Int. J. Epidemiol. Advance Access published September 26, 2007

Published by Oxford University Press on behalf of the International Epidemiological Association
International Journal of Epidemiology 2007;1–8
doi:10.1093/ije/dym191

Survey error in measuring socio-economic risk factors of health status: a comparison of a survey and a census

Vincent Lorant,[1]* Stefaan Demarest,[2] Pieter-Jan Miermans[2] and Herman Van Oyen[2]

Accepted 22 August 2007

Background Individuals of lower socio-economic status (SES) are less likely to participate in health surveys than individuals of a higher SES. It is, however, not known whether this difference in participation is associated with health status. This study sets out to assess whether a population health survey gives biased estimates of socio-economic inequalities in self-reported health.

Methods We compared two independent cross-national data collections, a national health interview survey ($n = 10\ 164$) and a census ($n = 8\ 491\ 528$), both carried out in Belgium in 2001 and posing the same health question. We computed the prevalence ratios of poor subjective health among socio-economic groups. To estimate the bias, a relative odds ratio (ROR) was computed as the ratio of the survey prevalence ratio to the census prevalence ratio.

Results Less-educated individuals had a lower risk of poor health status in the survey [Prevalence ratio = 1.66, 95% confidence interval (CI): 1.48–1.86] than in the census (Prevalence ratio = 2.23) leading to an underestimation of the risk associated with low education (ROR = 0.74, 95% CI 0.66–0.83). Compared with better-off groups, those who were not working or who were less educated were generally less likely to participate in the survey when they had a poor health status.

Conclusions Overall, the health survey underestimated the effects of low SES on poor health status, due to selection bias. We conclude that strategies to improve participation among disadvantaged socio-economic groups should be identified.

Keywords Selection bias, health surveys, censuses, socio-economic factors, public policy

Introduction

Over the last decade, the number of new health surveys has increased, with 37 national health interview surveys being carried out in Europe between 2000 and 2004.[1] These surveys provide useful information for designing and assessing the implementation of public health policies;[2] more recently, health surveys have been used to track health inequalities in the United States[2] and in Europe.[3]

However, the participation rates in such surveys have been shown to vary widely.[4,5] A recent review of European health interview surveys reported participation rates ranging from 52% to 95%.[1] Besides, participation rates have been declining in recent decades, particularly for population-based studies.[6] Although the US National Health Interview Survey has had rather higher participation rates, these have also been declining steadily over the last decade, from 96% in 1990 to 87% in the 2004 survey. If the participation rate is low to moderate, the risk of selection bias becomes serious.[7] However, the link between participation and survey error is not that straightforward and has recently been questioned. Comparing participation rates between subgroups provides poor information about possible biases[8] and studies with low responses may even be less biased than studies with high response rates.[9] If participation difference between socio-economic groups is unrelated to health status, then participation will not affect results in terms of socio-economic inequalities in health. Conversely, if individuals of lower socio-economic status

[1] Public Health School, Université Catholique de Louvain, Belgium.

[2] Epidemiology Unit, Scientific Institute of Public Health, Belgium.

* Corresponding author. Public Health School, Université Catholique de Louvain, Belgium. E-mail: vincent.lorant@uclouvain.be

(SES) are less likely to participate in a survey than well-off individuals when they have poorer health status, then these socio-economic inequalities are likely to be underestimated.

Previous studies have shown that well-off socio-economic groups are more likely to participate than disadvantaged groups.[10–14] In relation to health status, the results are, however, inconsistent: participation tends to be higher for individuals who have used health care[10,15–17] or are severely ill;[18] but other studies have evidenced lower participation in individuals with a poorer health status[11,12,14,19] and follow-up studies have shown that individuals refusing to participate have a higher mortality risk than participants.[11,20,21] It is, therefore, unclear whether health surveys provide unbiased estimates of socio-economic risk factors for poor health status.

In 2001, Belgium carried out both a census and a health interview survey with the same health status questions delivered to the same target population. Using this unique opportunity, we assess whether a health survey provides biased estimates of socio-economic inequalities in self-reported health. We aimed, firstly, to compare the results of the census with those of the survey, and, second, to measure the selection bias in self-reported health status among different socio-economic groups.

Methods

Study population

The study compared data collected by means of a health interview survey to those collected by a census. The health interview survey and the census were both carried out in 2001 and the target population for both included all individuals with their main residence in Belgium. The health survey was a multi-stage stratified household sampling; the number of households contacted was 9011 and 5533 participated (participation rate = 61.4%). For the census, 4 248 502 households were contacted and 4 101 890 replied (participation rate = 96.5%). The health survey was a voluntary, face-to-face interview at the residence of the household, whereas the census was delivered by post and was compulsory. The detailed methodology of the survey is provided elsewhere.[22] The census and health survey questionnaires can be accessed at http://www.sesa.ucl.ac.be/questionnaires Analysis of the census and health survey data was restricted to individuals over 15 years of age and living in private households (census $n = 8\ 491\ 528$ vs health survey $n = 10\ 164$).

Measures

Both the census and the health survey questionnaire included a widely used and validated question about subjective health status ('How is your health in general? Very bad, bad, fair, good, very good').[23] Subjective health was classified into two groups (very bad, bad and fair vs good and very good). We selected socio-economic covariates known to be important predictors of self-rated health and available in both questionnaires: housing tenure, educational level, working status, nationality and household type. Although education, working status and housing tenure are commonly used to measure SES, we cannot discount the possibility that their association with self-reported health may be vulnerable to residual confounding, particularly as income data was not available to us. However, our aim in this study is to compare SES risk factors from the survey and the census: although residual confounding may exist, it should apply to both the survey and the census.

Data analysis

Our analytical strategy involved four steps. First, the distributions of the health survey and the census were compared. Differences were tested using a Chi-squared, taking the census as the theoretical distribution. Second, using a log-binomial regression,[24] we compared the prevalence ratios for poor health status among the different socio-economic groups for both the census and the survey, controlling for age and sex. Following Kleinbaum, we computed the relative odds ratio (ROR) as the ratio of the prevalence ratio in the survey to the prevalence ratio in the census.[25] The confidence interval (CI) of the ROR was computed using the formula of Nohr and colleagues.[26] The CDC guidelines for the measurement of inequalities state that disparities should be measured in both absolute and relative terms.[27] We thus computed rate difference as the difference between the standardized prevalence in the census and the standardized prevalence in the survey. The prevalences were directly standardized using the census 2001 age and sex distribution.

Next, we computed the participation fraction for each socio-economic group, stratified by health status. If α is the participation fraction (m/N) of those exposed (low SES) and having poor health status, β the participation fraction of those non-exposed (high SES) and having poor health status, δ the participation fraction of those exposed and having good health and γ the participation fraction of those non-exposed and having good health, with m and N, respectively, the number of observations in the sample and the number of observations in the whole population, then, according to Kleinbaum and colleagues, selection bias arises whenever the participation odds of the healthy ($R_h = \delta/\gamma$) are different from the participation odds of the non-healthy ($R_{nh} = \alpha/\beta$)[25] so that

$$\text{Bias} = \frac{\alpha\gamma}{\delta\beta} - 1 = \left(\frac{R_{nh} - R_h}{R_h}\right)$$

However, for legal reasons, we were not able to match each individual record of the health survey with its corresponding individual census record. Matching, therefore, had to be carried out at the group level: combining the selected socio-demographic and health variables yielded 2048 groups. For each group, we computed the number of sampling observations from the survey (m, weighted according to the health survey sampling design) and the number of observations in the population ($=N$). Merging these two tables allowed us to compute the sampling fraction for each group. We then used logistic regression to investigate the effect of the socio-economic covariates on the sampling fraction, stratifying our analyses by health status. The dependent variable was the sampling fraction, that is the ratio ($(m_k/N_k)\ k = 1,\ldots,2048$) of the number of individuals participating in the health survey (m_k) to the number of individuals registered in the census (N_k) within each k group. The independent variables were the socio-economic covariates, controlled for age and sex. All computations were carried out using SAS version 9.

Table 1 Demographic distributions of the participants in the census and in the health survey, aged 15 years or more, Belgium, 2001: percentage[a] and chi-square test

Label	Percentage health survey[b] ($n = 10\,164$)	Percentage census ($n = 8\,491\,528$)	Ratio of relative frequencies[d]	Chi-square[c]	*P*-value
Sex				0.1	0.796
Male	48.6	48.4	1.00		
Female	51.4	51.6	1.00		
Age group				538.6	<0.001
15–24 years	14.5	22.5	0.64		
25–44 years	35.8	27.6	1.30		
45–64 years	29.8	29.4	1.01		
65+ years	19.9	20.5	0.97		
Household type				0.1	0.796
Alone	15.9	16.0	0.99		
Not alone	84.1	84.0	1.00		
Nationality				106.7	<0.001
Belgian	94.2	91.3	1.03		
Other	5.8	8.7	0.67		
Housing tenure				170.3	<0.001
Owner	74.8	68.8	1.09		
Tenant	25.2	31.2	0.81		
Working status				3.0	0.082
Working	49.6	48.7	1.02		
Not working	50.4	51.3	0.98		
Education				199.6	<0.001
Primary	21.6	20.8	1.04		
Lower secondary	20.4	26.5	0.77		
Upper secondary	30.4	29.4	1.03		
Superior	27.6	23.3	1.18		
Subjective health				129.4	<0.001
Poor health	22.8	28.0	0.81		
Good health	77.2	72.0	1.07		

[a]Non-standardized percentages.
[b]The health survey distribution is weighted according to the sampling design.
[c]The census is the reference distribution.
[d]%Survey/%census.

Health survey observations were weighted according to the sampling design. Analysis weighted in this way provides valid point estimates but may lead to slight underestimation of variance, because of clustering at the household level. Sensitivity analysis showed however that the CIs of our results were minimally affected. This is mainly due to the fact the number of clusters (=households) is big and the number of units (=individuals) within a cluster (=household) limited (maximum 4).

Quality of the data sources

In these initial steps, the census was considered as the benchmark, an assumption that needed to be verified. Indeed, it could be that a trade-off occurred between two types of bias:[28] while a survey is more at risk of selection bias when participation is modest, conversely, a census may be at risk of information bias due to the manner in which the information is collected.[29, 30] Indeed, because the census was compulsory, and was delivered and returned by post, there was concern that respondents would return their questionnaire with more data missing or inconsistent data.[31] As a consequence, in our final fourth step, we computed the percentages of item non-response for the whole sample or population as well as for hard-to-reach groups, such as the elderly, less-educated individuals and non-Belgian nationals.[32] We also proceeded to check the consistency of the responses by comparing the percentage of unexpected responses, i.e. the percentage of participants who reported continuous daily restrictions on their activities but nonetheless declared that they had very good health.

Results

The census and the health survey had a slightly different distribution for age, nationality, housing tenure and education (Table 1). Compared with the census, the survey had a lower

Table 2 Socio-economic risk factors of poor subjective health among the participants in the survey and the participants in the census, aged 15 years or more, Belgium, 2001: prevalence ratios for the census and the health survey and relative odds ratios[a]

	Survey PR ($n=9254$) 95% CI	Census PR ($n=7\,874\,126$)	ROR[b] 95% CI
Household type			
Not alone (ref.)	1.0	1.0	1.0
Alone	1.07 (0.99, 1.17)	1.12	0.96 (0.88, 1.04)
Nationality			
Belgian (ref.)	1.0	1.0	1.0
Other	1.23 (1.07, 1.43)	1.18	1.04 (0.90, 1.21)
Housing tenure			
Owner (ref.)	1.0	1.0	1.0
Tenant	1.25 (1.17, 1.34)	1.25	1.00 (0.94, 1.08)
Working status			
Working (ref.)	1.0	1.0	1.0
Not working	1.88 (1.69, 2.10)	2.52	0.75 (0.67, 0.83)
Education			
Primary	1.66 (1.48, 1.86)	2.23	0.74 (0.66, 0.83)
Lower secondary	1.72 (1.52, 1.93)	1.86	0.92 (0.82, 1.04)
Upper Secondary	1.13 (1.00, 1.29)	1.45	0.78 (0.69, 0.89)
Superior (ref.)	1.0	1.0	1.0

[a]PR, Prevalence ratios controlled for age group and sex; CI, confidence interval.
[b]ROR, relative risk ratio = PR survey/PR Census.

percentage of younger participants, of participants with a lower educational level, of non-Belgians and of tenants. There were no or few differences in distribution for sex, household type, or working status (Chi-squared *P*-values >0.08). Compared with the census, the health survey had a lower prevalence of poor subjective health (Table 1).

Table 2 shows the prevalence ratio of poor subjective health by socio-economic group for both the survey (first column) and the census (second column), controlling for age and sex. The last column gives the ROR: a value below one indicates that the survey had a lower prevalence ratio than the census's prevalence ratio. Poor health status was less strongly associated with not working in the survey than in the census (ROR = 0.75, 95% CI 0.67–0.83). Likewise, participants with primary education only had a lower risk of poor health status in the survey than in the census (ROR = 0.74, 95% CI 0.66–0.83). For household type, nationality and housing tenure there were few differences in the prevalences ratio.

We stratified this analysis by sex (results not shown), for primary education, and for those not working. For education, women had a higher ROR of 0.86 (95% CI 0.72–1.03), compared with a ROR of 0.68 for men (95% CI 0.58–0.80); the reverse was found among women not working, with a lower ROR of 0.69 (95% CI 0.60–0.80), compared with 0.82 for men (95% CI 0.70–0.95).

The standardized prevalences of poor subjective health are displayed in Table 3 for both the survey and the census, with the rate difference computed as the difference between the census and the survey. All rate differences but one was positive, indicating that the survey had a lower prevalence than the census. Absolute differences were higher for individuals living alone (rate difference: 6.6%), for tenants

Table 3 Socio-economic risk factors of poor subjective health among the participants in the survey and the participants in the census, aged 15 years or more, Belgium, 2001: age and sex standardized[a] prevalence for the census and the health survey and rate difference

	Poor subjective health standardised prevalence[a]		
	Health survey (%) ($n=9254$)	Census (%) ($n=7\,874\,126$)	Rate difference (census-survey)
Household type			
Not alone (ref.)	24.1	27.0	2.9
Alone	28.4	35.0	6.6
Nationality			
Belgian (ref.)	24.1	27.6	3.5
Other	32.8	33.5	0.7
Housing tenure			
Owner (ref.)	22.9	25.6	2.7
Tenant	29.3	34.5	5.2
Working status			
Working (ref.)	18.7	18.0	−0.7
Not working	31.7	39.4	7.7
Education			
Primary	33.0	38.9	5.9
Lower secondary	29.8	29.8	0.0
Upper Secondary	21.3	23.7	2.4
Superior (ref.)	16.7	17.8	1.1

[a]Age and sex direct standardization with census 2001 population distribution as reference.

Table 4 Participation risk in the health survey by subjective health and socio-economic status among the participants in the census, aged 15 years or more, Belgium, 2001: odds ratio and selection bias with 95% CI

	Participation in the survey among those with poor health	Participation in the survey among those with good health	Selection bias
	OR[a] 95% CI	OR[a] 95% CI	Bias[b] 95% CI
Household type			
Not alone (ref)	1.00	1.00	
Alone	0.99 (0.90–1.10)	1.06 (0.99–1.14)	−0.07 (−0.18, 0.04)
Nationality			
Belgian (ref)	1.00	1.00	
Other	0.60*** (0.50–0.73)	0.67*** (0.61–0.74)	−0.11 (−0.42, 0.21)
Housing tenure			
Owner (ref)	1.00	1.00	
Tenant	0.74*** (0.68–0.82)	0.89*** (0.85–0.94)	−0.17** (−0.29, −0.05)
Working status			
Working (ref)	1.00	1.00	
Not working	0.71*** (0.63–0.79)	1.32*** (1.25–1.40)	−0.47*** (−0.56, −0.38)
Education			
Primary	0.55*** (0.48–0.63)	1.67*** (1.57–1.79)	−0.67*** (−0.75, −0.59)
Lower secondary	0.63*** (0.55–0.72)	0.68*** (0.63–0.73)	−0.07 (−0.30, 0.15)
Upper Secondary	0.73*** (0.63–0.84)	0.96 (0.90–1.02)	−0.24** (−0.40, −0.09)
Superior (ref)	1.00	1.00	

[a]OR, odds ratios controlled for age and sex; CI, confidence interval.

[b]Bias = (OR among poor health − OR among good health)/OR among good health; Standard error of the ratio computed by way of the delta method: $\mathrm{Var}(OR_{nh}/OR_h) = (1/OR_h^2) \times (\mathrm{Var}(OR_{nh}) + (OR_{nh}/OR_h)^2 \times \mathrm{Var}(OR_h)$.

$^{*}P < 0.05$; $^{**}P < 0.01$; $^{***}P < 0.001$

(rate difference: 5.2%), for individuals not working (rate difference: 7.7%) and for those with primary education only (rate difference: 5.9%).

The odds of participation by health and SES are provided in Table 4. The first column provides the participation odds ratios (ORs) among those declaring a poor health status, the second column among those declaring good health. The last column displays the bias as the relative difference between the first and the second column (as a percentage of the second) and its 95% CI. A negative difference implies an underestimation of poor health status among the group in comparison with the reference category (the participation OR is greater among healthier individuals), while a positive difference implies an overestimation (the participation odds OR is greater in the non-healthy category). All estimates were controlled for age group and sex. Compared with Belgians, foreigners were slightly less likely to participate when they had poor health (−0.11, 95% CI −042, 0.21). Tenants with a poor health status were less likely to participate than tenants with a good health status, resulting in an underestimation of the association between housing tenure and health (bias = −0.17, 95% CI −0.29, −0.25). Individuals not working were more likely to participate when they had good subjective health (OR = 1.32, 95% CI 1.25–1.40) than when they had a poor health status (OR = 0.71, 95% CI 0.63–0.79), resulting in an underestimation of the risk of poor health status among those not working (bias = −0.47, 95% CI −0.56, −0.38). Finally, compared with the more highly educated, participants with primary education only were less likely to participate when they had poor subjective health (OR = 0.55, 95% CI: 0.48–0.63) and were more likely when they reported good subjective health (OR = 1.67, 95% CI 1.57–1.79). As a consequence, the risk of poor health was underestimated among the less educated (bias = −0.67, 95% CI −0.75, −0.59).

Finally, we compared the quality of information provided by the census and the survey. For the subjective health question, there were fewer non-responses in the census than in the survey (census: 2.6% vs survey: 7.8%). Among the elderly, the risk of non-response to any of these three questions was higher in the census (OR = 1.12, 95% CI 1.11–1.14, Table 5) than in the survey (OR = 0.92, 95% CI 0.65–1.31). The opposite was true for the less educated and for foreigners: the risk of any non-response was higher in the survey than in the census. Finally, there were similar degrees of inconsistency in the responses to the census and the survey: among those declaring continuous limitation due to a long-term restricting illness, 0.9% declared very good health in the census and 2.6% in the survey.

Discussion

We found that lower educated and not working individuals were less likely to participate in the survey when they had a poor health status compared with the better-off groups, leading to an underestimation of the effect of low SES on poor subjective health.

The lower participation for disadvantaged socio-economic groups when they have a poorer health status fits the

Table 5 Item non-response to subjective health question according to age group, educational status, and nationality among the participants in the census and the participants in the health survey, aged 15 years or more, Belgium, 2001: odds ratio with 95% CI

	Non-response to the subjective health question in the census ($n = 9254$)	**Non-response to the subjective health question in the health survey ($n = 787426$)**
Variable	OR[a] 95% CI[a]	OR[a] 95% CI
Age group (years)		
15–24	1.10*** (1.07–1.12)	2.60*** (1.65–4.09)
25–44 (ref.)	1.0	1.0
45–64	0.97** (0.96–0.99)	0.84 (0.58–1.20)
65+	1.12*** (1.11–1.14)	0.92 (0.65–1.31)
Education		
Primary/low secondary	1.27*** (1.25–1.29)	2.27*** (1.51–3.42)
Upper secondary	1.05*** (1.03–1.08)	0.69 (0.40–1.17)
Superior (ref.)	1.0	1.0
Nationality		
Belgian (ref.)	1.0	1.0
Other	1.14*** (1.12–1.16)	1.75* (1.07–2.87)

[a]Multivariate odds ratios controlled for all variables included in the table. CI, confidence interval.
*$P < 0.05$; **$P < 0.01$; ***$P < 0.001$

leverage-saliency theory. This predicts that a survey attribute (such as the topic or the question of privacy) will have a different effect on the decision to participate in different persons.[33] The reduced participation rate of the disadvantaged socio-economic groups when they have a poorer health status may be explained by two factors: an involvement effect and a privacy effect. Although low socio-economic groups are big health care consumers, this is mainly because of their poor health status. For a given health status, groups with a lower SES tend to use health care less, particularly specialized care.[34,35] These timid health care consumers might thus display a lower willingness to participate in health surveys, presumably for the same reasons that patients under using health care are less willing to get involved in research addressing health status.[18] Another explanation may relate to lifestyle factors, such as smoking or being sedentary. These risk factors are known to be more common in less well-educated individuals and to reduce survey participation.[21,36,37]Given the importance of lifestyle in the health survey and the current increasing public pressure on smoking, particularly in public facilities, it is possible that lower SES individuals may have declined participation for fear of stigma. This is consistent with the fact that privacy concerns are the leading cause for non-response in health surveys[38] and with other studies showing that those with less education or with lower SES are less likely to respond to surveys addressing health behaviours.[13,39] This hypothesis was checked by computing the non-response percentage for the smoking question in the health interview survey: item non-response for this question was higher in less educated than in more highly educated individuals (1.3% vs 0.5% $\chi^2 = 16.6$, $P < 0.01$). This hypothesis is also supported by the lower prevalence of poor subjective health in the survey compared with the census prevalence of poor subjective health: if, as found in previous research, riskier lifestyles increase both the risk of non-participation and the risk of poor health status, then the survey may also be more likely to underestimate the prevalence of poor subjective health.

Up to now, increasing the participation rate has been the main objective of population surveys. Our study, however, suggests that survey participation is not necessarily a good predictor of survey error: e.g. although working individuals participated as much as inactive individuals, the health status of the latter was underestimated. Conversely, tenants had a lower participation rate but their risk ratio of poor health was not particularly biased. This finding is consistent with recent survey research: as suggested by a recent review by Robert Groves, the slope between participation rate and survey error is mostly flat.[8] Although there is evidence that non-participation bias does occur, it tends to vary greatly within a survey.[8] An overall survey participation rate is thus a poor indicator of survey bias, which can be quite small for some groups and important for others, and depends, as we have shown, on their health status.

The selection bias was computed on the assumption that the census was a complete and valid denominator. Since the participation rate for the census was not 100%, this assumption needs to be verified and we need to acknowledge the possibility of a selection bias for the census itself. Indeed, our results could be affected if the census non-participants had a much higher risk of poor health status than the non-participants in the health survey. Since the census was linked to the national register, it was at least possible to check the association between age and non-participation in the census. It turned out that non-participation in the census had, as in the health survey, a U-shaped association with age: it decreased from 4% in the 15–24-year group to 1.2% in the 25–44-year group and then increased to 2.1% for the 65+ group, presumably because the elderly are more likely to be institutionalized. As a consequence, because non-participation in the census was small, and higher in the younger age group than in the elderly, and because this age pattern was similar to the age pattern observed in the health survey, we are confident that non-participation in the census would not jeopardize our conclusions.

A second limitation concerns the study design. The census was a self-completed questionnaire delivered by post, while the survey was face-to-face. As a consequence, differences in socio-demographic risk factors for poor health status may be due not only to selection bias but also to the difference in administration mode. There is indeed some evidence that the mode of questionnaire administration influences the distribution of health status responses: participants are less likely to choose extreme categories of responses to the health status questions (very good or very bad) in a self-completed questionnaire than in a face-to-face interview, so that the former mode of administration will display a more concentrated distribution than the latter.[40] However, for two reasons, we are confident that the administration mode will have influenced our conclusions only modestly. First, a previous study suggested that the administration mode only slightly influenced the results in terms of health inequalities.[41]

Moreover, because a self-completed questionnaire would display a more concentrated distribution than a face-to-face interview, at worst we may have slightly underestimated the differences between the census (a self-completed questionnaire) and the survey (face-to-face) regarding socio-demographic inequalities in health.

Another limitation comes from the matching of the census and the survey at the group level (Table 4). The validity of the selection bias results may be vulnerable either to ecological error or misclassification. Ecological error would occur if the association at the group level did not reflect a true association at the individual level. Two elements, however, makes this risk rather low: first, the sizes of the selection biases were in line with the relative risk ratios, so that the analysis at the group level (Table 4) was consistent with the individual-level analysis (Table 2); second, with 2048 groups, the level of aggregation was quite low, reducing the risk of ecological error. Nevertheless, with 2048 groups, the risk of misclassification cannot be discounted: indeed, if the definitions of the variables in the survey and in the census were not strictly the same, there could be a risk that individuals classified, e.g. as having low educational status in the survey might be classified differently in the census. Is this risk important? We think this risk is rather low, for three reasons: first, we carefully chose limited socio-demographic variables whose definitions matched perfectly in both the survey and the census. Second, for some covariates (age, sex, nationality and household composition), information came directly from the national register for both the census and the survey and was thus not vulnerable to reporting error. Third, although some measurement error may exist, our results would be affected only if both of the following two conditions are met: (i) misclassification of SES is linked to health status; (ii) this difference is not a random individual error but is systematic for an important group of individuals. To assess the threat to validity, we carried out a sensitivity analysis: we compared the selection bias for low education in a table with only 32 groups (two genders, four age groups, four educational groups) with the initial results estimated for the 2048 groups. In the 32-group analysis, OR of selection for the less educated (compared with the more highly educated) were 0.56 and 1.68, respectively for the unhealthy and the healthy, leading to a bias of −0.67; these results were much the same as those that we found with 2048 groups.

Finally, we may question whether our Belgian setting can be generalized to other populations, particularly as the participation rate for our health interview survey was lower than the rates reported for the 2004 US Health Interview Survey (87%) or the 2003 UK Health Survey (73%). However, such participation rates are common in Europe[1] and the design of the Belgian health interview survey is largely consistent with other European surveys. A study reviewing participation in epidemiological studies has found, moreover, that lower SES is also likely to affect participation in epidemiological studies, both in the United States and in a number of European countries.[39,42]

Conclusions

Inequalities in health were placed at the forefront of the European Union's National Action Plans at the Lisbon European Council.[43] Our study shows that caution is required regarding the magnitude of socio-demographic inequalities in self-rated health as measured by surveys. Strategies to improve participation in health surveys have been identified.[44,45] In order to improve relevance and acceptance among hard-to-reach audiences, one possible avenue would be to avoid having the same script for all participants and, instead, to address more specifically their concerns, e.g. by addressing privacy concerns or providing incentives.

Acknowledgements

Belgian Scientific Policy, grant no TA/00/15; Observatoire de la Santé et du Social, convention 2003-2005. The authors are grateful to Geert Molenbergh, of the Hasselt Centre for Statistics, for his comments and encouragement; all remaining errors are ours.

Conflict of interest: None declared.

KEY MESSAGES

- Individuals of lower SES are less likely to participate in health surveys than individuals of a higher SES. It is, however, not known whether this difference in participation is associated with health status.
- Compared with better-off groups, lower socio-economic groups, particularly those who were not working or who were less educated, were generally less likely to participate in the survey when they had a poor health status, leading to an underestimation of the socio-economic risk factors.
- Our study shows that caution is required regarding the magnitude of socio-demographic inequalities in self-rated health as measured by health surveys.

References

[1] Aromaa A, Koponen P, Tafforeau J, Vermeire C. Evaluation of health interview surveys and health examination surveys in the European Union. *Eur J Public Health* 2003;**13:**67–72.

[2] United States, Dept.of Health and Human Services. *Tracking Healthy People 2010*. Washington, DC: U.S. Department of Health and Human Services, 2000.

[3] Kunst AE, Bos V, Lahelma E *et al*. Trends in socioeconomic inequalities in self-assessed health in 10 European countries. *Int J Epidemiol* 2005;**34:**295–305.

[4] Wolf HK, Kuulasmaa K, Tolonen H *et al*. *Participation Rates, Quality of Sampling Frames and Sampling Fractions in the MONICA Surveys*. Helsinky: National Public Health Institute, 1998.

[5] O'Neill TW, Marsden D, Matthis C, Raspe H, Silman AJ. Survey response rates: national and regional differences in a European multicentre study of vertebral osteoporosis. *J Epidemiol Community Health* 1995;**49:**87–93.

[6] Hartge P. Participation in population studies. *Epidemiology* 2006;**17:**252–54.

[7] Armstrong BK, White E, Saracci R. Principles of exposure measurement in epidemiology. *Monographs in Epidemiology and Biostatistics*. New York: Oxford University Press, 1995.

[8] Groves RM. Nonresponse rates and nonresponse bias in household surveys. *Public Opin Q* 2006;**70:**646–75.

[9] Stang A, Jockel KH. Studies with low response proportions may be less biased than studies with high response proportions. *Am J Epidemiol* 2004;**159:**204–10.

[10] Freudenstein U, Arthur AJ, Matthews RJ, Jagger C. Community surveys of late-life depression: who are the non-responders? *Age Ageing* 2001;**30:**517–21.

[11] Goldberg M, Chastang JF, Leclerc A *et al*. Socioeconomic, demographic, occupational, and health factors associated with participation in a long-term epidemiologic survey: a prospective study of the French GAZEL cohort and its target population. *Am J Epidemiol* 2001;**154:**373–84.

[12] 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.

[13] Purdie DM, Dunne MP, Boyle FM, Cook MD, Najman JM. Health and demographic characteristics of respondents in an Australian national sexuality survey: comparison with population norms. *J Epidemiol Community Health* 2002;**56:**748–53.

[14] van den Akker M, Buntinx F, Metsemakers JF, Knottnerus JA. Morbidity in responders and non-responders in a register-based population survey. *Fam Pract* 1998;**15:**261–63.

[15] Kjoller M, Thoning H. Characteristics of non-response in the Danish Health Interview Surveys, 1987-1994. *Eur J Public Health* 2005;**15:**528–35.

[16] Lamers LM. Medical consumption of respondents and non-respondents to a mailed health survey. *Eur J Public Health* 1997;**7:**267–71.

[17] 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.

[18] Rupp I, Triemstra M, Boshuizen HC, Jacobi CE, Dinant HJ, van den Bos GA. Selection bias due to non-response in a health survey among patients with rheumatoid arthritis. *Eur J Public Health* 2002;**12:**131–35.

[19] 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.

[20] Andersen LB, Vestbo JÃ, Juel K *et al*. A comparison of mortality rates in three prospective studies from Copenhagen with mortality rates in the central part of the city, and the entire country. *Eur J Epidemiol* 1998;**14:**579–85.

[21] Jousilahti P, Salomaa V, Kuulasmaa K, Niemela M, Vartiainen E. Total and cause specific mortality among participants and non-participants of population based health surveys: a comprehensive follow up of 54 372 Finnish men and women. *J Epidemiol Community Health* 2005;**59:**310–15.

[22] Demarest S, Tafforeau J, Van Oyen H *et al*. *Health Interview Survey 2001: Protocol for the Sampling Design*. Brussels: Scientific Institute of Public Health, 2001.

[23] Idler EL, Benyamini Y. Self-rated health and mortality: a review of twenty-seven community studies. *J Health Soc Behav* 1997;**38:**21–37.

[24] Spiegelman D, Hertzmark E. Easy SAS calculations for risk or prevalence ratios and differences. *Am J Epidemiol* 2005;**162:**199–200.

[25] Kleinbaum DG, Kupper LL, Morgenstern H. *Epidemiologic Research: Principles and Quantitative Methods*. Belmont, Calif: Lifetime Learning Publications, 1982.

[26] Nohr EA, Frydenberg M, Henriksen TB, Olsen J. Does low participation in cohort studies induce bias? *Epidemiology* 2006;**17:**413–18.

[27] Keppel KG, National Center for Health Statistics. *Methodological Issues in Measuring Health Disparities*. Hyattsville, Md: U.S. Dept. of Health and Human Services, Centers for Disease Control and Prevention, National Center for Health Statistics, 2005.

[28] Rothman KJ, Greenland S. *Modern Epidemiology*. Philadelphia, PA: Lippincott-Raven, 1998.

[29] Rothman KJ, Greenland S. Precision and validity in epidemiologic studies. In: Rothman KJ, Greenland S (eds). *Modern Epidemiology*. Philadelphia, PA: Lippincott-Raven, 1998.

[30] Boyle P, Dorling D. Guest editorial: the 2001 UK census: remarkable resource or bygone legacy of the 'pencil and paper era'? *Area* 2004;**36:**101–10.

[31] Groves RM, Couper M. *Nonresponse in Household Interview Surveys*. New York: Wiley, 1998.

[32] Freimuth VS, Mettger W. Is there a hard-to-reach audience? *Public Health Rep* 1990;**105:**232–38.

[33] Groves RM, Singer E, Corning A. Leverage-saliency theory of survey participation - description and an illustration. *Public Opin Q* 2000;**64:**299–308.

[34] van Doorslaer E, Wagstaff A, Rutten F. *Equity in the Finance and Delivery of Health Care: An International Perspective*. Oxford: Oxford Medical Publications, 1993.

[35] van Doorslaer E, Wagstaff A, van der BH *et al*. Equity in the delivery of health care in Europe and the US. *J Health Econ* 2000;**19:**553–83.

[36] Hill A, Roberts J, Ewings P, Gunnell D. Non-response bias in a lifestyle survey. *J Public Health* 1997;**19:**203–7.

[37] Korkeila K, Suominen S, Ahvenainen J *et al*. Non-response and related factors in a nation-wide health survey. *Eur J Epidemiol* 2001;**17:**991–99.

[38] Bates N, Piani A. *Participation in the National Health Interview Survey: Exploring Reasons for Reluctance Using Contact History Process Data*. Washington: U.S.Census Bureau, 2005.

[39] Helasoja V, Prattala R, Dregval L, Pudule I, Kasmel A. Late response and item nonresponse in the Finbalt Health Monitor Survey. *Eur J Public Health* 2002;**12:**117–23.

[40] Crossley TF, Kennedy S. The reliability of self-assessed health status. *J Health Econ* 2002;**21:**643–58.

[41] Clarke PM, Ryan C. Self-reported health: reliability and consequences for health inequality measurement. *Health Econ* 2006;**15:**645–52.

[42] Galea S, Tracy M. Participation Rates in Epidemiologic Studies. *Ann Epidemiol* 2007;**17:**643–53. Epub 2007 June 6.

[43] Atkinson AB. *Social Indicators: The EU and Social Inclusion*. Oxford: Oxford University Press, 2002.

[44] Groves RM. *Survey Nonresponse*. New York: Wiley, 2002.

[45] Edwards P, Roberts I, Clarke M *et al*. Increasing response rates to postal questionnaires: systematic review. *Br Med J* 2002;**324:**1183.