Public Health Volume 120, Issue 4 , April 2006, Pages 297-308

doi:10.1016/j.puhe.2005.09.008
Copyright © 2005 The Royal Institute of Public Health Published by Elsevier Ltd.

Non-response in a survey of cardiovascular risk factors in the Dutch population: Determinants and resulting biases

H.C. Boshuizen[a,*], A.L. Viet[b], H.S.J. Picavet[b], A. Botterweck[c] and A.J.M. van Loon[b]

[a]Department of Information Technology and Methodology, National Institute for Public Health and the Environment, P.O. Box 1, 3720 BA Bilthoven, The Netherlands
[b]Department of Prevention and Healthcare Research, National Institute for Public Health and the Environment, Bilthoven, The Netherlands
[c]Statistics Netherlands, Division of Social and Spatial Statistics, Statistical Analysis, Heerlen, The Netherlands

[*] Corresponding author. Tel.: +31 30 274 2944; fax: +31 30 274 4456.

Summary

Background

The aim of the research was to study the determinants of participation in a health examination survey (HES) which was carried out in a population that previously participated in a health interview survey (HIS) of Statistics Netherlands, and to estimate the effect of non-participation on both the prevalence of the main HES outcomes (risk factors for cardiovascular disease) and on relationships between variables.

Methods

Logistic regression was used to study the determinants of participation in the HES (n=3699) by those who had previously participated in the HIS (n=12 786). Linear models were used to predict the main outcomes in non-participants of the HES. Item non-response was handled by multiple imputation.

Results

HES participants had a higher socio-economic status and comprised more 'worried well', while the rural population were less likely to participate in the HES. Most predicted values of outcomes in HES non-participants differed from those in HES participants, but much of this was due to differences in the age and gender composition of both groups. Taking age and gender differences into account, most predicted values of outcomes in the entire HIS population were within the 95% confidence intervals of the HES values, with the exception of body height in men and high-density lipoprotein cholesterol, fasting glucose and body weight in women. These differences are most likely to be due to the higher socio-economic status of HES participants. Relationships between HIS variables did not change significantly when using HES participants alone compared with all HIS participants.

Public Health Volume 120, Issue 4 , April 2006, Pages 297-308

Conclusions

Despite a high rate of non-participation, some bias, mostly small, was seen in the prevalence rates of the main outcome variables. Bias in the relationships between variables was negligible.

Introduction

Monitoring of endogenous risk factors for cardiovascular disease (blood pressure, cholesterol, weight etc.) in a population is carried out by means of health examination surveys (HES). One approach is to link a HES to an existing health interview survey (HIS) by recruiting HIS respondents for participation in the HES. This method was used in the US NHANES study, and was adopted in 1998 in the Regenboog Study (Risk Factors and Health in The Netherlands; a Survey in Municipal Health Services). Linking the HES to the HIS carried out by Statistics Netherlands made it possible to use a nationwide representative population sample. Several studies have been undertaken on the non-response to the Dutch HIS.[1, 2, 3, 4 and 5] In agreement with other studies on non-response,[6, 7, 8, 9, 10 and 11] they have found that non-responders to the HIS have a lower socio-economic status and a lower healthcare use than responders.

In this HIS/HES design, participants had several opportunities to decline to participate. At the end of a lengthy interview, they were asked to give consent for their address to be forwarded to the organization carrying out the HES. After being invited to participate, they could have decided not to. Finally, they had to visit the municipal health centre after accepting this invitation. As such, only 29% of the interviewed HIS participants participated in the HES.

Relatively little is known about selection mechanisms in second-stage non-response in a HIS/HES design. This selection takes place in a population that has already been willing to participate in health research. Therefore, other mechanisms may operate compared with those operating when an individual refuses an initial interview. Also, the reasons for being willing to participate in a physical examination could differ from those for being willing to respond to a questionnaire survey. A few studies have investigated non-participation in the HES part of NHANES,[12] and have indicated a higher response rate among the 'worried well' (those with high healthcare use despite good health), those with more cars in the household, and those living near the location of the examination. In the Welsh Heart Health Survey,[13] subjects who consented to a medical examination had a higher socio-economic status and a healthier lifestyle (diet, physical activity) than non-consenters. However, as the participation rates in NHANES were much larger than in the present study, the non-participation mechanisms may differ. In the Welsh Heart Health Survey, the consent rates were of equal magnitude to the examination consent rates in the first stage of the present study, but only differences between consenters and non-consenters were studied. In the present study, only half of the consenters actually participated.

As extensive health information was available in the present study for all subjects who participated in the HIS, regardless of whether they participated in the HES, it was possible to study the determinants and consequences of second-stage non-participation.

The aims of this paper are:

1. to describe the determinants of non-participation in the HES amongst HIS responders;

2. to quantify how estimates of the outcomes monitored are influenced by selective non-response; and

3. to describe the magnitude of bias due to non-response in associations as observed in HES participants by comparing associations between HIS variables in analyses using all HIS participants, and analyses using HES participants alone.

Methods

Survey design

The HIS carried out by Statistics Netherlands is a nationwide survey on health, use of health services and occupational health. The population comprised HIS participants (1998–2000) aged at least 12 years in regions with municipal health centres participating in the HES (n=12,786). At the end of a full interview, they were asked for their consent for an additional health examination. In The Netherlands, 46 municipal health centres supply public health services to the entire population, including rural areas. Most centres commenced participation in the latter half of 1998. The number of participating municipal health centres increased from five in May 1998 to 42 in December 2000. Consent was obtained from 7431 of 12786 (58%) people who were subsequently invited by telephone to the municipal health centre for a physical examination. In total, 3699 subjects participated in the HES (29%).

Health interview survey

Detailed information on the design of the HIS can be found elsewhere.[14] In short, a multistage probability sample was drawn of the entire Dutch population after stratification by geographic area and degree of urbanization. The sampling frame excluded those living in institutions (prisons, homes for the elderly, mentally handicapped etc.). Only subjects fluent in Dutch were included and they were visited at home by the interviewer. In the first 4 months of 2004, lack of fluency in Dutch was the reason for non-response in 0.4% of those of Dutch descent or second-generation immigrants in the sample frame, for 7.9% of first-generation immigrants from Western countries, and for 18.2% of immigrants from other countries,[15] so this criterion will have decreased the presence of ethnic-minority groups in the HIS participants. Those who refused to take part or who could not be contacted after three visits were contacted by telephone in order to administer, if possible, a short interview and (in case of no earlier contact) to make an appointment for a full interview.

During 1998–2000, the average response to the full HIS interview was 56.2% (of those in the sample frame), and 4.3% only answered the short telephone questionnaire. Only the respondents who participated in the full interview are included in the present paper. The full HIS interview used two data-collection methods: a computer-aided face-to-face interview and a self-administered mail-back questionnaire (for the more personal questions on health). Due to the computer-aided interview format, item non-response was less than 1% for most items. Exceptions were questions on smoking (1.1%, including inconsistent answers), educational level (1.2%), body weight (3.4%) and household income (22.4%). The mail-back questionnaire, which included questions on the presence of chronic diseases such as diabetes, was returned by 75% of the 12,786 interviewees.

Health examination survey

Participants completed an informed consent form. When the examination was scheduled before 10.30 a.m., participants were asked to fast from 10.00 p.m. the day before. The examination was carried out by a trained nurse and consisted of the measurement of blood pressure (duplicate measurement with the automatic Omron HEM 711 with 5 min of rest before and between the measurements), weight, height, waist and hip circumferences, and 10 joint function tests. Blood samples were collected for measuring fasting glucose, total cholesterol and high-density lipoprotein (HDL) cholesterol. The number of missing values was <1% in the direct measurements and 8% in the laboratory values. Fasting glucose concentration was not present for 37%, mainly because of participants examined after 10.30 a.m., but also because some participants forgot to fast.

This paper is restricted to the main outcomes: blood pressure, weight, height, total and HDL cholesterol, and fasting glucose.

Statistical analyses

Using existing knowledge and preliminary analyses, 86 variables from the HIS were selected for this analysis. Variables selected were either potential determinants of non-response or of one of the outcomes of interest (Table 1).

Public Health Volume 120, Issue 4 , April 2006, Pages 297-308

Table 1.

Areas from which items were used in the analyses (sometimes multiple variables were used per area).

Sociodemographic	**Lifestyle and antro-pometry**	**Work**	**Health state/wellbeing**	**Healthcare use**
Age	Smoking	Number of hours paid work (S+HH)	Problems with ADL	Alternative health care
Gender	BMI (calculated from self-reported weight and height)	Receives social benefit (S+HH)	Long-standing illness	Use of OTC drugs during last 2 weeks
Education (S+HH)	Weight (self-reported)	Occupational class (S+HH)	Limiting longstanding illness	Type of OTC drugs
Type of household	Height (self-reported)	Shift/evening/weekend work	Diabetes	Use of prescribed drugs during last 2 weeks
Number of persons in household	Alcohol use	Heavy work	Self-perceived health	Type of prescribed drugs
Household income	Hours doing sport per week	Noise exposure at work	Satisfied with life	Visited specialist in last 2 months
Marital status	Hours cycling per week	Stress at work	Happiness	Type of specialist visited
Nationality	Hours walking per week	Has disability pension		Visited general practitioner in last 2 weeks
Type of health insurance		S is main wage earner (=highest income)		General practitioner prescribes alternative treatments
Month of HIS participation				Nights in hospital in last year
Religion				Sickness leave in last 2 months
				Uses mental health care

S, subject; HH, head of household (person in household who earns the highest wage); ADL, activities of daily living; OTC, over the counter; HIS, health interview survey; BMI, body mass index.

In order to use all available information and to prevent further bias (e.g. caused by selective refusal to answer the question on income), the authors dealt with item non-response in the HIS by multiple imputation,[16] carried out with an adapted version of predictive mean matching[17] with a SAS macro making partial use of SAS PROC MI. The macro can be obtained from the authors. Multiple imputations yielded five versions of the datasets, in which the non-missing data are the same but the values imputed for the missing data vary from dataset to dataset. Analyses were performed five times, and the results were pooled using methods yielding valid confidence intervals (taking the uncertainty due to missing data into account).

To describe the determinants of non-participation in the HES by HIS responders, logistic regression was used with participation as the dependent variable. Stepwise backward selection was used, pooling the parameter values of the five imputed datasets after each selection step. The pool of variables contained all variables derived from Table 1, fractional polynomial terms for age, and interaction terms between gender and these polynomial age terms. After a first selection step using $P>0.05$ as the criterion for removal, interaction terms were added between all the remaining variables, and the selection procedure was repeated with this set and $P>0.01$ as criterion. $P>0.01$ was used rather than $P>0.05$ as the final criterion in order to reduce the number of false-positive findings. If an interaction term was retained, its component terms were added to the model to improve interpretability. The indicator variable 'holds paid job' was also forced into the model to improve interpretation because many variables are only applicable to those who work (e.g. stress at work, type of occupation, number of hours worked). With the indicator variable in the model, the odds ratio for, for example, hours worked can be interpreted as the odds ratio of an extra hour worked within the group of those holding jobs.

To quantify bias in the outcomes, the average outcome values (blood pressure, weight, height, total and HDL cholesterol, and fasting glucose) were predicted in HES non-participants. The prediction model (a linear regression model) was fit on the HES participants, using different covariates for each outcome. Model selection was carried out using backward selection, pooling the parameter values of the five imputed datasets after each selection step and using the pooled parameter values from the last step to predict the outcome values. Cross-validation was used to determine which P value for removal gave the best predictions. The pool of variables from which to select the model contained fractional polynomial terms for age and self-reported body weight, as well as interaction terms between gender and the polynomial age terms.

The pool of variables for predicting height and weight also contained fractional polynomial terms of self-reported body height; for predicting the other outcomes, the pool contained fractional polynomial terms of self-reported body mass index. Both crude average outcomes and age-standardized, sex-specific average outcomes were compared between HES participants and HES non-participants.

The magnitude of bias in associations between variables due to non-response may be small even if there is bias in the prevalence of averages of the individual variables. This bias in associations was studied by comparing associations between HIS variables in the entire HIS population with those associations in HES participants alone. For this analysis, three 'exposure' HIS variables and three 'outcome' HIS variables that were strongly associated with non-participation were selected in order to represent the worst-case scenarios.

Results

Determinants of non-participation

Figure 1 shows that the rate of participation increased gradually with age until approximately 60 years of age, and decreased strongly with age thereafter. There was some interaction with gender, which disappeared after inclusion of other covariates.

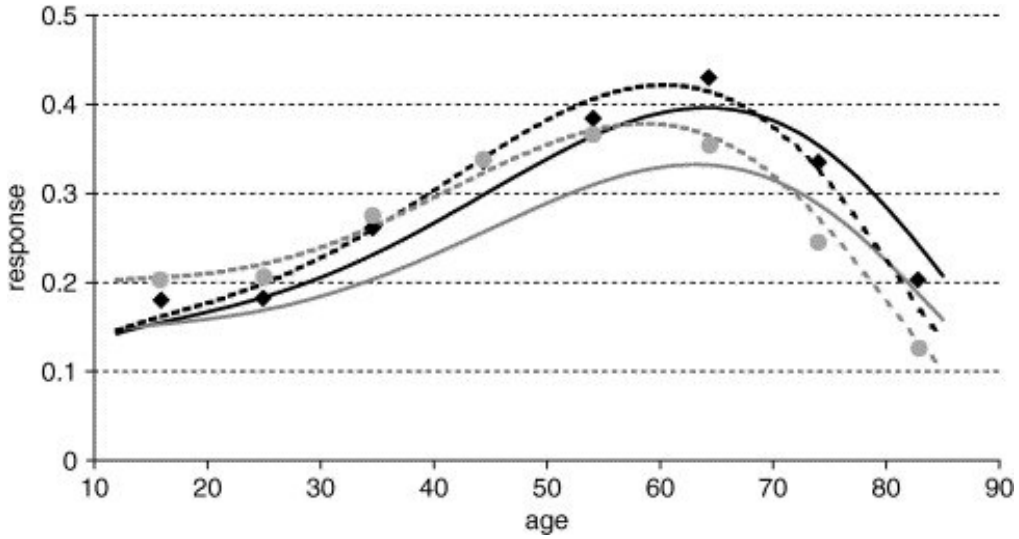

Figure 1. Participation rates in the health examination survey among eligible health interview survey respondents by age and gender. Diamonds, crude rate for men; circles, crude rate for women; broken line, fitted values (unadjusted); solid line, fitted values adjusted for all variables in Table 2; black, men; grey, women.

Table 2 shows the variables that were significantly related to non-response after adjusting for all other variables related to non-response. Although this table includes many variables, many more are not included as they were not associated with non-response. The table shows that socio-economic variables are important determinants of HES participation, with those from lower socio-economic strata having lower participation rates. When adjusting for all other variables, including occupation and income, differences in participation rates are only seen for those with either a very high or a very low educational level. People without paid work were less likely to participate than those with a part-time job, but of those with a job, the participation rate decreased with the number of hours worked. The results of the multivariate analysis imply that, other things (including health) being equal, those working in unskilled work or working for at least 22 h/week in skilled work were less likely to participate than those without a job. Women, Muslims and people living in rural areas were less likely to participate, while those from the western part of the country (where the large cities are located) were more likely to participate.

Public Health Volume 120, Issue 4 , April 2006, Pages 297-308

Table 2.

Health examination survey (HES) participation rates among eligible health interview survey (HIS) participants, and multivariate odds ratio (adjusted for all other variables in the table and age) of participating in the HES.

Crude variable	**Level**	***n***	**Participation rate (%)**	**Model category**	**Adjusted odds ratios of participating**
All	All	12,786	28.9		
Gender	Male	6235	29.5	Male	1.00
	Female	6551	28.4	Female	0.82 [0.75–0.90]
Type of occupation	Skilled work[a]	4963	31.6	Skilled[a]	1.00 (reference)
	Unskilled work	2702	24.6	Unskilled	0.76 [0.67–0.88]
	Occupational status				
Occupational status	In work force	7665	29.1	In workforce	1.00
	Not in workforce	5121	28.6	Not in workforce	0.83 [0.71–0.97]
Number of hours worked per week	Not in workforce	5121	28.6	Per 10 h worked (only when working)	0.87 [0.83–0.91]
	1–4	203	33.5		
	5–12	727	26.1		
	13–29	1623	32.8		
	30–39	2162	30.8		
	40 or more	2950	26.4		
Religion	Muslim	294	9.9	Muslim	0.57 [0.38–0.86]
	Other/none	12,492	29.4	Other/none	1.00
Educational level	Primary education	2904	21.7	Primary education	0.81 [0.71–0.92]
	3 years of secondary education	3375	28.8	3 years of secondary education	0.96 [0.86–1.07]
	High school	4079	30.7	High school	1.00
	Bachelor's degree	1713	35.0	Bachelor	1.01 [0.89–1.14]
	Master's degree	626	35.5	Master	0.97 [0.80–1.17]
	PhD	88	24.0	Ph D	0.55 [0.33–0.92]
Decile of household income	10th (highest income)	1258	32.4	Highest vs lowest decile[b] in[c]:	
	9th	1230	32.6		
	8th	1256	30.9	Those who do not participate in sports	1.64 [1.33–2.01]
	7th	1275	30.0	Those who participate in sports	1.02 [0.41–2.54]
	6th	1296	31.5		
	5th	1205	29.4		
	4th	1325	28.7		
	3rd	1387	27.5		
	2nd	1244	25.0		
	1st	1310	21.7		
Urbanity	Very strong	1539	30.2	Very strong	0.98 [0.85–1.13]
	Strong	3857	31.5	Strong	1.00
	Moderate	2889	28.8	Moderate	0.83 [0.74–0.93]
	Little	2797	26.7	Little	0.75 [0.67–0.85]
	None	1704	25.9	None	0.74 [0.65–0.85]
Region	Western part	4925	31.4	Western part	1.15 [1.05–1.25]

Crude variable	Level	*n*	Participation rate (%)	Model category	Adjusted odds ratios of participating
	Other	7861	27.4	Other	1.00
Long-term health condition	No	8369	28.0	No	1.00
	Yes	4417	30.6	Yes	0.87 [0.79–0.97]
ADL (only available for age >54 years)	Age <55 years	9294	27.1	Odds ratio per extra ADL point	0.94 [0.91–0.98]
	No problems	2396	37.4		
	Score 1–2	497	30.6		
	Score 3+	599	21.7		
Self-rated health	Very good	2755	30.1	Very good	1.09 [0.98–1.21]
	Good	7097	29.4	Good	1.00
	Reasonable	1722	28.0	Reasonable	0.85 [0.74–0.97]
	Sometimes good, sometimes bad	866	27.0	Sometimes good, sometimes bad	0.87 [0.73–1.05]
	Bad	346	18.8	Bad	0.59 [0.43–0.81]
Last visit to medical specialist	<2 months ago	2089	29.7	<2 months ago	1.03 [0.91–1.17]
	2–12 months ago	3067	31.0	2–12 months ago	1.05 [0.95–1.17]
	>12 months ago	6690	29.2	>12 months ago	1.00
	Never visited a medical specialist	940	18.5	Never visited a medical specialist	0.76 [0.63–0.91]
Visit to internist in last 2 months	Yes	257	23.7	Yes	0.66 [0.48–0.91]
	No	12,529	29.0	No	1.00
Last visit to GP	<2 months ago	4708	30.5	<2 months ago	1.23 [1.09–1.39]
	2–12 months ago	5169	29.1	2–12 months ago	1.13 [1.01–1.27]
	>12 months ago	2884	26.3	>12 months ago	1.00
	Never visited a GP	25	4.0	Never visited a GP	0.14 [0.02–1.10]
Visited alternative medical practitioner in last 12 months	No	11,951	28.5	No	1.00
	Yes	835	35.7	Yes	1.25 [1.07–1.47]
Sports participation	No	5919	25.7	In lowest income	1.73 [0.69–4.3]
	Yes	6867	31.7	In highest income	1.08 [0.95–1.22]
Returned self-administered HIS questionnaire	Yes	9547	34.8	Yes	1.00
	No	3239	11.7	No	0.29 [0.25–0.32]

ADL, activities of daily living; GP, general practitioner.
[a] Including the self-employed.
[b] Calculated from quadratic variable in model.
[c] Due to significant interaction between income and doing sport, odds ratios are calculated within strata of the other variable.

The relationship between health and participation rates is complex. Generally, those who used more health care (more frequent visits to general practitioners, medical specialists, alternative medical practitioners) were likely to participate in the HES, while those with better health (as measured by self-perceived health status and the absence of longstanding illness) were also likely to participate, leading to an overrepresentation of the 'worried well' (i.e. healthy people with frequent healthcare use). However, those who had visited an internist during the previous 2 months did not fit this pattern, as they were less likely to participate. This may indicate that visiting an internist (which is only possible after referral by a general practitioner in the Dutch healthcare system) reflects worse health rather than a tendency to use healthcare services more frequently.

Large differences in participation rates were seen between those who did and those who did not return the mail-back questionnaire that was part of the HIS.

Biases in outcomes to be monitored

Table 3 shows the strongest predictors for the seven outcomes in this study. As the prediction model contained many predictors, only the strongest predictors are shown. The percentage of variance explained by the prediction models was high (>90%) for height and weight (not surprising as self-reported height and weight were among the predictors), intermediate for systolic blood pressure (45%), and moderate for the other variables (24–29%).

Table 3.

Predictors with $P<0.001$ in the models for health examination survey (HES) outcomes; all predictors are based on self-reported data from the health interview survey (HIS).

HES outcome	**Predictors with *P*<0.001**	**Variance predicted by the model (%)**
Total cholesterol	Age, interaction age and gender, gender, BMI, diabetes, number of persons in the household, height[a]	25.7
High-density lipoprotein cholesterol	Gender, weight, alcohol on week days (number of glasses), frequency of alcohol use, height[a], number of cigarettes smoked, interaction between gender and age, age	27.8
Fasting glucose	Diabetes, weight[a], height[a], BMI[b], age, gender, 'other' living arrangement (not with family/partner/children or alone)	23.9
Diastolic blood pressure	Age, ever smoker, alcohol on week days (number of glasses)	28.6
Systolic blood pressure	Age, BMI[b], height[a]	44.5
Body height	Height[a], educational level, age, BMI[b]	91.2
Body weight	Weight[a], age	92.1

BMI, body mass index.
[a] As self-reported in the HIS interview.
[b] Calculated from self-reported height and weight in the HIS.

Table 4 compares the average value of the seven outcomes as measured in the HES participants with the predicted value in non-participants. Crude predicted values for the general population were lower than the lower boundary of the 95% confidence interval of the HES average for three of the five outcomes (blood pressure, cholesterol and body weight).

Table 4.

Outcomes in all health examination survey (HES) respondents, population SD in HES responders, and predicted outcomes in all health interview survey (HIS) responders eligible for HES participation who did not participate.

	Mean value measured in HES [95% CI]	**Population SD in HES**	**Predicted in HES non-responders[a]**
Total cholesterol	5.26 [5.23–5.30]	1.08	5.12
High-density lipoprotein cholesterol	1.25 [1.24–1.26]	0.34	1.23
Fasting glucose	4.81 [4.77–4.85]	1.00	4.79
Diastolic blood pressure	81.6 [81.2–82.0]	11.2	79.8
Systolic blood pressure	130.4 [129.7–131.1]	21.8	128.3
Body height	172.8 [172.5–173.1]	9.4	172.5
Body weight	75.2 [74.7–75.7]	14.5	73.8

CI, confidence intervals; SD, standard deviation.
[a] Using the outcome model described in Table 3.

However, age-standardized and sex-specific figures are more relevant in practical situations (Table 5). Here most estimates are within the 95% confidence intervals of the HES measurements, with the exception of body height in men (higher in participants) and HDL-cholesterol (higher in participants), fasting glucose levels (lower in participants) and body weight in women (lower in participants), indicating that the differences in Table 4 reflect that participants and non-participants differ in age and gender.

Table 5.

Age-standardized outcomes in health examination survey (HES) participants compared with predicted outcomes in non-participants.

	Men		**Women**	
	HES participants	**HES non-participants**	**HES participants**	**HES non-participants**
Total cholesterol	5.10 [5.05–5.15]	5.11	5.17 [5.12–5.23]	5.19
High-density lipoprotein cholesterol	1.12 [1.10–1.13]	1.10	1.37 [1.36–1.39]	1.35
Fasting glucose	4.88 [4.83–4.94]	4.92	4.64 [4.58–4.69]	4.70
Diastolic blood pressure	80.8 [80.3–81.4]	80.7	79.4 [79.0–80.0]	79.7
Systolic blood pressure	132.0 [131.0–133.0]	131.8	125.4 [124.5–126.4]	125.7
Body height	179.4 [179.0–179.7]	178.7	166.6 [166.3–167.0]	166.3
Body weight	80.0 [79.4–80.7]	79.4	67.9 [67.3–68.5]	68.9

Biases in associations between variables

The odds ratios for the associations between some HIS variables that are heavily influenced by non-participation were compared between analyses in all HIS participants and in HES participants alone (Table 6). This table shows that the 95% confidence intervals of the odds ratios calculated using HES participants

alone largely overlap those calculated using the entire HIS population, so no clear bias seems to result from non-participation.

Table 6.

Odds ratios between outcomes subject to selective non-response [having poor health, a longstanding illness, or having visited a general practioner (GP) in the last 2 months] and several factors also subject to selective non-response.

	Poor self-perceived health		**Longstanding illness**		**Visit to general practitioner in last 2 months**	
	HIS (*n*=12 786)	**HES (*n*=3699)**	**HIS (*n*=12 786)**	**HES (*n*=3699)**	**HIS (*n*=12 786)**	**HES (*n*=3699)**
Occupation						
Unskilled work	2.13 [1.28–3.54]	2.21 [0.74–6.57]	1.28 [1.15–1.44]	1.23 [1.00–1.52]	1.20 [1.07–1.32]	1.22 [1.00–1.50]
Educational level						
Primary education	2.76 [2.04–3.73]	3.80 [1.82–7.92]	1.61 [1.43–1.80]	1.56 [1.26–1.95]	1.42 [1.26–1.58]	1.34 [1.08–1.67]
3 years of secondary education	1.20 [0.86–1.66]	2.03 [0.97–4.24]	1.12 [1.01–1.24]	1.25 [1.04–1.49]	1.14 [1.03–1.26]	1.19 [0.99–1.42]
High school	1.00	1.00	1.00	1.00	1.00	1.00
Bachelor's degree	0.82 [0.53–1.27]	1.15 [0.45–2.96]	0.88 [0.77–0.99]	0.95 [0.77–1.18]	1.00 [0.88–1.12]	1.07 [0.87–1.31]
Master's degree/PhD	0.67 [0.35–1.31]	1.13 [0.31–4.06]	0.79 [0.66–0.95]	0.82 [0.61–1.12]	0.97 [0.82–1.16]	0.98 [0.73–1.31]
Urbanity						
Very strong	1.40 [1.02–1.92]	1.74 [0.90–3.37]	1.02 [0.90–1.16]	0.96 [0.76–1.20]	1.03 [0.91–1.16]	1.00 [0.80–1.25]
Strong	1.00	1.00	1.00	1.00	1.00	1.00
Moderate	0.69 [0.50–0.95]	0.51 [0.23–1.10]	0.92 [0.83–1.02]	0.84 [0.69–1.01]	1.02 [0.92–1.13]	1.00 [0.83–1.20]
Little	0.71 [0.51–0.97]	0.62 [0.29–1.31]	0.92 [0.83–1.03]	0.86 [0.71–1.04]	0.90 [0.81–1.00]	0.80 [0.66–0.97]
None	0.94 [0.66–1.32]	0.72 [0.31–1.69]	0.90 [0.79–1.02]	1.00 [0.79–1.26]	0.91 [0.80–1.02]	1.00 [0.80–1.26]
Longstanding illness					2.51 [2.32–2.72]	2.21 [1.92–2.55]
Poor self-rated health					4.30 [3.36–5.50]	5.61 [2.08–10.2]

A comparison between odds ratios in the entire population participating in the health interview survey (HIS) and in those participating in the health examination survey (HES) alone.

Discussion

This study found that HES participants differed from non-participants in many aspects. Most notable was that participants had a higher socio-economic status and were more likely to be 'worried well'. Analysis of the biases caused by this selective participation found that it does not generally lead to large biases in prevalence figures, and it has little influence on the relationship between variables.

Initially, the variables determining participation in a second-step examination (HES) by those who had participated in the first step (HIS) were investigated. As hardcore non-responders are not present in the HIS sample, the mechanisms of non-participation in a second-step examination may differ from the mechanisms of initial non-participation. However, most of the results indicate that the mechanisms are not very different from those in non-response in the first step. As in most non-response studies looking at initial non-response, people from a higher socio-economic strata were more likely to participate, as were the 'worried well', characterized by using more healthcare services,[10, 11 and 18] being healthier,[19, 20, 21 and 22] and following a healthier lifestyle.[23, 24 and 25] However, the latter aspect (healthier lifestyle) was only partially observed in the present study; participation in sports only made a difference for subjects with a low income, while smoking did not influence participation rates at all.

Another difference from previous findings is the lower participation rate observed in rural areas, where surveys usually observe higher rates of response.[26 and 27] This may be due to the longer distance to travel to the examination centre in rural areas. Participation in the NHANES physical examination was observed to be higher in those living near the location of the examination.[12]

People who did not return the self-administered questionnaire of the HIS participated much less frequently in the HES. A similar finding has been observed previously for those who returned a questionnaire only after receiving a reminder; they were considerably less likely to respond to a questionnaire sent 11 years later.[28] This supports the notion of a general tendency to respond as postulated in the 'continuum of resistance' theory.[29 and 30] According to this theory, the more resistant a subject is to participation, the more he/she resembles people who never respond. An non-testable sequitur of this theory is that the difference between respondents who are 'easy' and 'hard' to recruit can be extrapolated to predict how non-responders would have reacted, and statistical methods have been developed that perform such an extrapolation.[31] In the present case, this theory implies that a lack of bias in outcomes due to the selective response of HES participants within the HIS sample indicates that there is also no bias due to selective non-response of HIS responders within the general population. This theory, however, remains to be verified.

The authors also looked at the magnitude of the bias due to selective non-response in the population average values of the main outcomes. Although non-participation rates were high in this study and depended on socio-economic position and health status (over-representation of the 'worried well'), bias could not be demonstrated for age-standardized and sex-specific average outcomes, except for body height in men and HDL cholesterol, fasting glucose and body weight in women. This is probably due to a lack of concordance in the variables explaining non-response (health status and socio-economic position, see Table 2) and those predicting the outcomes (Table 3). Apart from age and gender, only education was strongly related to both non-response and one outcome (body

height). Based on this lack of two-way associations, it was anticipated that no substantial bias could be detected with the variables available in this study.

This is particularly true for outcomes for which the models predicted less than 30% of the variance in outcomes, indicating that the HIS variables are only weakly related to the outcomes. If other variables (not included in the HIS) exist that are related to both participation and an outcome, this would result in bias not picked up by the modelling. One example would be dietary habits, which influence serum cholesterol levels. It has been observed previously that those using margarine and skimmed milk were more likely to consent to clinical examination.[13] Apart from a lack of dietary information, the HIS covered a large range of health-related topics and socio-economic information, so most other important areas were probably included in the present analyses. Height and weight could be predicted accurately and therefore, their prediction is unlikely to be biased by omitted predictors. For the variables that were predicted accurately, a bias was only observed for body weight in women and body height in men.

The differences observed between HES participants and non-participants are probably related to the under-representation of participants with lower socio-economic status in the HES. An earlier survey in The Netherlands looking at the relationship between socio-economic status and blood pressure, cholesterol and overweight showed that the relative index of inequality was only above five for overweight and HDL-cholesterol in women; the same two factors for which bias was observed in women in the present study. The relative index of inequality for these factors was clearly lower in men, and was even lower for blood pressure and total cholesterol in both men and women.[32]

Another explanation for the lack of bias in outcomes could be that non-responders are a mixture of groups with their own reasons for not participating. As these reasons may be connected to health outcomes in different ways, their influences could partially cancel each other out.

These results are largely similar to those of an earlier study on non-response on a questionnaire following a health examination (instead of a health examination following an interview, as in the present case). This study showed no differences in total serum cholesterol, HDL-cholesterol and systolic blood pressure between respondents and non-respondents, and only a slightly higher diastolic blood pressure (0.7 mmHg) in non-respondents.[33]

Lastly, potential bias in relationships between some variables strongly related to non-participation was studied. In all cases, age- and gender-adjusted associations hardly changed when the HES participants alone were used in the analysis instead of all the HIS respondents. These findings are in agreement with previous studies which showed (with a single exception[34]) no significant bias in associations.[6, 10, 19, 28, 35, 36, 37 and 38]

Although these results are reassuring in that large biases could not be observed and the biases observed can be attributed to socio-economic differences (and thus, can be adjusted for), the possibility of selection dependent on non-observed factors still exists with the large non-participation rate of the HES. Also, this study only addressed the bias by non-participation among HIS respondents, who may already be a selective sample from the general population as the response rate of the HIS was only 56%. Moreover, the sampling framework of the HIS excluded the institutional and homeless population, and those with difficulty speaking Dutch are also under-represented among HIS respondents. If those groups were included in the HIS and were invited to take part in the HES, their non-

participation rates may further increase bias as these groups may have high non-response rates and poorer health.

A preceding HES in municipal health centres recruited participants directly from the general population and had a response rate of 45%.[6] This is only slightly lower than the participation rate of 49% in the present HES, which was observed in subjects that had already participated in the HIS and had consented to being invited. This suggests that having a population of subjects willing to participate does not have a marked effect on participation rates. Together with the finding that factors influencing participation in the second stage of a survey seem to be similar to those influencing the first stage, this suggests that each time subjects are invited to take part in a study, they refuse with a certain probability. This probability depends on characteristics of the subject, e.g. socio-economic status, but also seems to be the result of a spur of the moment, random decision. This is reassuring in that it suggests that second-stage non-participation is partly a random process and that those participating are representative of those with similar characteristics who are not participating. It also implies that when high participation rates are required, multistage surveys are not the optimum method for data collection.

In conclusion, although the subjects in the HES sample are clearly different from the HIS population in terms of sociodemographic characteristics, socio-economic status, healthcare use and health status, no resulting bias could be demonstrated for measures of association. For absolute age-standardized outcome measures, some bias, mostly small, was seen for body height in men and for HDL-cholesterol, fasting glucose and body weight in women. Therefore, caution is needed when using these outcome measures.

References

1 W.J.M.J. Appelboom, Non respons onderzoek gezondheidsenquete 1985, CBS, Voorburg (1984).

2 M. Van Baal, Respons analysis survey van POLS—een tussenrapportage, CBS, Heerlen (2000).

3 S. Te Riele, Vertekening door non-respons: hoe nauwkeurig zijn de uitkomsten van persoonsenquêtes?, CBS, Voorburg/Heerlen (2002).

4 B.F.M. Bakker and J. van Rooijen, One figure for the supply and demand of services, Netherlands official stat 15 (2000), pp. 40–46.

5 H.S.J. Picavet, National health surveys by mail or home interview: effects on response, J Epidemiol Community Health 55 (2001), pp. 408–413.

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

7 G. Turrell, C. Patterson, B. Oldenburg, T. Gould and M.A. Roy, The socio-economic patterning of survey participation and non-response error in a multilevel study of food purchasing behaviour: area- and individual-level characteristics, Public Health Nutr 6 (2003), pp. 181–189.

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

9 M. van den Akker, F. Buntinx, J.F. Metsemakers and J.A. Knottnerus, Morbidity in responders and non-responders in a register-based population survey, Fam Pract 15 (1998), pp. 261–263.

10 S. Reijneveld and K. Stronks, The impact of response bias on estimates of health care utilization in a metropolitan area: the use of administrative data, Int J Epidemiol 28 (1999), pp. 1134–1140.

11 J.F. Etter and T.V. Perneger, Analysis of non-response bias in a mailed health survey, J Clin Epidemiol 50 (1997), pp. 1123–1128.

12 R.N. Forthofer, Investigation of nonresponse bias in NHANES II, Am J Epidemiol 117 (1983), pp. 507–515.

13 E. Pullen, D. Nutbeam and L. Moore, Demographic characteristics and health behaviours of consenters to medical examination. Results from the Welsh heart health survey, J Epidemiol Community Health 46 (1992), pp. 455–459.

14 J.W. Winkels and P.C.J. Everaers, Design of an integrated survey in the Netherlands. The case of POLS, Netherlands official stat 13 (1998), pp. 8–11.

15 H. Smeets, Non-response van eerste generatie allochtonen in het POLS [Non-response of first generation immigrants in the periodical survey of conditions of living]. Heerlen/Voorburg, CBS, Bevolkings Trends (2004) 52(4): 92–97.

16 D.B. Rubin, Multiple imputation for nonresponse in surveys, Wiley, New York (1987).

17 Lazzeroni LG, Schenker N, Taylor JMG. Robustness of multiple imputation techniques to model misspecification. American Statistical Association's 1990 Proceedings of the Survey Research Methods Section 1990; 260-265.

18 L.M. Lamers, Medical consumption of respondents and non-respondent to a mailed health survey, Eur J Public Health 7 (1997), pp. 267–271.

19 L.K. Heilbrun, A. Nomura and G.N. Stemmermann, The effects of non-response in a prospective study of cancer: 15-year follow-up, Int J Epidemiol 20 (1991), pp. 328–338. Abstract-EMBASE | Abstract-MEDLINE | Order Document | Abstract + References in Scopus | Cited By in Scopus

20 E. Ronmark, A. Lundqvist, B. Lundback and L. Nystrom, Non-responders to a postal questionnaire on respiratory symptoms and diseases, Eur J Epidemiol 15 (1999), pp. 293–299.

21 A.H. Mihelic and E.M. Crimmins, Loss to follow-up in a sample of Americans 70 years of age and older: the LSOA 1984–1990, J Gerontol B Psychol Sci Soc Sci 52 (1997), pp. S37–S48.

22 E. Shahar, A.R. Folsom and R. Jackson, The effect of nonresponse on prevalence estimates for a referent population: insights from a population-based cohort study. Atherosclerosis risk in communities (ARIC) study investigators, Ann Epidemiol 6 (1996), pp. 498–506.

23 R.C. Klesges, J.E. Williamson, G.W. Somes, G.W. Talcott, H.A. Lando and C.K. Haddock, A population comparison of participants and nonparticipants in a health survey, Am J Public Health 89 (1999), pp. 1228–1231.

24 A. Hill, J. Roberts, P. Ewings and D. Gunnell, Non-response bias in a lifestyle survey, J Public Health Med 19 (1997), pp. 203–207.

25 C.A. Macera, K.L. Jackson, D.R. Davis, J.J. Kronenfeld and S.N. Blair, Patterns of non-response to a mail survey, J Clin Epidemiol 43 (1990), pp. 1427–1430.

26 S. Eaker, R. Bergstrom, A. Bergstrom, H.O. Adami and O. Nyren, Response rate to mailed epidemiologic questionnaires: a population-based randomized trial of variations in design and mailing routines, Am J Epidemiol 147 (1998), pp. 74–82.

27 G.M. Jay, J. Liang, X. Liu and H. Sugisawa, Patterns of nonresponse in a national survey of elderly Japanese, J Gerontol 48 (1993), pp. S143–S152.

28 T.M. Eagan, G.E. Eide, A. Gulsvik and P.S. Bakke, Nonresponse in a community cohort study: predictors and consequences for exposure-disease associations, J Clin Epidemiol 55 (2002), pp. 775–781.

29 K. Kypri, S. Stephenson and J. Langley, Assessment of nonresponse bias in an internet survey of alcohol use, Alcohol clin Exp Res 28 (2004), pp. 630–634.

30 N. Bates and K. Creighton, The last five percent: what can we learn from difficult/late interviews?, Proceedings of the Annual Meetings of the American Statistical Association, US Census Bureau, Washington DC (2000).

31 J.R. Hochstim, A critical comparison of three strategies of collecting data from households, J Am Stat Assoc 62 (1967), pp. 976–989.

32 Fvd. Lucht and H.S.J. Picavet, Sociaal-economische verschillen in persoonskenmerken, [Socio-economic differences in personal characterics], RIVM, Bilthoven (2003).

33 B.K. Jacobsen and D.S. Thelle, The Tromsø heart study: responders and non-responders to a health questionnaire, do they differ?, Scand J Soc Med 16 (1988), pp. 101–104.

34 L.J. Launer, A.W. Wind and D.J. Deeg, Nonresponse pattern and bias in a community-based cross-sectional study of cognitive functioning among the elderly, Am J Epidemiol 139 (1994), pp. 803–812.

35 J. Brogger, P. Bakke, G.E. Eide and A. Gulsvik, Contribution of follow-up of nonresponders to prevalence and risk estimates: a Norwegian respiratory health survey, Am J Epidemiol 157 (2003), pp. 558–566.

36 J. Siemiatycki and S. Campbell, Nonresponse bias and early versus all responders in mail and telephone surveys, Am J Epidemiol 120 (1984), pp. 291–301.

37 P. Bakke, A. Gulsvik, P. Lilleng, O. Overa, R. Hanoa and G.E. Eide, Postal survey on airborne occupational exposure and respiratory disorders in Norway: causes and consequences of non-response, J Epidemiol Community Health 44 (1990), pp. 316–320.

38 M.A. Austin, M.H. Criqui, E. Barrett-Connor and M.J. Holdbrook, The effect of response bias on the odds ratio, Am J Epidemiol 114 (1981), pp. 137–143.