistent with previous research using the web-based method at a single university examining alcohol use alone (Kypri et al., 2004b), and with the findings of a pen-and-paper survey of a national household sample of alcohol use and intimate partner violence (Meiklejohn, 2010). In both of those studies, late respondents drank more than early respondents. In the latter study, the prevalence of binge drinkers in the New Zealand population was underestimated by 4.0 percentage points (17.6 vs. 21.6%) or 19% in relative terms. Also consistent with other studies are findings showing that late respondents tend to have a higher prevalence of smoking (Korkeila et al., 2001; Tolonen et al., 2005; Van Loon et al., 2003; Verlato et al., 2010) overweight/obesity (Tolonen et al., 2005; Van Loon et al., 2003) and physical inactivity (Van Loon et al., 2003).

The findings suggest that non-response bias seen in telephone, postal, and face-to-face surveys is also present in the web-based modality. Estimates of health compromising behaviours from surveys should be generally considered under-estimates and the degree of under-estimation probably worsens with lower response rates. Variability in the degree of bias according to health behaviour, and by gender, seen in this study suggests that simple adjustment of estimates to correct for non-response error e.g., post-weighting to the population, is likely to introduce error, by magnifying existing non-response biases in the data. Urgent work is needed to increase response rates in population health behaviour surveys.

## Author contributions

KK designed and oversaw the implementation of the study. KK and JL obtained funding. AS conducted the analysis. All authors contributed to interpretation of the results. KK led the writing of the paper and all authors contributed to and approved the final version of the paper.

**Conflict of interest statement**
The authors declare they have no conflict of interest.

## References


Barchielli, A., Balzi, D., 2002. Nine-year follow-up of a survey on smoking habits in Florence (Italy): higher mortality among non-responders. Int. J. Epidemiol. 31, 1038–1042.

Cohen, G., Duffy, J.C., 2002. Are nonrespondents to health surveys less healthy than respondents? J. Off. Stat. 18, 13–23.

Edwards, P., Roberts, I., Clarke, M., DiGuiseppi, C., Pratap, S., Wentz, R., Kwan, I., 2002. Increasing response rates to postal questionnaires: systematic review. Br. Med. J. 324, 1183.

Gundgaard, J., Ekholm, O., Hansen, E.H., Rasmussen, N.K., 2008. The effect of non-response on estimates of health care utilisation: linking health surveys and registers. Eur. J. Public Health 18, 189–194.

Hardie, J.A., Bakke, P.S., Mørkve, O., 2003. Non-response bias in a postal questionnaire survey on respiratory health in the old and very old. Scand. J. Public Health 31, 411–417.

Kjoller, M., Thoning, H., 2005. Characteristics of non-response in the Danish Health Interview Surveys, 1987–1994. Eur. J. Public Health 15, 528–535.

Korkeila, K., Suominen, S., Ahvenainen, J., Ojanlatva, A., Rautava, P., Helenius, H., Koskenvuo, M., 2001. Non-response and related factors in a nation-wide health survey. Eur. J. Epidemiol. 17, 991–999.

Kotaniemi, J.-T., Hassi, J., Kataja, M., Jonsson, E., Laitinen, L.A., Sovijarvi, A.R.A., Lundback, B., 2001. Does non-responder bias have a significant effect on the results in a postal questionnaire study? Eur. J. Epidemiol. 17, 809–817.

Kypri, K., Gallagher, S.J., Cashell-Smith, M.L., 2004a. An internet survey method for college student drinking research. Drug Alcohol Depend. 76, 45–53.

Kypri, K., Stephenson, S.C.R., Langley, J.D., 2004b. Assessment of non-response bias in an internet survey of alcohol use. Alcohol. Clin. Exp. Res. 28, 630–634.

Kypri, K., Paschall, M.J., Langley, J., Baxter, J., Cashell-Smith, M., Bourdeau, B., 2009. Drinking and alcohol-related harm among New Zealand university students: findings from a national Web-based survey. Alcohol. Clin. Exp. Res. 33, 307–314.

Lin, I.-F., Schaeffer, N.C., 1995. Using survey participants to estimate the impact of nonparticipation. Public Opin. Q. 59, 236–258.

Martikainen, P., Laaksonen, M., Piha, K., Lallukka, T., 2007. Does survey non-response bias the association between occupational social class and health? Scand. J. Public Health 35, 212–215.

Meiklejohn, J., 2010. Drinking patterns, drinking in partnerships and informal social controls on drinking in New Zealand. Master of Public Health Thesis. In: *Department of Preventive & Social Medicine, University of Otago*, Vol. Master of Public Health, Dunedin.

Ministry of Education, 2011. Education counts. http://educationcounts.edcentre.govt.nz/statistics/tertiary/index.html2011Accessed 22 June 2011. Ministry of Education, New Zealand Government.

Ministry of Health, 1998. Taking the Pulse: The 1996/97 New Zealand Health Survey, Report HP3261. Author, Wellington.

Tolonen, H., Dobson, A., Kulathinal, S., 2005. Effect on trend estimates of the difference between survey respondents and non-respondents: results from 27 populations in the WHO MONICA project. Eur. J. Epidemiol. 20, 887–898.

Van Loon, A.J.M., Tijhuis, M., Picavet, H.S.J., Surtees, P.G., Ormel, J., 2003. Survey non-response in the Netherlands: effects on prevalence estimates and associations. Ann. Epidemiol. 13, 105–110.

Verlato, G., Melotti, R., Olivieri, M., Corsico, A., Bugiani, M., Accordini, S., Villani, S., Migliore, E., Marinoni, A., Pirina, P., Carrozzi, L., Bortolami, O., Rava, M., de Marco, R., 2010. Asthmatics and ex-smokers respond early, heavy smokers respond late to mailed surveys in Italy. Respir. Med. 104, 172–179.

Wellington School of Medicine and Health Sciences, 2002. Mana Whakamarama- Equal Explanatory Power: Maori and Non-Maori Sample Size in National Health Surveys. Public Health Intelligence, Wellington.

Wild, T.C., Cunningham, J., Adlaf, E., 2001. Non-response in a follow-up to a representative telephone survey of adult drinkers. (Statistical Data Included). J. Stud. Alcohol 62, 257.