

Research

Open Access

The Oslo Health Study: The impact of self-selection in a large, population-based survey

Anne Johanne Sjøgaard*¹, Randi Selmer¹, Espen Bjertness² and Dag Thelle³

Address: ¹Norwegian Institute of Public Health, Oslo, Norway, ²Institute of General Practice and Community Medicine, Faculty of Medicine, University of Oslo, Norway and ³Akershus University Hospital, Faculty of Medicine, University of Oslo, Norway

Email: Anne Johanne Sjøgaard* - ajso@fhi.no; Randi Selmer - randi.selmer@fhi.no; Espen Bjertness - Espen.Bjertness@samfunnsmed.uio.no; Dag Thelle - Dag.Thelle@hjl.gu.se

* Corresponding author

Published: 06 May 2004

Received: 26 September 2003

International Journal for Equity in Health 2004, **3**:3

Accepted: 06 May 2004

This article is available from: <http://www.equityhealthj.com/content/3/1/3>

© 2004 Sjøgaard et al; licensee BioMed Central Ltd. This is an Open Access article: verbatim copying and redistribution of this article are permitted in all media for any purpose, provided this notice is preserved along with the article's original URL.

Abstract

Background: Research on health equity which mainly utilises population-based surveys, may be hampered by serious selection bias due to a considerable number of invitees declining to participate. Sufficient information from all the non-responders is rarely available to quantify this bias. Predictors of attendance, magnitude and direction of non-response bias in prevalence estimates and association measures, are investigated based on information from all 40 888 invitees to the Oslo Health Study.

Methods: The analyses were based on linkage between public registers in Statistics Norway and the Oslo Health Study, a population-based survey conducted in 2000/2001 inviting all citizens aged 30, 40, 45, 59–60 and 75–76 years. Attendance was 46%. Weighted analyses, logistic regression and sensitivity analyses are performed to evaluate possible selection bias.

Results: The response rate was positively associated with age, educational attendance, total income, female gender, married, born in a Western county, living in the outer city residential regions and not receiving disability benefit. However, self-rated health, smoking, BMI and mental health (HCSL) in the attendees differed only slightly from estimated prevalence values in the target population when weighted by the inverse of the probability of attendance.

Observed values differed only moderately provided that the non-attending individuals differed from those attending by no more than 50%. Even though persons receiving disability benefit had lower attendance, the associations between disability and education, residential region and marital status were found to be unbiased. The association between country of birth and disability benefit was somewhat more evident among attendees.

Conclusions: Self-selection according to sociodemographic variables had little impact on prevalence estimates. As indicated by disability benefit, unhealthy persons attended to a lesser degree than healthy individuals, but social inequality in health by different sociodemographic variables seemed unbiased. If anything we would expect an overestimation of the odds ratio of chronic disease among persons born in non-western countries.

Background

Since the Black report was published in 1982 [1] a considerable number of papers have described social inequality in health and discussed possible reasons for the invert association between social position and mortality [1-4]. Even in the egalitarian Scandinavian countries there has been reported a substantial difference in health between the least and the most privileged groups [5]. Ecological analysis in Oslo also showed large differences in mortality rates between residential areas characterised by social inequalities [6]. In order to identify variables at individual level to explain these health differences, the Oslo Health Study, an age stratified population-based health survey, was carried out in 2000–2001.

Valid estimates of health inequality, however, and especially prevalence figures depend heavily upon representative attendance. The low response rates observed of certain exposed and affected groups such as lower social classes, elderly, single-households, third-world country immigrants, receivers of social security benefits, urban city dwellers, people with poor self-reported health and unhealthy lifestyle, have been a major concern in many population-based health surveys [7-18]. The Oslo Health Study (HUBRO), (HUBRO = eagle owl, acronym for the Norwegian title of the Oslo Health Study), provided the possibility to compare the prevalence and associations between already known background variables in the total study population with that of the attendees. We are not aware of any large population-based study using a complete set of data on both exposure and outcome variables concerning all invited subjects, analysed to assess whether the associations observed in health surveys are influenced by selection bias.

Some studies have compared own data with large-scale survey data for the total population [17,19,20], others have analysed fairly small samples [9,21,22] or compared a limited number of linked variables [8,9]. Three Swedish population-based studies and one Dutch study have used administrative data linkage to analyse non-response bias. The first three did not, however, discuss possible bias in associations [10,16,23], which is of major interest, and the last paper studied only health care utilisation [15].

In the present analysis public register data from Statistics Norway (education, income, social- and disability benefits) covering all Norwegian citizens, were linked with data from the Oslo Health Study, using the individual's unique 11-digit personal identification number.

Thus the objectives of the present analysis were to:

1. Identify subgroups with low response rates and predictors of response

2. Investigate the magnitude and direction of errors in prevalence estimates of selected exposure and outcome variables in the health survey

3. Investigate the magnitude and direction of selection bias of association measures (odds ratio) between selected sociodemographic variables and disability benefit.

Methods

The Oslo Health Study, a joint collaboration between the Oslo City Council, the University of Oslo and the Norwegian Institute of Public Health, was conducted in Oslo from May 2000 to September 2001. An invitation for participation in the health survey was sent to all residents born in the following years: 1924, 1925, 1940, 1941, 1955, 1960 and 1970. Additional groups, not included in the present study, were also invited to HUBRO [24].

Invitation and procedure

The Oslo Health Study consists of a central core and 70 supplementary projects. The data collection for the core, which we report in the present paper, was undertaken following the procedure mentioned below:

A letter of invitation – containing an information brochure and the main questionnaire – was mailed two weeks prior to the appointment at the screening station. Information included, among other things, that participants could avail themselves of the brochure and the questionnaire, in 11 other languages.

At the screening station a simple clinical examination was conducted: A venous non-fasting blood sample was analysed for serum total cholesterol, HDL cholesterol, glucose and triglycerides. Automatic device (DINAMAP) measured pulse recordings, systolic- and diastolic blood pressures. Body weight (in kilograms), height (in cm) and waist-hip-ratio (cm) were measured with a standard procedure according to the protocol [25].

At the screening station the main questionnaire was handed in and the participants were given two supplementary questionnaires, which they were requested to fill in at home and return by mail in pre-addressed pre-stamped envelopes (more details are presented at HUBRO's web site [24]).

Four weeks after attending the examination, a letter with the results of the examination and blood tests was sent to all participants. Those presenting with high risk of cardiovascular disease [25] were offered a new clinical examination at Ullevål University Hospital.

An extensive information campaign was developed in order to motivate those invited to attend [24].

Table 1: Number and participation rate according to age and gender in The Oslo Health Study in 2000–2001

Age (yrs)	Number of participants *			Participation rate (%)		
	Men	Women	All	Men	Women	All
30	1826	2288	4114	32.2	39.9	36.1
40	1547	1905	3452	37.6	50.1	43.7
45	1389	1757	3146	39.9	53.5	46.5
59–60	2117	2357	4474	53.4	57.3	55.4
75–76	1525	2059	3584	58.2	50.0	53.2
All	8404	10366	18770	42.4	49.3	45.9

* Number attending the survey and/or submitting at least one of the questionnaires.

Participants and attendance

Of the 40 888 citizens invited, a total of 18 770 individuals (46 %) participated in the survey.

Almost all those attending the screening station responded to the main questionnaire, while the response rates to the first and second supplementary questionnaire were 82 %-87 % of those attending. Table 1 shows the participation rates in different subgroups of age and gender. The notation "attendees" and "non-attendees" used later refer to participants and non-participants. Further detailed information concerning participation in the Oslo Health Study can be obtained from HUBRO's web site [24].

Ethics and approvals

All the participants of the Oslo Health Study have given their written consent. The participant's names and personal ID numbers are omitted when data are used for research purposes. The Norwegian Data Inspectorate has approved the Oslo Health Study, the Regional Committee for Medical Research Ethics has evaluated it, and it has been conducted in full accordance with the World Medical Association Declaration of Helsinki.

Linkage to Statistics Norway

Sociodemographic information from public registers in Statistics Norway [26] was linked to data from the clinical examination, the main- and first supplementary questionnaire and to the population file used for invitation to the Oslo Health Study – through the individual's personal identification number. All personal identification has been erased before the data is analysed.

Variables used in the non-response analyses

The following variables were added from the Statistics Norway's event-history database: highest education completed, personal income and information about disability -, rehabilitation -, sickness -, unemployment and single parent benefit. From the invitation file, also obtained

from Statistics Norway, we used the following variables in the present paper: age, gender, marital status, country of birth and residential region.

From the clinical examination and the main questionnaire we have used:

Body mass index: Body weight (kilogram)/(height (meter))²

Self-evaluated general health status: How would you describe your present state of health? (poor, not very good, good, very good)

Self-reported daily smoking: Have you smoked or do you smoke daily? (yes – now, yes – earlier, never)

Mental distress: Below is a list of various problems. Have you suffered from any of the following during the last week (including today)? (Put a cross for every problem)

The 10 items asked [27] are an abridged version of the Hopkins Symptom Check List (HSCL) [28] which is a widely used, self-administered instrument designed to measure psychological distress in population surveys. The HSCL-10 consists of 10 items on a 4-point scale ranging from "not at all" to "extremely" [27]. The average score is calculated by dividing the total score by number of items – i.e. ten. Missing values are replaced with the sample mean values for each item. Records with three or more missing items are, however, excluded.

Statistical analyses

The attendance rates according to previously listed background variables were calculated using data from public registers. The crude and adjusted odds ratios for attendance were estimated in logistic regression models including all the sociodemographic variables as covariates.

To demonstrate the effect of selective attendance on prevalence estimates of selected variables (good/excellent self-reported health, proportion of daily smokers, proportion with obesity (BMI ≥ 30) and proportion with mental distress score 1.85 or above), we used the inverse of the probability of attendance based on logistic regression models as weights. In the first step we weighted with age, sex and education. These variables are known to be strong predictors of health and health behaviour and are commonly used in analyses of social inequality in health. In the second step we included all background variables significantly associated with attendance to see if more adjustments would give a better estimate of the true prevalence. This procedure assumes that the prevalence is similar among the attendees and non-attendees provided they belong to the same sociodemographic categories.

Sensitivity analyses are useful to assess the consequences of dissimilarities (e.g. prevalence) between attendees and non-attendees within each stratum of background variables. We calculated estimates of the "true" prevalence assuming that the prevalence was 10 %, 25 %, 50 % and 100 % higher in non-attendees compared to attendees. Variables with low, medium and high prevalence (diabetes, obesity, symptoms of mental distress, daily smoking and good/excellent health) were included in this analysis.

The possible effects of selected attendance on associations were also assessed based on data from the public registers. Odds ratios for the four exposure variables; education, residential region, country of birth and marital status – and the outcome variable disability benefit, were estimated separately for attendees, non-attendees and for the total invited population. Testing for interactions was done by logistic regression.

Results

Attendance by different subgroups

Compared to the invited population, the following subgroups were under-represented among the attendees: males, young age-group, unmarried and separated/divorced, those not born in Norway, inner city dwellers, those with unknown or lower secondary education, low income groups and receivers of disability benefit (Table 2). The other social security benefits, i.e. single parent, rehabilitation and sickness, were not significantly associated with attendance, and are not included in further analyses. The odds ratios of attendance according to country of birth turned non-significant in a model with all covariates included. There were, on the other hand, significant interactions between age, education and country of birth: In the two youngest age groups with lower or upper secondary education the attendance was higher among people born in non-western countries compared to people born in Norway. In the other sub groups of age and education, the attendance was higher among people born in Norway (not shown).

Prevalence estimates

The prevalence estimates of the four selected variables, crude and weighted, are shown in Table 3. The prevalence of good/excellent self-rated health, daily smokers, obesity and symptoms of mental distress, changed only marginally when weighted according to the response model based on age, sex and education (model 1). Similar results were achieved in a response model which also included marital status, residential region, total income and disability benefit (model 2). Response models including interactions between country of birth, education and age did not change these results (data not shown). When we included the oldest age group and adjusted according to

the response model 2 without disability benefit, we received similar results (data not shown).

The weighting in table 3 is based on the assumption that attendees and non-attendees are similar within each stratum of sociodemographic variables. Sensitivity analyses are useful if one doubts this similarity. Thus we made different assumptions about prevalence among non-attendees compared to attendees. In Table 4 different examples are shown for variables with low, medium and high prevalence. If the percentage reporting diabetes among non-attending men 30–76 years old is 25 % higher than among the attending men, the true prevalence is 4.3 % compared to 3.8 % in those attending. A 50 % higher diabetes prevalence among the non-attendees implies that the true prevalence will be 4.9 %. Similar calculations for the other four variables showed only moderate changes of the prevalence figures in the target population compared to the attendees if the non-attendees do not differ from the attendees by more than 50%. With more substantial differences between non-attendees and attendees, in the region of a ratio of 2, the true prevalence in the population will be around 50% higher than the prevalence estimated from the attendees only.

Associations

A higher percentage of non-attendees received disability benefit compared to people attending – independent of level of education (Table 5). Equal odds ratios appeared among attendees, non-attendees and the total invited population, when analysing the outcome measure disability benefit and the exposure variable level of education. This indicates no selection bias. Performing the same analysis for residential regions and marital status as exposure variables gave similar results – the associations were similar among the invited and the attending population (data not shown).

The pattern changed, however, when country of birth was used as an exposure variable (Table 6). The percentage of individuals born in Norway receiving disability benefit was higher among non-attendees than among attendees, with a maximum ratio of 2 in some sub-groups, but for those born in non-western countries the difference between attendees and non-attendees was much smaller. The ethnic differences were more pronounced in men. The odds ratios of disability benefit comparing non-western and Norwegian born attendees, were higher than the odds ratios among all invited from these countries (interaction between attendance and country of birth in logistic regression, $p = 0.003$) (Table 6). Thus the difference between individuals born in Norway and in non-western countries was overestimated in analyses of the attendees only, indicating possible selection bias with a 57% over-

Table 2: The association (odds ratio) between selected sociodemographic variables and attendance (yes/no) in the Oslo Health Study 2000–2001. Crude odds ratio and adjusted for all variables by logistic regression.

	Invited N	Attendance (%)	Crude odds ratio **	Adjusted odds ratio **	(95% CI)
Sex					
Men	19839	42.4	Ref.	Ref.	
Women	21035	49.3	1.32	1.51	(1.44–1.59)
	40874†	45.9			
Age					
30	11404	36.1	Ref.	Ref.	
40+45	14668	45.0	1.45	1.38	(1.30–1.47)
59–60	8072	55.4	2.20	2.37	(2.19–2.55)
75–76	6730	53.2	2.02		
	40874	45.9			
Marital status					
Unmarried	14067	39.1	Ref.	Ref.	
Married	18083	51.6	1.66	1.31	(1.23–1.39)
Widowed	2545	48.7	1.48	1.08	(0.90–1.30)
Separated/divorced	5900	43.2	1.19	0.89	(0.82–0.96)
Reg. partnership	111	46.8	1.38	1.38	(0.92–2.06)
	40706	45.8			
Country of birth					
Norway	33519	47.2	Ref.	Ref.	
Western countries	2422	42.2	0.82	1.06	(0.96–1.17)
Non-western	4933	39.3	0.72	0.99	(0.92–1.07)
	40874	45.9			
Residential region					
Outer east	16696	50.3	Ref.	Ref.	
Outer west	10167	49.9	0.98	0.75	(0.71 – 0.80)
Inner west	5939	40.4	0.67	0.62	(0.58 – 0.67)
Inner east	6215	40.2	0.66	0.79	(0.74 – 0.85)
	39017	47.1			
Education					
Lower secondary	6766	39.5	Ref.	Ref.	
Upper secondary	16808	46.5	1.33	1.42	(1.32 – 1.53)
College/University	14906	48.7	1.46	1.86	(1.71 – 2.01)
Unknown	1504	24.2	0.49	0.58	(0.50 – 0.68)
	39984	45.3			
Total income (NOK)					
< 100 000	5014	34.3	Ref.	Ref.	
- 199 000	9490	42.9	1.44	1.46	(1.33 – 1.60)
- 399 000	19473	49.2	1.85	1.82	(1.68 – 1.98)
400 000+	6178	44.3	1.52	1.44	(1.30 – 1.59)
	40155	45.1			
Disability benefit*					
No	30918	44.0	Ref.	Ref.	
Yes	2689	38.7	0.80	0.66	(0.60 – 0.73)
	33607	43.6			
Single parent benefit*					
No	33138	43.7	Ref.	Ref.	
Yes	469	35.8	0.72	0.89	(0.73 – 1.09)
	33607	43.6			
Rehabilitation benefit*					
No	33156	43.6	Ref.	Ref.	
Yes	451	42.4	0.95	0.93	(0.76 – 1.14)
	33607	43.6			
Sickness benefit*					
No	32166	43.5	Ref.	Ref.	
Yes	1441	44.9	1.06	0.89	(0.79 – 1.00)
	33607	43.6			

Table 2: The association (odds ratio) between selected sociodemographic variables and attendance (yes/no) in the Oslo Health Study 2000–2001. Crude odds ratio and adjusted for all variables by logistic regression. (Continued)

Unemployment benefit*					
No	32716	43.7	Ref.	Ref.	
Yes	891	38.5	0.81	0.95	(0.82 – 1.10)
	33607	43.6			

† Numbers differ a little from text due to updating of data register with respect to number of deaths before screening * The numbers are based on invited individuals below 75 years, to be able to include social security benefits ** The crude odds ratios are based on the total sample, while the adjusted odds ratios are estimated based on the total sample excluding those 75–76 years of age to be able to include social security benefits in the model.

estimation of OR in men aged 40 and 45 years and 33% in men aged 59–60 years.

The odds ratios were slightly changed when adjusted for marital status, residential region and education, but the pattern remained the same (data not shown).

For variables not available for non-attendees we are left with assessing possible bias in associations in sensitivity analyses. What would be the possibility that the increased odds found for certain conditions is explained totally by self-selection bias? As an example we have chosen women aged 59–60 years and studied the association between daily smoking and mental health in those 51% of the invited who answered these questions. Observed odds ratio of mental distress (HCSL ≥ 1.85) was 1.6. In sensitivity analyses we assume that the percentage with mental distress and percent daily smoking are somewhat higher in the total population (e.g. 20.0 % with mental distress, 30.0% daily smokers) and odds ratio = 1. Based on these assumptions the calculated odds ratio for non-attendees will be 0.66 (mental distress: 25.0%, daily smoking: 35.0%). Thus, if the positive association in total were to be explained by selection bias, the non-smokers should have more mental distress than smokers should among the non-attendees. This seems as a very unlikely situation and we conclude that our finding of a positive association between smoking and mental distress is valid.

Discussion

The non-attendees in our study were characterised by being young, unmarried, males, inner city dwellers, and belong to the lower income and educational echelons. They also received, to a larger extent, disability benefit than did the attendees. This corresponds to results from other studies where the same sub-groups tend to be under-represented [8,10-18,22,23,29]. Also in accordance with some studies [7,10,12,13,15,16,22] we found those persons born in non-western countries to be under-represented.

The participation rate in HUBRO was 42.4% in men and 49.3% in women, resulting in a total number of respond-

ers of 18 770. The attendance rate was lower than in previously comparable surveys in Norway [30], but only slightly lower than reported in the population-based Oslo Study of men in 1972–73 where 42.8% attended among 20–39 years old and 62.6% in those 40–49 years old [31]. Since the 1970's the Oslo population has become more heterogeneous with regard to ethnic origin and socioeconomic situation. During the last years the response rates have declined in Norway as well as in other countries [18,32-36].

It is not evident that a higher response rate in HUBRO would have prevented selection bias. Several other studies have demonstrated only moderate changes in prevalence estimates and sociodemographic distribution when comparing results by increasing the response rates in the range from around 30% to 70% [18,37-41].

The register linkage to data in Statistics Norway made it possible to estimate prevalence values of selected health variables based on information regarding certain socioeconomic and demographic variables for all invited. The observed values of self-rated health, smoking, BMI and mental health (HCSL) in the attendees differed only slightly from the estimated prevalence values in the target population when weighted by the inverse of probability of attendance. Thus, self-selection by sociodemographic variables did not influence prevalence estimates. This is reassuring and in accordance with other studies weighted in similar ways [18]. The analyses were, however, based on the assumption of similar prevalence among non-attendees and attendees within each sociodemographic group.

Assuming that the prevalence in non-attendees differs by no more than 50% from the prevalence in the attendees, the calculated prevalence for the target population do not differ much from the estimated prevalence values. Most studies report differences of less than 25–30%. In the Rancho Bernardo study [42] diabetes was about 30% higher in non-responders (5.5%) compared to responders (4.2%) whereas the difference in smoking was below 20% (29.5 % in non-responders and 24.7% in responders).

Table 3: Percentage with good or excellent health, daily smoking, body mass index 30 kg/m² or higher and percentage scoring 1.85 or higher on the HSCL scale in The Oslo Health Study 2000–2001. Number answering the question (N), crude- and weighted percentages according to the inverse of the probabilities of participation from logistic regression models. All men and women aged 30, 40, 45, 59 and 60 years of age.

	N	Crude percent	(95% CI)	Model 1*	Model 2**
Weighted percent					
Reporting good or excellent health					
Men					
30	1813	90.3	(89.0–91.7)	89.6	90.3
40+45	2899	80.0	(78.6–81.5)	78.8	78.3
59–60	2075	72.5	(70.6–74.5)	71.5	69.9
	6787	80.5	(79.6–81.4)	80.5	80.2
Women					
30	2257	85.7	(84.3–87.2)	84.9	84.8
40+45	3607	77.8	(76.5–79.2)	76.9	76.4
59–60	2307	64.5	(62.5–66.4)	63.4	61.9
	8171	76.2	(75.3–77.1)	76.3	75.6
Daily smokers					
Men					
30	1813	22.9	(21.0–24.8)	23.8	23.9
40+45	2903	32.0	(30.3–33.7)	33.6	34.5
59–60	2094	25.9	(24.0–27.8)	26.7	28.0
	6810	27.7	(26.6–28.8)	28.7	29.4
Women					
30	2250	22.5	(20.8–24.3)	22.6	23.1
40+45	3610	32.9	(31.4–34.4)	33.3	33.4
59–60	2312	25.0	(23.2–26.7)	25.5	26.4
	8172	27.8	(26.8–28.8)	27.9	28.4
Body mass index 30 kg/m² or higher					
Men					
30	1801	11.1	(9.6–12.6)	11.3	10.9
40+45	2848	14.3	(13.0–15.6)	14.6	13.9
59–60	2039	18.6	(16.9–20.3)	18.8	18.5
	6688	14.8	(13.9–15.7)	14.5	14.1
Women					
30	2204	9.1	(7.9–10.3)	9.7	9.7
40+45	3556	13.5	(12.4–14.6)	14.3	14.5
59–60	2291	17.9	(16.3–19.5)	18.5	19.2
	8051	13.5	(12.8–14.2)	13.8	14.1
HSCL 1.85 or higher					
Men					
30	1755	5.9	(4.8–7.0)	6.4	6.4
40+45	2736	10.3	(9.2–11.4)	10.7	11.7
59–60	1952	7.6	(6.4–8.8)	7.8	8.7
	6443	8.3	(7.6–9.0)	8.6	9.2
Women					
30	2166	10.6	(9.3–11.9)	10.6	11.2
40+45	3452	13.4	(12.3–14.5)	13.8	14.5
59–60	2120	15.5	(14.0–17.0)	16.3	17.4
	7738	13.2	(12.4–14.0)	13.3	14.1

*Model 1: Age, sex and education **Model 2: Age, sex, education, marital status, residential region, total income and disability benefit

The utilisation of various types of health care was 3–32% lower in a Dutch study [15]. The corresponding difference in non-fatal stroke and myocardial infarction between non-responders and responders in a Swedish follow-up study with linkage to public registers, were less than 15% [16].

For specific diseases or conditions the difference between attendees and non-attendees could, however, be higher than 50%, which means that the true prevalence in the population will diverge from the prevalence estimated from the attendees only. In our study the ratio between non-attendees and attendees with respect to disability benefits reached a maximum of 2 in some sub-groups of

Table 4: Sensitivity analyses: Estimates of prevalence in target population under different assumptions about the ratio between prevalence in attendees and non-attendees. Rates of attendance: men 42.4 % and women 49.3 %. The Oslo Health Study 2000–2001.

Ratio between prevalence in non-attendees and attendees	Diabetes	Percent daily smokers	Percent BMI \geq 30 kg/m ²	Percent HSCL score \geq 1.85	Ratio between prevalence in non-attendees and attendees	Percent reporting good or excellent health
Men					Men	
1.00	3.8	27.7	14.8	8.3	1.00	80.5
1.10	4.0	29.3	15.7	8.8	0.90	81.4
1.25	4.3	31.7	16.9	9.5	0.75	74.5
1.50	4.9	35.7	19.1	10.7	0.67	70.8
2.00	6.0	43.7	23.3	13.1	0.50	62.9
Women					Women	
1.00	2.6	27.8	13.5	13.2	1.00	76.2
1.10	2.7	29.2	14.2	13.9	0.90	72.3
1.25	2.9	31.3	15.2	14.9	0.75	66.5
1.50	3.3	34.8	16.9	16.5	0.67	63.5
2.00	3.9	41.9	20.3	19.9	0.50	56.9

Table 5: Number (n) and percentage receiving disability benefit, association (odds ratio) with level of education among attendees and all invited. The Oslo Health Study 2000–2001.

Education *	Attendees				All invited			
	n	%	Odds ratio	(95% CI)	n	%	Odds ratio	(95% CI)
Men								
Age 40+45								
Lower secondary	30	9.3	Ref.		143	14.4	Ref.	
Upper secondary	41	3.3	0.34	(0.21 – 0.55)	159	5.0	0.31	(0.25 – 0.40)
College/University	15	1.2	0.12	(0.06 – 0.23)	58	2.0	0.12	(0.09 – 0.17)
Age 59–60								
Lower secondary	116	31.3	Ref.		318	38.9	Ref.	
Upper secondary	142	15.5	0.40	(0.30 – 0.53)	338	20.0	0.39	(0.33 – 0.47)
College/University	36	4.8	0.11	(0.08 – 0.17)	83	6.5	0.11	(0.08 – 0.14)
Women								
Age 40+45								
Lower secondary	38	10.1	Ref.		119	13.8	Ref.	
Upper secondary	77	5.0	0.47	(0.31 – 0.70)	199	6.7	0.45	(0.36 – 0.58)
College/University	37	2.4	0.22	(0.14 – 0.36)	85	3.0	0.20	(0.15 – 0.26)
Age 59–60								
Lower secondary	170	35.0	Ref.		396	39.7	Ref.	
Upper secondary	224	19.9	0.46	(0.37 – 0.59)	449	23.7	0.47	(0.40 – 0.55)
College/University	65	10.6	0.22	(0.16 – 0.30)	128	13.2	0.23	(0.18 – 0.28)

* Testing of interaction between attendance and education in a logistic regression model including all men and women aged 40, 45, 59 and 60 years, and age and sex as covariates in addition to attendance and education, was not significant ($p > 0.05$).

men born in Norway. Previous analyses have shown that the prevalence in early responders differed insignificantly from the prevalence in late responders in age groups less than 70 years [43]. In the oldest age group, however, the prevalence of diabetes, daily smoking and symptoms of mental distress was significantly higher in late responders, with the highest prevalence ratio (2.2) for mental distress in men. This suggests that a prevalence ratio of 2 could

occur in the elderly leading to an underestimation of the prevalence in the population when based on the attendees. For the younger age groups we conclude that, except for special conditions obstructing people from attending health screenings physically or mentally, the prevalence estimates in the present survey, are valid for public health and administrative purposes.

Table 6: Number (n) and percentage receiving disability benefit, association (odds ratio) with country of birth among attendees and all invited. The Oslo Health Study 2000–2001.

Country of birth *	Attendees				All invited			
	n	%	OR	(95% CI)	n	%	OR	(95% CI)
Men								
Age 40+45								
Norway	58	2.7	Ref.		278	4.9	Ref.	
Western	2	1.4	0.52	(0.13–2.12)	6	1.3	0.26	(0.12–0.60)
Non-western	32	6.0	2.33	(1.50–3.62)	101	7.1	1.48	(1.17–1.88)
Age 59–60								
Norway	230	12.8	Ref.		601	18.0	Ref.	
Western	12	10.9	0.83	(0.45–1.54)	25	11.8	0.61	(0.40–0.94)
Non-western	57	36.8	3.96	(2.77–5.64)	141	39.5	2.97	(2.36–3.74)
Women								
Age 40+45								
Norway	123	4.4	Ref.		340	6.2	Ref.	
Western	6	2.9	0.65	(0.28–1.49)	14	3.4	0.53	(0.31–0.91)
Non-western	27	5.6	1.31	(0.85–2.01)	73	7.2	1.18	(0.91–1.53)
Age 59–60								
Norway	431	21.3	Ref.		901	25.7	Ref.	
Western	12	9.2	0.38	(0.21–0.69)	30	13.5	0.45	(0.31–0.67)
Non-western	25	22.7	1.09	(0.69–1.72)	70	23.9	0.91	(0.69–1.20)

* Adjusting for marital status, residential region and education gave similar results. Testing of interaction between attendance and country of birth in a logistic regression model including all men and women aged 40, 45, 59 and 60 years, and age and sex as covariates in addition to attendance and country of birth, gave $p = 0.003$.

The associations between education, residential region and marital status – and disability benefit, measured by odds ratios, were similar among the invited population and the attendees. This demonstrates the robustness of the effect measurements. The only divergent finding was an overestimation of the odds ratios for disability benefit in non-western born compared to Norwegian born, when calculated from attendees only. Kleinbaum et al. [44] have described how selection can be characterised conceptually by comparison of the fourfold table describing the actual sample (i.e. education low/high \times disability yes/no of those who attended the screening) with the corresponding table describing the total invited population. There is no bias in odds ratio if the cross product of the selection probabilities (attendance rates) for each cell is one. Based on the numbers presented in tables 5 and 6 we have calculated these cross products. They were close to one for education and disability, whereas for country of birth and disability, the cross products were generally larger than one. This was due to a collectively over-representation of persons born in Norway without disability benefit and persons born in non-western countries with disability benefit. The overestimation of the odds ratio was, however, moderate (33% in men aged 59–60 years and 57% in men aged 40+45 years). In general, we may assume little bias in odds ratio estimates of variables of the same nature as disability benefit when the observation is limited to the attendees.

Several other studies [19,45–47], including the few with information on all or almost all individuals [15,48,49], show a rather small impact of non-response on risk estimates of health or disease with regard to various background characteristics. In the Rancho Bernardo study the mean odds ratio error was about 15% and most error terms were reasonably close to 1 [48]. Of the 54 OR of health care utilisation that were estimated for various background characteristics in a Dutch study, only 11 differed by, at the most, 10% between the respondents only and the entire sample [15]. In a previous Norwegian community cohort study [49] no overt differences were observed in associations between gender, age and smoking – and respiratory disorders when the analyses were based on initial (65% attendance) compared to "all" respondents (89% attendance).

The main emphasis of the Oslo Health Study was to provide survey data to identify and monitor social, ethnic and geographical differences in health and associated risk factors for disease, to assess the need for health services, and to initiate etiological research. The demand for representative study groups is valid when the inferential goal with the data is a description of the target population. However, a highly representative sample of participants is no longer considered essential for generalisability in etiological studies that report risk estimates rather than prevalence estimates [15,19,46]. Generalisability depends on

the ability to abstract universal scientific hypotheses or theories from a set of observations and not only from the statistical framework of these observations [50]. In biological science, in which we include etiologic medical research, we select subjects with certain characteristics enabling us to make valid comparisons, but not necessarily being representative for the population from which they have been recruited. Rothman and Greenland [50] argue against the notion that generalisation from a study group depends on the study group being a representative sub-group of the target population, in the sense of survey sampling.

There are a number of striking examples of "unrepresentative" studies in the epidemiological literature, studies which have contributed significantly to the domain of medical knowledge [51-55]. All these studies are characterised by being highly unrepresentative of the general population, but the etiological issues being studied are likely to represent valid contributions to the current scientific domains.

It is, of course, important to obtain as large sample size and response rate as possible to avoid selection bias. However, because the presence of selection bias is determined from the selection probabilities within the fourfold table, selection bias may occur even with a relatively large overall response rate, according to Kleinbaum et al. [44]. Conversely, there may be no selection bias despite small response rates.

Conclusions

The attendance in the Oslo Health Study varied according to sociodemographic variables, which is in agreement with other population-based studies in the Western world. It is therefore likely that the results from the present study have a general validity corresponding to similar studies, with the same social distribution of the attendees and the same attendance rate. Furthermore, the weighted prevalence estimates, based on background variables from the total population, were close to unweighted estimates. This indicates that self-selection by sociodemographic background does not influence prevalence estimates to any great degree, assuming the same prevalence between attendees and non-attendees within each stratum of background variables. Assuming dissimilarities, sensitivity analyses have shown a high robustness of the prevalence estimates for prevalence ratios of < 1.5 in non-attendees versus attendees. Unhealthy and sick persons may participate to a less extent than healthy persons as illustrated by those receiving disability benefit. But this selection was independent of education, marital status and residential region leaving the associations between disability benefit and these sociodemographic variables unbiased. If any bias were present, we would

expect an overestimation of the association between ethnicity (non-western born) and chronic disease. But sensitivity analysis established as probable that the association of selected variables could not be totally explained by selection bias.

The main strength of this study is that we have complete information from both attendees and non-attendees for selected variables. The main weakness is that it is difficult to generalise to other populations, although we may assume little bias in odds ratio estimates of variables of the same nature as disability benefit when the observation is limited to the attendees.

Competing interest

None declared.

Authors' contributions

AJS was the project manager of the Oslo Health Study, participated in design and co-ordination of the analyses, drafted parts of the manuscript and co-ordinated the writings, RS participated in planning of the study, in design and co-ordination of the analyses, did all the statistical analyses and drafted parts of the manuscript, EB participated in planning of the study, in design and co-ordination of the analyses and drafted parts of the manuscript, DT conceived of the Oslo Health Study, participated in planning of the study, in design and co-ordination of the analyses, and drafted parts of the manuscript.

All authors read and approved the final manuscript.

Acknowledgement

We would like to thank the population of Oslo for their willingness to contribute to this important study. The National Health Screening Service of Norway – now the Norwegian Institute of Public Health, conducted the practical part of the data collection.

References

1. Townsend P, Davidson N: *The Black report* Harmondsworth: Pelican Books; 1982.
2. Macintyre S: **The Black Report and beyond: what are the issues?** *Soc Sci Med* 1997, **44**:723-45.
3. Lynch JW, Smith GD, Kaplan GA, House JS: **Income inequality and mortality: importance to health of individual income, psychosocial environment, or material conditions.** *BMJ* 2000, **320**:1200-4.
4. Marmot M, Wilkinson RG: **Psychosocial and material pathways in the relation between income and health: a response to Lynch et al.** *BMJ* 2001, **322**:1233-6.
5. Mackenbach JP, Kunst AE, Cavelaars AE, Groenhouf F, Geurts JJ, and the EU Working Group on Socioeconomic Inequalities in Health: **Socioeconomic inequalities in morbidity and mortality in western Europe.** *Lancet* 1997, **349**:1655-9.
6. Rognerud MA, Kruger O, Gjertsen F, Thelle DS: **Strong regional links between socio-economic background factors and disability and mortality in Oslo, Norway.** *Eur J Epidemiol* 1998, **14**:457-63.
7. Wilhelmsen L, Tibblin G, Werko L: **A primary preventive study of Gothenburg, Sweden.** *Prev Med* 1972, **1**:153-60.

8. Jacobsen BK, Thelle DS: **The Tromsø Heart Study: responders and non-responders to a health questionnaire, do they differ?** *Scand J Soc Med* 1988, **16**:101-4.
9. Osler M, Schroll M: **Differences between participants and non-participants in a population study on nutrition and health in the elderly.** *Eur J Clin Nutr* 1992, **46**:289-95.
10. Bostrom G, Hallqvist J, Haglund BJ, Romelsjo A, Svanstrom L, Diderichsen F: **Socioeconomic differences in smoking in an urban Swedish population. The bias introduced by non-participation in a mailed questionnaire.** *Scand J Soc Med* 1993, **21**:77-82.
11. Launer LJ, Wind AW, Deeg DJ: **Nonresponse pattern and bias in a community-based cross-sectional study of cognitive functioning among the elderly.** *Am J Epidemiol* 1994, **139**:803-12.
12. Jackson R, Chambless LE, Yang K, Byrne T, Watson R, Folsom A, Shahar E, Kalsbeek W, for the Atherosclerosis Risk in Communities (ARIC) Study Investigators: **Differences between respondents and nonrespondents in a multicenter community-based study vary by gender and ethnicity.** *J Clin Epidemiol* 1996, **49**:1441-46.
13. Belsby L, Vedø A: **Non-response analysis of the Norwegian Health Survey 1995 [Frafallsanalyse av Helseundersøkelsen 1995].** Notater 3/98. Oslo: Statistics Norway; 1998. (in Norwegian)
14. Hoeymans N, Feskens EJ, Van Den Bos GA, Kromhout D: **Non-response bias in a study of cardiovascular diseases, functional status and self-rated health among elderly men.** *Age Ageing* 1998, **27**:35-40.
15. Reijneveld SA, Stronks K: **The impact of response bias on estimates of health care utilization in a metropolitan area: the use of administrative data.** *Int J Epidemiol* 1999, **28**:1134-40.
16. Berglund G, Nilsson P, Eriksson KF, Nilsson JA, Hedblad B, Kristenson H, Lindgarde F: **Long-term outcome of the Malmo preventive project: mortality and cardiovascular morbidity.** *J Intern Med* 2000, **247**:19-29.
17. Korkeila K, Suominen S, Ahvenainen J, Ojanlatva A, Rautava P, Heleinius H, Koskenvuo M: **Non-response and related factors in a nation-wide health survey.** *Eur J Epidemiol* 2001, **17**:991-9.
18. van Loon AJ, Tijhuis M, Picavet HS, Surtees PG, Ormel J: **Survey non-response in the Netherlands. Effects on prevalence estimates and associations.** *Ann Epidemiol* 2003, **13**:105-10.
19. van den Brandt PA, Goldbohm RA, van 't Veer P, Volovics A, Hermus RJ, Sturmans F: **A large-scale prospective cohort study on diet and cancer in The Netherlands.** *J Clin Epidemiol* 1990, **43**:285-95.
20. Livingston PM, McCarty CA, Taylor HR: **Visual impairment and socioeconomic factors.** *Br J Ophthalmol* 1997, **81**:574-7.
21. der Wiel AB, van Exel E, de Craen AJ, Gussekloo J, Lagaay AM, Knook DL, Westendorp RG: **A high response is not essential to prevent selection bias: results from the Leiden 85-plus study.** *J Clin Epidemiol* 2002, **55**:1119-25.
22. Bergstrand R, Vedin A, Wilhelmsson C, Wilhelmsen L: **Bias due to non-participation and heterogeneous sub-groups in population surveys.** *J Chronic Dis* 1983, **36**:725-8.
23. Wilhelmsen L, Ljungberg S, Wedel H, Werko L: **A comparison between participants and non-participants in a primary preventive trial.** *J Chronic Dis* 1976, **29**:331-9.
24. Norwegian Institute of Public Health. **The Oslo Health Study** [<http://www.fhi.no/tema/helseundersokelse/oslo/index.html>]
25. The Oslo Health Study: Protocol Oslo: Norwegian Institute of Public Health; 2002. (in Norwegian)
26. Statistics Norway [http://www.ssb.no/english/subjects/05/01/inntind_en/about.html]
27. Strand BH, Dalgard OS, Tambs K, Rognerud M: **Measuring the mental health status of the Norwegian population: a comparison of the instruments SCL-25 SCL-10, SCL-5 and MHI-5 (SF-36).** *Nordic J Psychiatry* 2003, **57**:113-8.
28. Derogatis LR, Lipman RS, Rickels K, Uhlenhuth EH, Covi L: **The Hopkins Symptom Checklist (HSCL): a self-report symptom inventory.** *Behav Sci* 1974, **19**:1-15.
29. Hara M, Sasaki S, Sobue T, Yamamoto S, Tsugane S: **Comparison of cause-specific mortality between respondents and nonrespondents in a population-based prospective study: ten-year follow-up of JPHC Study Cohort I. Japan Public Health Center.** *J Clin Epidemiol* 2002, **55**:150-6.
30. Bjartveit K, Foss OP, Gjervig T: **The cardiovascular disease study in Norwegian counties. Results from first screening.** *Acta Med Scand Suppl* 1983, **675**:1-184.
31. Lund-Larsen PG: **ECG in health and disease. ECG findings in relation to CHD risk factors, constitutional variables and 16-year mortality in 2990 asymptomatic Oslo men aged 40-49 years in 1972.** PhD thesis ISM-skriftserie, nr. 30. University of Tromsø, Institute of Community Medicine; 1994.
32. Bergdahl M, Ekman S, Lindberg A, Lundquist P, Rennermalm M: *The non-response monitor no. 6 [Bortfallsbarometern nr 6]* (R&D Report nr 13). Stockholm: Statistics Sweden; 1991. (in Swedish)
33. Smith T: **Changes in non-response on the US general social surveys, 1975-94.** Presented at the Fifth International Workshop on Household Survey Non-Response, Ottawa, Ontario, Canada. September 26-28 1994
34. Belsby L: **The non-response problem increases.** [Frafallsproblemet øker]. *Samfunnsspeilet* 1997, **2**:10-13. (in Norwegian)
35. Bjartveit K, Wøien G: *Cardiovascular disease risk factors in Norway. Results from surveys in 18 countries* Oslo: National Health Screening Service; 1997.
36. Helakorpi S, Uutela A, Prättälä R, Puska P: *Health behaviour among Finnish adult population, B19/1999* Helsinki: Publications of the National Public Health Institute; 1999. (in Finnish, English abstract)
37. O'Neill TV, Marsden D, Silman AJ: **Differences in the characteristics of responders and non-responders in a prevalence survey of vertebral osteoporosis.** *European Vertebral Osteoporosis Study Group.* *Osteoporos Int* 1995, **5**:327-34.
38. Lund E, Gram IT: **Response rate according to title and length of questionnaire.** *Scand J Soc Med* 1998, **26**:154-60.
39. Iglesias C, Torgerson D: **Does length of questionnaire matter? A randomised trial of response rates to a mailed questionnaire.** *J Health Serv Res Policy* 2000, **5**:219-221.
40. Keeter S, Miller C, Kohut A, Groves RM, Presser S: **Consequences of reducing nonresponse in a national telephone survey.** *Public Opin Q* 2000, **64**:125-48.
41. Curtin R, Presser S, Singer E: **The effects of response rate changes on the index of consumer sentiment.** *Public Opin Q* 2000, **64**:413-28.
42. Criqui MH, Barrett-Connor E, Austin M: **Differences between respondents and non-respondents in a population-based cardiovascular disease study.** *Am J Epidemiol* 1978, **108**:367-72.
43. Selmer S, Sogaard R, Bjertness E, Thelle D: **The Oslo Health Study. Reminding the non-responders - effects on prevalence estimates.** *Nor J Epidemiol* 2003, **13**:89-94.
44. Kleinbaum DG, Kupper LL, Morgenstern H: **Selection Bias.** In: *Epidemiologic Research* Edited by: Kleinbaum DG, Kupper LL, Morgenstern H. New York: Van Nostrand Reinhold Company Inc; 1982:194-219.
45. Heilbrun LK, Nomura A, Stemmermann GN: **The effects of nonresponse in a prospective study of cancer.** *Am J Epidemiol* 1982, **116**:353-63.
46. Benfante R, Reed D, MacLean C, Kagan A: **Response bias in the Honolulu Heart Program.** *Am J Epidemiol* 1989, **130**:1088-100.
47. Vernon SW, Roberts RE, Lee ES: **Ethnic status and participation in longitudinal health surveys.** *Am J Epidemiol* 1984, **119**:99-113.
48. Austin MA, Criqui MH, Barrett-Connor E, Holdbrook MJ: **The effect of response bias on the odds ratio.** *Am J Epidemiol* 1981, **114**:137-43.
49. Eagan TM, Eide GE, Gulsvik A, Bakke PS: **Nonresponse in a community cohort study: predictors and consequences for exposure-disease associations.** *J Clin Epidemiol* 2002, **55**:775-81.
50. Rothman KJ, Greenland S: **Precision and validity in epidemiological studies.** In: *Modern Epidemiology* 2nd edition. Edited by: Rothman KJ, Greenland S. Philadelphia: Lippincott-Raven; 1998:115-34.
51. The Steering Committee of the Physicians' Health Study Research Group: **Preliminary Report: Findings from the aspirin component of the ongoing Physicians' Health Study.** *Engl J Med* 1988, **318**:262-4.
52. Guallar E, Hennekens CH, Sacks FM, Willett WC, Stampfer MJ: **A prospective study of plasma fish oil levels and incidence of myocardial infarction in U.S. male physicians.** *J Am Coll Cardiol* 1995, **25**:387-94.
53. Cummings SR, Nevitt MC, Browner WS, Stone K, Fox KM, Ensrud KE, Cauley J, Black D, Vogt TM: **Risk factors for hip fracture in white women. Study of Osteoporotic Fractures Research Group.** *N Engl J Med* 1995, **332**:767-73.
54. Thun MJ, Peto R, Lopez AD, Monaco JH, Henley SJ, Heath CW, Doll R: **Alcohol consumption and mortality among middle-aged and elderly U.S. adults.** *N Engl J Med* 1997, **337**:1705-14.

55. Lee IM, Hennekens CH, Berger K, Buring JE, Manson JE: **Exercise and risk of stroke in male physicians.** *Stroke* 1999, **30**:1-6.

Publish with **BioMed Central** and every scientist can read your work free of charge

"BioMed Central will be the most significant development for disseminating the results of biomedical research in our lifetime."

Sir Paul Nurse, Cancer Research UK

Your research papers will be:

- available free of charge to the entire biomedical community
- peer reviewed and published immediately upon acceptance
- cited in PubMed and archived on PubMed Central
- yours — you keep the copyright

Submit your manuscript here:
http://www.biomedcentral.com/info/publishing_adv.asp

