

BMJ Open Registry-based analysis of participator representativeness: a source of concern for sickness absence research?

Marit Knapstad,^{1,2} Jesper Löve,³ Kristina Holmgren,³ Gunnel Hensing,³ Simon Øverland^{1,4}

To cite: Knapstad M, Löve J, Holmgren K, *et al.* Registry-based analysis of participator representativeness: a source of concern for sickness absence research?. *BMJ Open* 2016;**6**:e012372. doi:10.1136/bmjopen-2016-012372

► Prepublication history for this paper is available online. To view these files please visit the journal online (<http://dx.doi.org/10.1136/bmjopen-2016-012372>).

Received 21 April 2016

Revised 20 June 2016

Accepted 30 August 2016



CrossMark

¹Department of Health Promotion, Norwegian Institute of Public Health, Bergen, Norway

²Department of Clinical Psychology, University of Bergen, Bergen, Norway

³Department of Public Health and Community Medicine, The Sahlgrenska Academy, University of Gothenburg, Gothenburg, Sweden

⁴Department of Psychosocial Science, University of Bergen, Bergen, Norway

Correspondence to

Dr Marit Knapstad;
marit.knapstad@fhi.no

ABSTRACT

Objectives: Selective participation can bias results in epidemiological surveys. The importance of health status is often suggested as a possible explanation for non-participation but few empirical studies exist. In a population-based study, explicitly focused on sickness absence, health and work, we examined whether a history of high levels of sickness absence was associated with non-participation.

Design: The study is based on data from official sickness absence registers from participants, non-participants and the total target population of the baseline survey of the Health Assets Project (HAP).

Setting: HAP is a population-based cohort study in the Västra Götaland region in South Western Sweden.

Participants: HAP included a random population cohort (n=7984) and 2 cohorts with recent sickness absence (employees (n=6140) and non-employees (n=990)), extracted from the same overall general working-age population.

Primary outcome measures: We examined differences in participation rates between cohorts (2008), and differences in previous sickness absence (2001–2008) between participants (individual-level data) and non-participants or the target population (group-level data) within cohorts.

Results: Participants had statistically significant less registered sickness absence in the past than non-participants and the target population for some, but not all, of the years analysed. Yet these differences were not of substantial size. Other factors than sickness absence were more important in explaining differences in participation, whereby participants were more likely to be women, older, born in Nordic countries, married and have higher incomes than non-participants.

Conclusions: Although specifically addressing sickness absence, having such experience did not add any substantial layer to selective participation in the present survey. Detailed measures are needed to gain a better understanding for health selection in health-related surveys such as those addressing sickness absence, for instance in order to discriminate between selection due to ability or motivation for participation.

Strengths and limitations of this study

- Selective participation by history of sickness absence was examined employing official registries of sickness absence across 8 years. Such health data have rarely been applied in former studies on survey representativeness.
- The sickness absence data on participants, non-participants and the target population alike are based on all reimbursements from the Social Insurance Agency, and are not self-reported, which is a strength with regard to common methodological problems such as attrition and recall bias.
- Since data from a population-based survey were employed, the observed results may reflect general tendencies concerning selective survey participation.
- Both recent and more distant sickness absence were included as predictors for participation, which may provide evidence on representativeness of participants concerning both recent time and recurrent sickness absence.
- The study does not investigate mechanisms driving an association between sickness absence and survey participation, such as obstacles or motivations, which also are important to clarify to provide decision support for how to best approach potential participants.

INTRODUCTION

Sickness absence is a major challenge and policy development requires high-quality and unbiased data. In sickness absence research, surveys and cohort studies remain important to gain better understanding of variations in level, causes, consequences and mechanisms of sickness absence across social groups and gender. A crux of any survey is to ensure sample representativeness; if participants are different from non-participants in the variables of interest, estimates may suffer from bias.^{1 2} The declining participation rates in epidemiological surveys observed across

Western countries in the past 30 years are therefore worrying.^{3 4} Registry data can circumvent issues regarding participation, but often lack the required depth of information for sickness absence research to move forward. Consequently, knowledge about selective survey participation and, in particular, concerning the key variable, sickness absence, is needed to provide researchers with decision support in how to contact participants and, perhaps more importantly, to evaluate the accuracy of results from such surveys.

In surveys across topics, demographic factors such as female gender, being married and higher socio-economic position are consistently found to predict survey participation,^{5–9} whereas the evidence regarding age groups and ethnicity are less conclusive.¹⁰ Existing evidence further suggests health selection whereby participants have better general^{6 8 11 12} and mental health,¹³ are less likely to be on^{6–8} or at risk for disability pension award,¹¹ and also have a higher life expectancy¹⁴ than non-participants. Studies of health status and survey participation have mostly examined rare health-related events (such as hospitalisation), or severe or long-lasting illness (like disability pension award and mortality). Barriers and selection mechanisms may be different in these cases than for sickness absence, which is common in the entire population, fluctuate and, in the majority of cases, concern common musculoskeletal and mental illnesses. Sickness absence is moreover a measure of health that reflect aspects related to functional and working ability, which might be more relevant than diagnoses in explaining survey behaviour.

If and how sickness absence predicts survey participation is uncertain. Linkages to administrative registries are expedient, as they enable unbiased and complete data from participants and non-participants.¹⁵ Of the few studies having employed such data, some have found that participants have lower sickness absence rates than non-participants, in line with health selection to survey participation.^{5 16–18} Others have found this among men only,⁹ or report weak¹⁹ or no⁷ association between sickness absence and survey participation. The unequivocal findings may relate to variations in measures and follow-up time, as well as complex selection mechanisms involving reachability, ability and motivation to participate.²⁰

Concerning motivation, it is commonly proposed that people will be more prone to participate if the survey topic is relevant for them personally.^{20–23} In interviews with participants and responding non-participants, perceived value or personal gain of contributing to advances in research in the topic has been highlighted as decisive.^{22 24} Following this line of thought, studies addressing sickness absence should lead to increased inclusion of current and previous sickness absentees. Direct measures of relevance are difficult to obtain in representative samples of study participants, and a feasible compromise is to match characteristics of sampled

individuals and the core topic of the survey, and infer topic relevance via these characteristics.²⁵ Based on this approach, personal relevance selection is found through randomised controlled designs,²⁶ observed by the general experience that cases are easier to recruit than controls in case-control studies¹⁰ and relating to consent giving in medical record follow-ups.²⁷

Only one study has addressed personal relevance selection in surveys on sickness absence specifically¹⁶ in which, in contrast to the personal relevance hypothesis, participants were found to have *less* sickness absence than non-participants. Owing to a small study population from one company only, the finding might not be generalisable to a general population context.

Taken together, it remains empirically unsettled whether sickness absence history influences survey participation and, in particular, in surveys where sickness absence is the main topic. The general decreasing participation rates call for studies that can provide a basis for how to approach potential participants in the future. In the current study, we analysed associations between registered sickness absence and survey participation in a large population-based survey-linkage study that explicitly focused on sickness absence (the Health Assets Project, HAP). HAP started in 2008 with the main aim of comparing workers with sickness absence experiences to those without such experience concerning health, work life and family affairs. To this end, a unique feature of HAP was the use of a 'case-control' sampling technique, sampling two cohorts with a recent, new sickness absence episode (employees and non-employees) in addition to a random population cohort (not recent sick-listed 'controls'), all extracted from a working age population of the Västra Götaland region in Sweden. This technique has, for example, enabled studies of differences in individual and structural factors between workers with and without sickness absence^{28 29} and predictors of return to work.^{30 31} The data collection included links to official registries covering demographics and sickness absence days per year across 9 years (2001–2009), extracted at an individual level for participants and group level for the target populations for each of the three cohorts. This specific design allowed for examining our research aim through the following research questions:

1. Were the participation rates higher in the two cohorts with a recent, new episode of sickness absence (employees and non-employees) than in the random population cohort?
2. Within each of the three cohorts, respectively, did participants have more sickness absence days annually in the years preceding the survey (2001–2008) than non-participants or the target population?
3. Within each of the three cohorts, respectively, were the proportions of individuals with registered sickness absence annually in the years preceding the survey (towards 2001) higher among participants than non-participants or the target population?

METHODS

This study is based on registry data from participants, non-participants and the target population of the baseline survey of HAP 2008. **Figure 1** depicts the sampling procedure in HAP, which specific components that are compared and data available for each component in the current study.

Target population and cohorts in HAP

The study base in HAP was the working age population (19–64 years) in Västra Götaland in Sweden, a region with both urban and rural areas comprising 17% of the Swedish population. In Sweden, all inhabitants are covered by the national sickness insurance. For employees, the employer covers the first 14 days of a sickness absence episode (except one qualifying day); thereafter, benefits are granted from the Social Insurance Agency (SIA). Non-employed (eg, self-employed, unemployed and students) can apply through self-report for benefit from SIA for sickness absence beyond 1 day. SIA thus has registries of all covered sickness absence beyond 14 days for employees and beyond 1 day for non-employees. With help from SIA and Statistics Sweden, the following three cohorts were extracted from the study base to obtain groups with and without recent sickness absence (see also **figure 1** and ref. 28 for more details):

1. A recent sick-listed cohort of *employees* (employee cohort), of which the target population consisted of all employed individuals with a *new* sickness absence episode >14 days during 18 February to 15 April 2008 (n=12 543).
2. A recent sick-listed cohort of *non-employees* (non-employee cohort), where the target population included all other insured with a *new* sickness absence episode >1 day during 18 February to 1 April 2008 (n=5004). The sampling frame for these cohorts only included those *registered* in SIA by 15 April 2008 (n=6140 in the employee cohort and n=4240 in the non-employee cohort), as the survey ideally should be conducted as close as possible to the current absence episode. In the employee cohort, the total sampling frame was invited to participate (n=6140), whereas a random sample of the non-employee-sampling frame was invited (n=990).
3. Finally, a *random population* cohort (population cohort, n=7984) was invited. A negative coordination was performed to ensure non-overlapping cohorts; thus, the population cohort included no cases with *new* registrations of sickness absence *during* inclusion.

Data collection: Eligible participants were invited through a postal survey, sent out on 15th and 25th of April 2008 with two reminders (ie, up until 2 months after onset of the registered sickness absence episode for

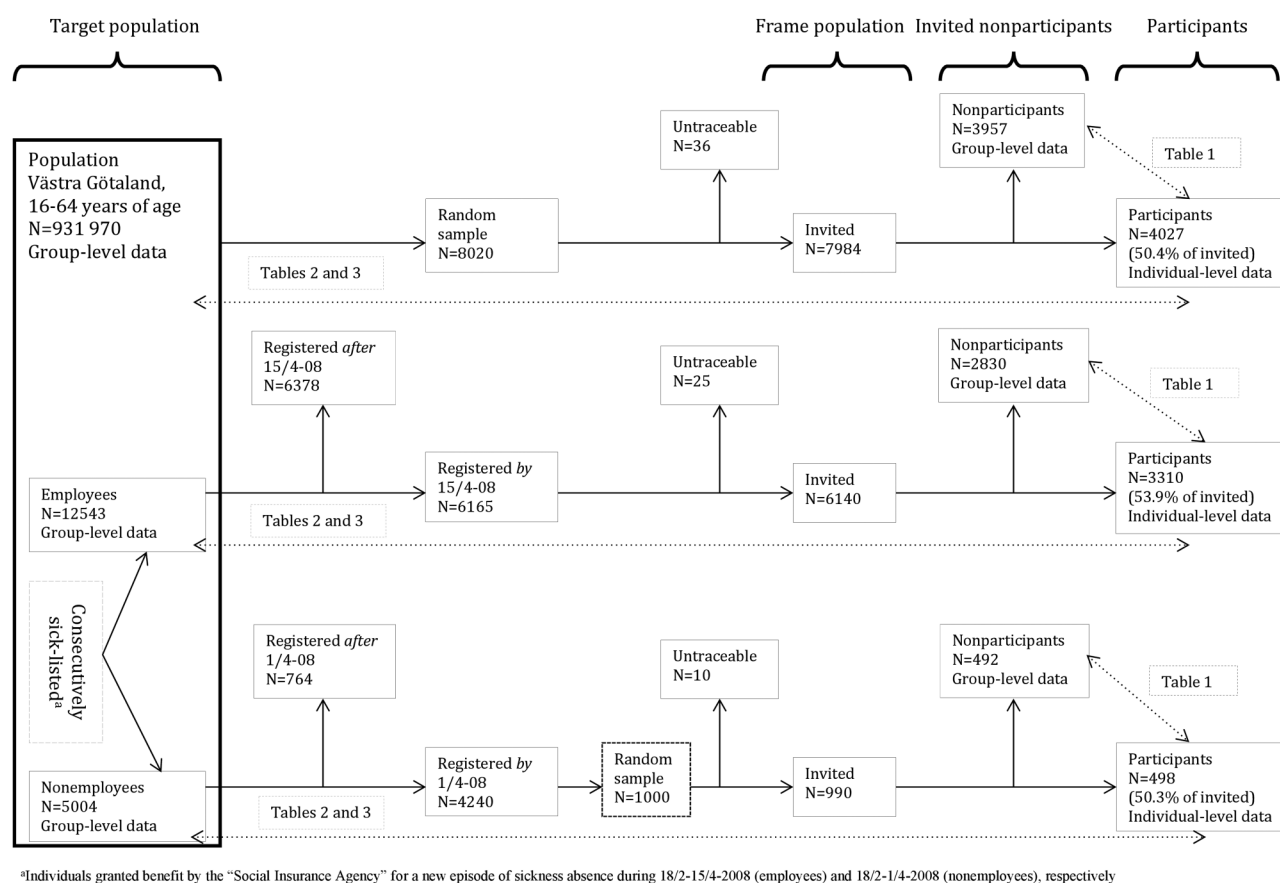


Figure 1 Flow chart of inclusion procedures in the Health Assets Project (HAP).

the 2 sick-listed cohorts). The invitation letter included a description of the study aim, data collection procedures, contact details and information that withdrawal from the study was possible at any time. It was explicitly stated that the SIA would not have access to information on participation status and that participation would not affect the invitee's sickness allowance. Participants gave informed consent to link survey data to official registry data on sociodemographic factors, sickness absence and employment status. For the current study, we extracted the corresponding registry data for each of the three cohorts' target populations, which are officially available at a grouped level.

In the following, the registry data employed in the current study will be described in more detail, including amendments made to enable comparisons between the individual-level data (participants) and group-level data (non-participants and target population).

Data source and measures on demographic variables

Regarding demographic variables, group-level data from all *invited* were extracted from Statistics Sweden: *Participation* (yes, no), *gender* (male, female), *age group* (19–30, 31–50, 51–64), *country of birth* (Nordic, others), *marital status* (married, not married (includes cohabitants)) and *gross income in intervals* (Swedish Krona (SEK) ≤149 000, 150 000–299 000, ≥300 000).

Data sources on registered sickness absence

Data on sickness absence benefit granted from SIA during the years 2001–2008 were extracted from the 'Longitudinal integrated database for sickness insurance and labour market research (LISA)'. The data included annual number of reimbursed sickness absence days (including sickness absence, rehabilitation and work injury allowance¹). Data on participants were available at an individual level and data on the target populations at a group level, distributed by gender and age groups (employee cohort and non-employee cohort: age groups 19–30, 31–50, 51–64; Västra Götaland population: age groups 20–29, 30–39, 40–49, 50–59 for data on sickness absence days and 16–19, 20–24, 25–29, 30–34, 35–39, 40–44, 45–49, 50–54, 55–59 for data on sickness absence cases).

To achieve appropriate comparison groups, the following accommodations were made: First, since the data from the target populations for the employee cohort and the non-employee cohort included those *granted* reimbursement, we excluded participants with no registered sickness absence days in 2008 from the participant groups. Second, to approximate non-participation groups, we subtracted participants in the employee cohort and non-employee cohort from their respective target populations. Finally, we handled problems with

age-related left censoring back in time (towards 2001) by only including those aged 31–64 in 2008 in the employee cohort and non-employee cohort. In the population cohort, to correspond to available official statistics, we included participants aged 20–59 per calendar year when comparing sickness absence days, and participants aged 16–59 per calendar year when comparing sickness absence cases.

Measures on registered sickness absence

1. *Participation rates between cohorts.* As a first crude step, we examined whether participation rate in the two cohorts with a recent registered sickness absence episode (employee cohort and non-employee cohort) differed from that in the population cohort.
2. *Days with registered sickness absence annually.* We compared mean number of registered sickness absence days per year (2001–2008) between participants and non-participants (employee cohort and non-employee cohort) and the target population (all 3 cohorts).
3. *Proportions with previous sickness absence annually.* Finally, we compared the proportion of individuals with registered sickness absence per year between participants and non-participants (employee cohort and non-employee cohort, 2001–2007) or the target population (population cohort, 2001–2008).

Statistical analyses

The data were analysed using Microsoft Excel 2010 and Stata V.12. Differences in participation rate and distribution across demographic characteristics between participants and non-participants in each of the three cohorts were examined as relative proportions and χ^2 tests for group-level data. Regarding sickness absence, we first compared participation rates with 95% CIs between the cohorts and performed χ^2 tests for group-level data. Second, we performed one sample mean comparison Student's t-test to examine differences in mean number of sickness absence days per year, and across years, from 2001 until 2008 between participants and their comparison groups in each cohort, respectively. To account for gender and age differences between the comparison groups, we calculated means weighted for the distribution in the respective participant groups. Finally, to compare proportions with registered sickness absence per year, gender-stratified ORs (95% CIs) were calculated comparing participants and their comparison groups in each cohort, respectively.

RESULTS

Demographic characteristics of participants and non-participants

Table 1 displays demographic characteristics and participation rates across groups between participants and invited non-participants in the three cohorts. Participants were more likely than non-participants to be women, older, born in Nordic countries, married and

¹The Västra Götaland general population statistics did not include work injury allowance, but this is regarded negligible for the analyses due to small numbers.

Table 1 Demographic distribution and participation rates across groups between participants and invited non-participants in the three cohorts included in the Health Assets Project

	Random population cohort				Recent sick-listed employee cohort				Recent sick-listed non-employee cohort			
	Participants n (%)	Invited non-participants n (%)	Participation rate Per cent	Difference*	Participants n (%)	Invited non-participants n (%)	Participation rate Per cent	Difference*	Participants n (%)	Invited non-participants n (%)	Participation rate Per cent	Difference*
Total	4027	3957	50.4		3310	2830	53.9		498	492	50.3	
Gender				$\chi^2=143.9$ df=1				$\chi^2=81.9$ df=1				$\chi^2=9.5$ df=1
Women	2234 (55.5)	1664 (42.1)	57.3		2196 (66.3)	1558 (55.1)	58.5		325 (65.3)	274 (55.7)	54.3	
Men	1793 (44.5)	2293 (57.9)	43.9	p<0.001	1114 (33.7)	1272 (44.9)	46.7	p<0.001	173 (34.7)	218 (44.3)	44.2	p=0.002
Age group, years				$\chi^2=129.8$ df=2				$\chi^2=121.4$ df=2				$\chi^2=2.4$ df=2
19–30	830 (20.6)	1175 (29.7)	41.4		380 (11.5)	516 (18.2)	42.4		114 (22.9)	116 (23.6)	49.6	
31–50	1803 (44.8)	1799 (45.5)	50.1	p<0.001	1479 (44.7)	1428 (50.5)	50.9	p<0.001	257 (51.6)	271 (55.1)	48.7	p=0.295
51–64	1394 (34.6)	983 (24.8)	58.6		1451 (43.8)	886 (31.3)	62.1		127 (25.5)	105 (21.3)	54.7	
Country of birth				$\chi^2=138.4$ df=1				$\chi^2=6.1$ df=1				$\chi^2=6.6$ df=1
Nordic	3642 (90.4)	3216 (81.3)	53.1		2985 (90.2)	2497 (88.2)	54.5		444 (89.2)	411 (83.5)	51.9	
Others	385 (9.6)	741 (18.7)	34.2	p<0.001	325 (9.8)	333 (11.8)	49.4	p=0.014	54 (10.8)	81 (16.5)	40.0	p=0.010
Marital status				$\chi^2=175.2$ df=1				$\chi^2=66.0$ df=1				$\chi^2=2.1$ df=1
Married	1877 (46.6)	1414 (35.7)	57.0		1705 (51.5)	1164 (41.1)	59.4		220 (44.4)	240 (48.8)	47.8	
Not married	2150 (53.4)	2543 (64.3)	45.8	p<0.001	1605 (48.5)	1666 (58.9)	49.1	p<0.001	278 (55.8)	252 (51.2)	52.5	p=0.146
Income (SEK)				$\chi^2=179.7$ df=2				$\chi^2=37.1$ df=2				$\chi^2=3.4$ df=2
≤149 000	987 (24.5)	1496 (37.8)	39.8		329 (9.9)	405 (14.3)	44.8		178 (35.7)	204 (41.5)	46.6	
150 000– 299 000	1920 (47.7)	1678 (42.4)	53.4	p<0.001	2219 (67.0)	1892 (66.9)	54.0	p<0.001	254 (51.0)	229 (46.5)	52.6	p=0.181
≥300 000	1120 (27.8)	783 (19.8)	58.9		762 (23.0)	533 (18.8)	58.8		66 (13.3)	59 (12.0)	52.8	

*Differences examined using Chi-square tests for group-level data.
SEK, Swedish Krona.

have higher incomes in both the population cohort and the recent sick-listed employee cohort. The demographic distribution was more even between participants and non-participants in the non-employee cohort, though participants were more likely than non-participants to be women and to be born in Nordic countries.

Differences in participation rates between cohorts

The participation rate was 3.5 percentage points higher in the employee cohort (53.9%, 95% CI 52.7% to 55.2%) than in the population cohort (50.4%, 95% CI 49.3% to 51.5%) ($\chi^2=16.75$, $df=1$ $p<0.001$). The participation rate in the non-employee cohort (50.3%, 95% CI 47.2% to 53.5%) was similar to that among the population cohort ($\chi^2=0.00$, $df=1$, $p=0.936$). As detailed in [table 1](#), there were more variations overall in participation rates across demographic groups within cohorts than between the cohorts.

Differences in mean days of registered sickness absence between participants and comparison groups within cohorts

Overall, there were no substantial differences in registered sickness absence between participants and their comparison groups across the three cohorts. Participants in the population cohort had a lower mean number of sickness absence days per year than the corresponding level in the population in the years 2001–2003. Weighted for gender and age distribution among participants, the differences were statistically significant through 2001–2008, except 2007. Yet the raw differences in annual mean number of registered sickness absence days only ranged from 1.7 to 5.3 days ([table 2](#)). The same tendency was found in the employee cohort; however, it was only statistically significant when comparing participants to non-participants in the years 2001–2003 and 2007, weighted for gender and age distribution ([table 2](#)). By contrast, participants in the non-employee cohort had a higher mean number of sickness absence days per year than non-participants and the target population in 2008 and 2007, gender and age weighted ([table 2](#)).

Differences in proportions with registered sickness absence between participants and comparison groups within cohorts

Regarding individuals with registered sickness absence per year, the proportions were lower overall among participants than non-participants or the target population. In the population cohort, compared with the target population, participants had statistically significant lower odds for having had an episode of sickness absence only in 2001 and 2003 for women, and in 2001, 2002 and 2003 for men (ORs ranging from 0.84 to 0.91 for women and 0.76 to 0.80 for men, [table 3](#)). In the employee cohort, compared with non-participants, participants had statistically significant lower odds for having had an episode of sickness absence most of the

comparisons per years from 2001 to 2007 (ORs ranging from 0.87 to 0.95 for women and 0.77 to 0.88 for men, see [table 3](#)). The corresponding comparisons in the non-employee cohort resulted in small and generally non-significant differences, and in opposing directions for men and women ([table 3](#)).

DISCUSSION

Main results

Participants in the HAP study, which specifically invited people to a survey on sickness absence, health and work, had less registered sickness absence in the past than non-participants and the target population in some but not all of the years analysed. The differences found in sickness absence were moreover not of substantial size. Secondary findings harmonise with commonly observed differences in sociodemographic characteristics as participants were more likely than non-participants to be women, older, born in Nordic countries, married and have higher incomes.

Strengths and limitations

The main strengths of this study were chiefly related to our application of objective registry data on sickness absence from participants and the target population. First, this enabled investigation of selection effects by sickness absence, which has rarely been achievable in prior research and restricted in many countries by lack of available registries. This study examined sickness absence history across more years than in previous studies. Many non-participation analyses on health variables are based on supplementary surveys of ‘participating non-participants’ willing to complete a shortened version of the survey, with the inherent risk of partly reproducing the same non-participation bias.³² Second, the use of registries reduced common methodological problems such as recall bias and missing responses.¹⁵ Third, since the registry data are based on financial reimbursement from the SIA, they are considered to be accurate and reliable. Finally, examining sickness absence several years before the survey is a particular advantage when studying selection by sickness absence, as the phenomenon on the one hand is common, with a 1 year cumulative incidence of 11.3% in the working population in Western Sweden in 2008³³ and, on the other hand, in some cases is prolonged and recurrent. Thus, the findings might inform representativeness of participants regarding both present time and prolonged or recurrent cases. Additionally, most studies on sickness absence as a predictor for survey participation have employed specific occupational^{5 9 16 18} or diagnostic groups.³⁴ These groups may have specific distributions of sickness absence and demography making the observed results not necessarily applicable to other groups. Since this study examined population-based cohorts, the results may to a greater extent be regarded as general tendencies. Despite considerable

Table 2 Differences in mean days of registered sickness absence, annually 2001–2008, between the participants and comparison groups within each of the three cohorts included in the HAP

Year	Participants			Non-participants		Target population	
	n	Mean days	95% CI	Mean days crude (raw difference†)	Mean days weighted‡ (raw difference†)	Mean days crude (raw difference†)	Mean days weighted‡ (raw difference†)
Random population cohort§							
2008	3379	8.5	7.0 to 9.9	–	–	9.6 (1.1)	10.2 (1.7)*
2007	3426	11.9	10.2 to 13.7	–	–	12.2 (0.3)	13.1 (1.2)
2006	3451	12.7	10.9 to 14.5	–	–	14.0 (1.3)	15.0 (2.3)*
2005	3477	14.2	12.4 to 16.1	–	–	15.7 (1.5)	16.8 (2.6)**
2004	3519	16.2	14.2 to 18.2	–	–	17.7 (1.5)	19.0 (2.8)**
2003	3538	17.6	15.4 to 19.8	–	–	20.4 (2.8)*	21.9 (4.3)**
2002	3468	17.0	14.9 to 19.2	–	–	21.0 (4.0)**	22.2 (5.2)**
2001	3384	14.9	12.9 to 16.8	–	–	19.3 (4.4)**	20.2 (5.3)**
Recent sick-listed employee cohort¶							
2008	2676	81.8	78.3 to 85.3	78.3 (–3.5)	78.8 (–3.0)	79.2 (–2.6)	79.7 (–2.1)
2007	2676	20.3	18.2 to 22.5	22.5 (2.2)*	23.0 (2.7)*	22.0 (1.7)	22.3 (2.0)
2006	2672	29.4	26.5 to 32.7	30.3 (0.9)	31.3 (1.9)	30.1 (0.7)	30.9 (1.5)
2005	2666	34.3	31.0 to 37.5	33.7 (–0.6)	35.1 (0.8)	33.8 (–0.5)	34.9 (0.6)
2004	2661	33.2	29.9 to 26.4	34.6 (1.4)	36.3 (3.1)	34.2 (1.0)	35.4 (2.2)
2003	2658	32.0	28.8 to 35.3	35.8 (3.8)*	37.7 (5.7)**	34.8 (2.8)	36.0* (3.8)
2002	2650	30.4	27.4 to 33.4	31.9 (1.5)	33.9 (3.5)*	31.4 (1.0)	32.7 (2.3)
2001	2644	24.4	21.7 to 27.1	26.6 (2.2)	28.4 (4.0)**	26.0 (1.6)	27.1 (2.7)
Total	2639	287.2	273.2 to 301.2	293.8 (6.6)	304.3 (17.1)	291.5 (4.3)*	299.0 (11.8)
Recent sick-listed non-employee cohort¶							
2008	277	68.3	57.9 to 78.7	55.7 (–12.6)*	56.8 (–11.5)*	56.6 (–11.1)*	57.6 (–10.7)*
2007	277	49.5	37.7 to 61.4	32.8 (–16.7)**	32.9 (–16.6)**	33.9 (–15.6)*	34.2 (–15.3)*
2006	276	47.0	35.0 to 58.9	36.1 (–11.5)	35.9 (–11.1)	36.8 (–10.2)	36.8 (–10.2)
2005	275	47.6	35.8 to 59.4	39.1 (–0.8)	39.4 (–8.2)	39.6 (–8.0)	39.8 (–7–8)
2004	275	39.9	28.8 to 51.0	40.7 (0.8)	41.2 (1.3)	40.6 (0.7)	40.9 (1.0)
2003	275	36.1	25.9 to 46.3	41.5 (4.1)	42.1 (6.0)	41.1 (5.0)	41.5 (5.4)
2002	273	37.4	26.7 to 48.1	35.6 (–1.8)	36.1 (–1.3)	35.7 (–1.7)	36.1 (–1.3)
2001	272	27.3	18.9 to 35.8	27.3 (0.0)	27.7 (0.4)	27.3 (0.0)	27.7 (0.4)
Total	271	358.2	304.7 to 411.7	308.8 (–45.4)	311.7 (–46.5)	311.7 (–46.5)	314.6 (–43.6)

*p<0.05; **p<0.01. Differences in means examined employing one-sample Student's t-tests.

†Raw difference=mean days non-participants or target population—mean days participants.

‡Weighted for gender and age distribution among HAP participants.

§Participants aged 20–59 in the respective calendar years are compared with the corresponding age groups in the Västra Götaland population (target population).

¶Only age group 31–64 (per 2008) included to avoid age-related left censoring when going back in time towards 2001. Among participants, only those having ≥1 day of registered sickness absence in 2008 are included to achieve equal inclusion criterion as for the non-participation group. Non-participants comprise all individuals granted benefit by the SIA for a new spell of sickness absence during the inclusion period (target population), excluding participants.

HAP, Health Assets Project; SIA, Social Insurance Agency.

Table 3 Gender-stratified proportions and OR (95% CI) for participants in each cohort compared with non-participants or target population for having had at least one registered sickness absence episode, each year 1–7 years prior to the HAP survey

	Women				Men			
	Part Per cent	Target population† or non-participants‡ Per cent	Difference		Part Per cent	Target population† or non-participants ‡ Per cent	Difference	
			OR	95% CI			OR	95% CI
Random population cohort†								
2008	10.5	11.7	0.89	0.77 to 1.03	7.4	7.1	1.05	0.86 to 1.28
2007	13.7	13.5	1.01	0.89 to 1.15	7.5	8.1	0.93	0.76 to 1.12
2006	14.3	15.0	0.95	0.84 to 1.08	8.2	8.9	0.91	0.75 to 1.09
2005	15.4	16.1	0.95	0.84 to 1.07	9.2	9.4	0.97	0.82 to 1.15
2004	14.9	16.2	0.90	0.80 to 1.02	8.6	9.3	0.92	0.77 to 1.10
2003	15.9	18.4	0.84	0.75 to 0.95**	8.4	10.7	0.76	0.64 to 0.91**
2002	18.8	20.2	0.91	0.82 to 1.02	9.9	12.1	0.80	0.68 to 0.94**
2001	17.5	19.7	0.84	0.76 to 0.95**	9.5	11.7	0.80	0.67 to 0.94**
Recent sick-listed employee cohort‡								
2008	100.0	100.0	–	–	100.0	100.0	–	–
2007	29.5	32.6	0.87	0.77 to 0.98*	26.8	29.5	0.87	0.74 to 1.04
2006	31.5	34.2	0.88	0.79 to 0.99*	26.2	28.8	0.88	0.74 to 1.04
2005	32.6	33.8	0.95	0.84 to 1.07	23.1	28.2	0.77	0.64 to 0.92**
2004	28.8	32.3	0.85	0.75 to 0.95**	20.6	25.6	0.75	0.63 to 0.91**
2003	30.0	33.7	0.84	0.75 to 0.95**	22.8	26.4	0.82	0.68 to 0.98*
2002	31.7	34.5	0.88	0.78 to 0.99*	21.3	25.9	0.77	0.64 to 0.93**
2001	29.3	31.0	0.92	0.81 to 1.04	20.0	24.4	0.77	0.64 to 0.93**
Recent sick-listed non-employee cohort‡								
2008	100.0	100.0	–	–	100.0	100.0	–	–
2007	52.4	53.4	0.96	0.69 to 1.34	54.0	49.2	1.21	0.81 to 1.81
2006	40.2	46.8	0.77	0.55 to 1.07	52.7	42.3	1.52	1.02 to 2.28*
2005	37.2	42.9	0.79	0.56 to 1.11	47.8	38.5	1.46	0.92 to 2.19
2004	33.5	40.3	0.75	0.52 to 1.05	36.0	36.1	1.00	0.65 to 1.51
2003	34.8	40.8	0.77	0.54 to 1.09	38.7	36.2	1.12	0.73 to 1.69
2002	35.0	38.1	0.87	0.62 to 1.23	44.6	32.9	1.64	1.08 to 2.47*
2001	32.1	32.6	0.98	0.68 to 1.39	36.4	30.5	1.32	0.86 to 2.02

*p<0.05, **p<0.01.

†Participants aged 16–59 in the respective calendar years are compared with the corresponding age groups in the Västra Götaland population (target population).

‡Only participants with ≥1 day of registered sickness absence in 2008 are included to achieve equal inclusion criterion as for the non-participation group. Non-participants comprise all individuals granted benefit by the SIA for a new spell of sickness absence during the inclusion period (target population), excluding participants. Only age group 31–64 included avoiding age-related left censoring (towards 2001).

HAP, Health Assets Project; SIA, Social Insurance Agency.

advantages in applying registries in research, the quality and accuracy of an analysis rest on the information available. First, some participants had either no days but one or more episode of registered absence or vice versa, whereas it was uncertain as to whether there were corresponding cases among non-participants, due to the use of group-level data. This uncertainty might have produced noise in the analyses. Our results were, however, quite robust across alternative analyses of the data, strengthening our confidence in the observed findings.

Second, the skewed distribution of sickness absence days makes median calculations more appropriate than means.³⁵ The use of group-level data on the target populations precluded calculating median values and SD estimates for the comparison groups. The one-sample Student's t-test was considered a valid approach based on the data available as the Student's t-test is very robust for comparing means, and as the distribution of means, according to 'the central limit theorem', will approximate a normal distribution when the sample size increases, even when the distribution in the population is non-normal.³⁶ That said, interpreting the mean values by themselves can be problematic when the distribution of the data is skewed. Though means of sickness absence days arguably is fairly meaningful, interpretations of results should focus more on the differences in means between groups than the mean values themselves.

Third, owing to the fluctuating nature of sickness absence and lag in registry administration, our comparison groups for research question 1 were inevitably somewhat overlapping concerning sickness absence status. The population cohort naturally included some ongoing cases and some cases with onset *after* inclusion (sampling procedures ensured no new cases during inclusion, but 6.7% of the population cohort participants self-reported being currently sickness-absent). Nevertheless, since the employee cohort and non-employee cohort *all* had recent sickness absence (ie required to be included in these cohorts), the comparison of participation rates between the cohorts were regarded appropriate. As for the within cohorts comparisons, non-participants in the sick-listed cohorts comprised the respective target populations minus participants. These target populations also included some non-invited individuals due to registration in SIA after the predefined inclusion period. Lagged registration in SIA is in general slightly skewed.³⁷ A sensitivity analysis, however, revealed no differences in outcome between those invited in the first and second rounds in the employee cohort, with late registrations presumably over-represented in the latter, indicating fairly comparable sickness absence histories between the invited and non-invited non-participants (numbers not shown).

Finally, we only had access to a limited amount of variables characterising the non-participation group. Hence, we cannot rule out an impact from residual confounding, especially from socioeconomic factors,^{9, 19} on our results. The data available on income, country of birth

and marital status were retrieved separately from the sickness absence data, precluding the possibility for making statistical adjustments. The registry data did not include information on medicolegal cause or specific timing of the sickness absence episodes beyond number of registered days per year, precluding some analyses on how sickness absence might influence survey participation.

Interpretation of the findings

Selection effects by topic relevance are assumed to be a particular statistical concern as associations are more prone to be biased if selection has to do with the key statistics.^{1 10 25} Empirical tests of this assumption have thus far not found consequential impact on survey estimates analysing associations,¹ in line with most,^{6 10 11 38} though not all,^{12 39} available studies on non-participation bias. Prevalence estimates are notably more vulnerable for selection bias. Levels of registered sickness absence among participants did not diverge substantially from the target populations in HAP, and selection by sickness absence is thus not likely to be any substantial source of bias in this particular survey.

As described in the introduction, selection mechanisms in surveys are complex and involve reachability, ability and motivation to participate. Sickness absence-related motivators and barriers may have influenced participation in the opposite direction, as will be elaborated on in the following, in concert contributing to the finding of relatively similar sickness absence histories between participants and non-participants. The study design did not allow for addressing these nuances directly, but the observed results might shed light on some aspects to be addressed in more detail in future studies. Personal relevance by recent or previous sickness absence seemed *not* to be a prominent selection mechanism for this survey. Notably, the participation rate was slightly higher in the recent sickness-listed employee cohort than in the population cohort. This could be interpreted as a 'recency effect' of personal relevance selection, as the finding contrasted the results regarding more distant sickness absence. The employee cohort nevertheless also included more women than the population cohort, and as women tend to participate more than men,¹⁰ this might have contributed to the observed result. The absolute difference of 3.5% may also be considered of little practical importance. Results for the non-employee cohort diverged somewhat from the two other cohorts as well. This might be explained by numerous factors specific for this cohort, such as absence registration schemes, huge heterogeneity including students and self-employed people, and finally the small size of this sample.

The overall finding in this study seemed more to reflect a reduced health and functional capacity among non-participants, as we found somewhat less previous sickness absence among those who participated than those who did not. According to the 'health selection

hypothesis', illness precludes participation in research.^{6 8 11} Several potentially opposing mechanisms may have contributed to this finding. Naturally, current or recent sickness absence can simply entail reduced ability to participate due to poor health, fatigue, motivation or hospitalisation, even though the person under normal circumstances would be inclined to participate. Besides, social inequalities are related to both sickness absence⁴⁰ and differential participation.^{8 9} Barriers and facilitators for survey participation across social groups are not well understood, but may involve structural barriers and differences in norms and perceived social value of research.^{10 41 42} Some barriers could be specific to sickness absence: First, 'oversurveying' is suggested to contribute to explaining falling participation rates in general.¹⁰ Recurrent or long-term sickness absences requires repeated assessments of work capacity to be eligible for sickness insurance, and being approached with yet another questionnaire might not have been welcomed by some of those invited. We do not know anything about "partial participation", for example, persons who start to answer the questionnaire, which was rather substantial, but gave up due to tiredness or lack of motivation. Second, sensitive questionnaire items decrease participation rates.²⁶ Stigma and shame related to some diagnoses such as mental illnesses^{10 32} or to the sickness absence status per se⁴³ could thus have made some more hesitant to participate. In concert with this interpretation, an epidemiological survey on mental health found participants to have fewer psychotropic prescriptions than non-participants, although using more medical services for somatic disorders.³² The assurance of confidentiality in the invitation letter in the HAP study, hereunder that the questionnaire was not related to the employer or SIA, probably partly counteracted nonparticipation due to fear of "exposure",²⁶ but how much is not easily quantifiable. Diagnoses may also have yielded differences in personal relevance motivation, as the survey overall was directed more towards mental than physical aspects of work, health and sickness absence. In sum, a more direct and specified measure of perceived relevance and attitude towards the topic, although challenging to obtain, could in theory have discriminated better between individual motivations and barriers for participation.

CONCLUSION

Selective participation remains a challenge in epidemiological surveys, yet again demonstrated by demographic differences between participants and non-participants in the HAP survey. Sickness absence did not seem to add any substantial layer to the selection, based on several registry-based comparisons in the current study. Registry data is a crucial resource for increasing knowledge on selective participation. Detailed measures are needed to gain a better understanding for health selection in health-related surveys such as those addressing sickness

absence, for instance in order to discriminate between selection due to ability or motivation for survey participation. Until such studies are performed, the overall findings of this study did not give rise to much concern about the representativeness of survey participants regarding sickness absence history.

Acknowledgements The authors thank Carl Högfeldt and Ulrik Lidwall at the Swedish Social Insurance Agency for availability and help in extracting and interpreting the group-level sickness absence registry data.

Contributors MK, JL, GH, SØ and KH designed the study. MK analysed the data and wrote the first draft and main revisions of the manuscript. All authors contributed in interpretation of the data and critical revision of the manuscript, and approved the final version of the manuscript.

Funding The data collection for the Health Assets Project was supported by the Swedish Social Insurance Agency.

Competing interests None declared.

Ethics approval The HAP study was approved by the Ethics Committee at the University of Gothenburg (registration number 039-08) and conducted in accordance with the latest version of the Helsinki protocol.

Provenance and peer review Not commissioned; externally peer reviewed.

Data sharing statement No additional data are available.

Open Access This is an Open Access article distributed in accordance with the Creative Commons Attribution Non Commercial (CC BY-NC 4.0) license, which permits others to distribute, remix, adapt, build upon this work non-commercially, and license their derivative works on different terms, provided the original work is properly cited and the use is non-commercial. See: <http://creativecommons.org/licenses/by-nc/4.0/>

REFERENCES

1. Groves RM, Peytcheva E. The impact of nonresponse rates on nonresponse bias a meta-analysis. *Public Opin Q* 2008;72:167–89.
2. Keiding N, Louis TA. Perils and potentials of self-selected entry to epidemiological studies and surveys. *J R Stat Soc Ser A Stat Soc* 2016;179:319–76.
3. Tolonen H, Helakorpi S, Talala K, *et al*. 25-year trends and socio-demographic differences in response rates: Finnish adult health behaviour survey. *Eur J Epidemiol* 2006;21:409–15.
4. Curtin R, Presser S, Singer E. Changes in telephone survey nonresponse over the past quarter century. *Public Opin Q* 2005;69:87–98.
5. Kaerlev L, Kolstad HA, Hansen AM, *et al*. Are risk estimates biased in follow-up studies of psychosocial factors with low base-line participation? *BMC Public Health* 2011;11:539.
6. Lundberg I, Damström Thakker K, Hällström T, *et al*. Determinants of non-participation, and the effects of non-participation on potential cause-effect relationships, in the PART study on mental disorders. *Soc Psychiatry Psychiatr Epidemiol* 2005;40:475–83.
7. Sogaard AJ, Selmer R, Bjertness E, *et al*. The Oslo Health Study: the impact of self-selection in a large, population-based survey. *Int J Equity Health* 2004;3:3.
8. Korkeila K, Suominen S, Ahvenainen J, *et al*. Non-response and related factors in a nation-wide health survey. *Eur J Epidemiol* 2001;17:991–9.
9. 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.
10. Galea S, Tracy M. Participation rates in epidemiologic studies. *Ann Epidemiol* 2007;17:643–53.
11. Knudsen AK, Hotopf M, Skogen JC, *et al*. The health status of nonparticipants in a population-based health study: the Hordaland Health Study. *Am J Epidemiol* 2010;172:1306–14.
12. Nummela O, Sulander T, Helakorpi S, *et al*. Register-based data indicated nonparticipation bias in a health study among aging people. *J Clin Epidemiol* 2011;64:1418–25.

13. Haapea M, Miettunen J, Läärä E, *et al.* Non-participation in a field survey with respect to psychiatric disorders. *Scand J Public Health* 2008;36:728–36.
14. Ferrie JE, Kivimäki M, Singh-Manoux A, *et al.* Non-response to baseline, non-response to follow-up and mortality in the Whitehall II cohort. *Int J Epidemiol* 2009;38:831–7.
15. Nielsen MB, Knardahl S. The healthy worker effect: do health problems predict participation rates in, and the results of, a follow-up survey? *Int Arch Occup Environ Health* 2016;89:231–8.
16. Van Goor H, Verhage AL. Nonresponse and recall errors in a study of absence because of illness: an analysis of their effects on distributions and relationships. *Qual Quant* 1999;33:411–28.
17. Bergstrand R, Vedin A, Wilhelmsson C, *et al.* Bias due to non-participation and heterogenous sub-groups in population surveys. *J Chronic Dis* 1983;36:725–8.
18. Martikainen P, Laaksonen M, Piha K, *et al.* Does survey non-response bias the association between occupational social class and health? *Scand J Public Health* 2007;35:212–15.
19. Laaksonen M, Aittomäki A, Lallukka T, *et al.* Register-based study among employees showed small nonparticipation bias in health surveys and check-ups. *J Clin Epidemiol* 2008;61:900–6.
20. Dillman DA, Eltinge JL, Groves RM, *et al.* Survey nonresponse in design, data collection, and analysis. In: Groves RM, Dillman DA, Eltinge JL, *et al.*, eds. *Survey nonresponse*. New York: Wiley & Sons, 2002:3–26.
21. Wenemark M, Persson A, Brage HN, *et al.* Applying motivation theory to achieve increased response rates, respondent satisfaction and data quality. *J Off Stat* 2011;27:393–414.
22. Sinicrope PS, Patten CA, Bonnema SM, *et al.* Healthy women's motivators and barriers to participation in a breast cancer cohort study: a qualitative study. *Ann Epidemiol* 2009;19:484–93.
23. Groves RM, Singer E, Corning A. Leverage-saliency theory of survey participation: description and an illustration. *Public Opin Q* 2000;64:299–308.
24. Harcombe H, Derrett S, Herbison P, *et al.* "Do I really want to do this?" Longitudinal cohort study participants' perspectives on postal survey design: a qualitative study. *BMC Med Res Methodol* 2011;11:8.
25. Groves RM, Presser S, Dipko S. The role of topic interest in survey participation decisions. *Public Opin Q* 2004;68(1):2–31.
26. Edwards P, Roberts I, Clarke M, *et al.* Methods to increase response rates to postal questionnaires. *Cochrane Database Syst Rev* 2007 (2):MR000008.
27. Dunn KM, Jordan K, Lacey RJ, *et al.* Patterns of consent in epidemiologic research: evidence from over 25,000 responders. *Am J Epidemiol* 2004;159:1087–94.
28. Hensing G, Holmgren K, Mårdbj AC. Harmful alcohol habits were no more common in a sample of newly sick-listed Swedish women and men compared with a random population sample. *Alcohol Alcohol* 2011;46:471–7.
29. Löve J, Holmgren K, Torén K, *et al.* Can work ability explain the social gradient in sickness absence: a study of a general population in Sweden. *BMC Public Health* 2012;12:163.
30. Holmgren K, Ekblad E, Hensing G, *et al.* The combination of work organizational climate and individual work commitment predicts return to work in women but not in men. *J Occup Environ Med* 2013;55:121–7.
31. Bertilsson M, Vaez M, Waern M, *et al.* A prospective study on self-assessed mental well-being and work capacity as determinants of all-cause sickness absence. *J Occup Rehabil* 2015;25:52–64.
32. Vercambre MN, Gilbert F. Respondents in an epidemiologic survey had fewer psychotropic prescriptions than nonrespondents: an insight into health-related selection bias using routine health insurance data. *J Clin Epidemiol* 2012;65:1181–9.
33. Armannsdóttir B, Mårdbj A-C, Haukenes I, *et al.* Cumulative incidence of sickness absence and disease burden among the newly sick-listed, a cross-sectional population-based study. *BMC Public Health* 2013;13:329.
34. Nielsen MBD, Bultmann U, Amby M, *et al.* Return to work among employees with common mental disorders: Study design and baseline findings from a mixed-method follow-up study. *Scand J Public Health* 2010;38:864–72.
35. Hensing G, Alexanderson K, Allebeck P, *et al.* How to measure sickness absence? Literature review and suggestion of five basic measures. *Scand J Public Health* 1998;26:133–44.
36. Pedhazur EJ, Schmelkin LP. *Measurement, design, and analysis: an integrated approach*. Hillsdale, NJ: Laurence Erlbaum associates, 1991.
37. Högföldt C. *Fördröjning av sjukpenningsdata: en utvärdering av tremånadersregeln [Lagged sickness insurance data: an evaluation of the three months regulation]*. Stockholm: Försäkringskassan, 2013.
38. Nilsen RM, Vollset SE, Gjessing HK, *et al.* Self-selection and bias in a large prospective pregnancy cohort in Norway. *Paediatr Perinat Epidemiol* 2009;23:597–608.
39. Langhammer A, Krokstad S, Romundstad P, *et al.* The HUNT study: participation is associated with survival and depends on socioeconomic status, diseases and symptoms. *BMC Med Res Methodol* 2012;12:143.
40. Marmot M, Feeney A, Shipley M, *et al.* Sickness absence as a measure of health-status and functioning: from the UK Whitehall II study. *J Epidemiol Community Health* 1995;49:124–30.
41. Markanday S, Brennan SL, Gould H, *et al.* Sex-differences in reasons for non-participation at recruitment: Geelong Osteoporosis Study. *BMC Res Notes* 2013;6:104.
42. Fry CL. A comprehensive evidence-based approach is needed for promoting participation in health research: a commentary on Williams. *Soc Sci Med* 2008;66:1457–60.
43. Knapstad M, Øverland S, Henderson M, *et al.* Shame among long-term sickness absentees: correlates and impact on subsequent sickness absence. *Scand J Public Health* 2014;42:96–103.