

ORIGINAL ARTICLE

Self-reported health problems and sickness absence in different age groups predominantly engaged in physical work

Simo Taimela, Esa Läärä, Antti Malmivaara, Jaakko Tiekso, Harri Sintonen, Selina Justén, Timo Aro



This paper is freely available online under the BMJ Journals unlocked scheme, see <http://oem.bmj.com/info/unlocked.dtl>

Occup Environ Med 2007;**64**:739–746. doi: 10.1136/oem.2006.027789

See end of article for authors' affiliations

Correspondence to: Dr S Taimela, Evalua International, PO Box 35, FIN-01531 Vantaa, Finland; simo.taimela@evalua.fi

Accepted 26 January 2007
Published Online First
15 February 2007

Objectives: To study the associations between self-reported health problems and sickness absence from work. **Methods:** The results of a questionnaire survey were combined with archival data of sickness absence of 1341 employees (88% males; 62% blue-collar) in the construction, service and maintenance work within one corporation in Finland. Sex, age and occupational grading were controlled as confounders. A zero-inflated negative binomial (ZINB) regression model was used in the statistical analysis of sickness absence data.

Results: The prevalence of self-reported health problems increased with age, from 23% in 18–30-year-olds to 54% in 55–61-year-olds. However, in those aged 18–30 years, 71% had been absent from work and in those aged 55–61 years this proportion was 53%. When health problems and occupational grading were accounted for in the ZINB model, age as such was not associated with the number of days on sick leave, but the young workers still had higher propensity for (any) sickness absence than the old. Self-rated future working ability and musculoskeletal impairment were strong determinants of sickness absence. Among those susceptible to taking sick leave, the estimated mean number of absence days increased by 14% for each rise of 1 unit of the impairment score (scale 0–10).

Conclusions: Young subjects had surprisingly high probability for sickness absence although they reported better health than their older colleagues. A higher total count of absence days was found among subjects reporting health problems and poorer working ability, regardless of age, sex and occupational grade. These findings have implications for both management and the healthcare system in the prevention of work disability.

Sickness absence means non-attendance by an employee at work due to a (certified) health complaint when the employer expects attendance. Despite the straightforward definition, sickness absence has proved to be a complex phenomenon. In addition to illness, it has been associated with, for example, demographical and socioeconomic factors, organisational features, job content and attitudes to work.¹ The key psychosocial predictors of sickness absence include individuals' own perceptions of health and working ability.^{2–3}

It is a common belief that older (supposedly in poorer health) employees are more absent from work than their younger (supposedly healthier) colleagues.^{4–5} However, the young seem to stay out of work due to minor health complaints more than older workers. Also some earlier studies have found that older age increases the risk of overall sickness absences, but decreases that of one-day absences.⁶

We investigated how age and self-reported health problems are associated with sickness absence within a cohort predominantly employed in physical work.

METHODS

Study design and ethics

The design was cross-sectional: data from questionnaires were combined with records of demographics and sickness absence from the employer's salary register. The Helsinki University Research Ethics Board approved the study, and it was performed according to the Declaration of Helsinki.

Participants

Inclusion criteria were permanent employment and age 18–60 years. Questionnaires were sent to a cohort of 3115 employees in one corporation in September 2004. The proposed study

design, implications of the trial and alternative options were explained in the cover letter. The letter also emphasised that taking part in the trial was voluntary and that employees would get the best treatment available and the full attention of the occupational doctor even if they did not want to participate. Those invited were told that they were free to withdraw from the trial at any point, and that this would not prejudice their treatment. At most two reminders were sent. The respondents signed an informed consent. Of the target group, 49% were employed in the field of construction industry: civil engineering, building contracting, technical building services and building materials industry. 51% were employed in installing, repairing, service and maintenance of buildings, industrial installations or communications networks.

Self-reported health problems

The self-administered questionnaire contained items about lifestyle, anthropometrics, sleep disturbances, work-related stress and fatigue, depression, pain, disability due to musculoskeletal problems and a prediction of future working ability. It included previously validated items^{7–14} (table 1).

The responses were interpreted on the basis of a priori defined cut-off limits. Subjects who reported problems with future working ability, pain, impairment due to musculoskeletal problems, insomnia or insufficient sleep, frequent stress or fatigue, or had a high depression score, were rated as having health problems (table 2). Furthermore, the presence of health problems were eventually classified as "none", "one" or "two

Abbreviations: AIC, Akaike information criterion; ZINB, zero-inflated negative binomial

Table 1 Questionnaire topics

Topic	Questions
Body anthropometrics	Height and weight, calculation of body mass index
Physical activity	Exercise, way to work, leisure-time activities. Modified from Laatikainen <i>et al</i> ⁷
Alcohol consumption	Frequency and dosage. Modified from Simpura <i>et al</i> ⁸
Smoking	Yes/no
Pain	Frequency and intensity
Impairment due to musculoskeletal problems at work and leisure time	Semi-continuous visual analogue scale (0–10) ⁹
Depression	Depression score, DEPS scale 0–30 ¹⁰
Stress and fatigue	Work-related stress and fatigue. ^{9,11}
Sleep disturbances	Modification of the Basic Nordic Sleep Questionnaire ¹²
Daytime sleepiness	Epworth Sleepiness Scale 0–24 ¹³
Future working ability	Self-rated ability to continue working in the present job due to health problems after two years ¹⁴

or more" in order to take into account coexisting health problems in each participant.

Sickness absence from work

Sickness absence data were obtained from the employer's records, covering a one-year period from 1 October 2003 to 30 September 2004 (although without medical diagnoses). Data privacy was strictly followed. Records were checked for inconsistencies. Overlapping and consecutive spells of sickness absence were combined. The employer records the sick leave periods, including the dates when each spell started and ended. In the company involved in our study, permanent employees are paid a full salary during their sick leave from the first day. The blue-collar employees cannot complete their own certificates for any sick leave. White-collar employees must provide a written explanation for short sick leaves and a medical certification for sick leaves longer than three days.

Maternity/paternity leave and absence from work to care for a sick child are not included in the sickness absences.

We also received the sickness absence records of the non-respondents in an anonymous manner, which made it possible

Table 2 "Health problems": findings in one or more of these topics. Percentages have been calculated within the group

Topic	Criteria	n (%)
Severe physical impairment at work (0–10)	≥5	270 (64)
Severe pain	At least "moderate" pain that "affects working ability" at minimum three times a week	81 (18)
Self-rated future working ability:	Uncertain of own ability ("Uncertain"), or quite sure ("Not able") not being able	244 (58)
Potential depression (0–30)	DEPS score ≥11	68 (16)
Severe insomnia	Problems in falling asleep or night awakenings AND daytime tiredness daily or almost daily	60 (14)
Work-related fatigue	"Very much" feeling of being squeezed empty because of work	35 (8)
Work-related stress	"Very much" feeling tense, strained, nervous and/or anxious because things are on one's mind all the time	30 (7)

Sum of percentages exceeds 100% because many subjects had more than one abnormal finding.

to compare the respondents and non-respondents as groups regarding sickness absence.

Statistics

Sickness absence was operationalised as the accumulated number of days on sick leave during the one-year study period. When analysing how sickness absence depends on covariates (explanatory variables and prognostic factors), we initially tried four different types of regression models: the simple Poisson regression model, the zero-inflated Poisson model, the simple negative binomial model, and the zero-inflated negative binomial model (ZINB). It turned out that (i) there was great overdispersion in relation to the Poisson model, and (ii) an essential excess of zero absences compared with what could be reasonably expected in the simple non-inflated Poisson and negative binomial models. Therefore, as it was necessary to allow for both of these features, we concentrated on using the ZINB model in subsequent analyses.

The ZINB model^{15,16} starts by postulating that the study population is latently divided into two subsets: A—subjects with a very high propensity to have zero days on sick leave, and B—subjects with substantial probability of at least one absence day. The zero-inflation part of the ZINB model predicts the odds of membership in the "immune" subpopulation A rather than in the "susceptible" subpopulation B. Dependency of these odds on covariates was modelled according to a logistic model, its regression coefficients describing the logarithms of the corresponding odds ratios associated with the covariates. The estimated odds ratios (with 95% CI) will also be presented in tabulated form. For easier interpretation and coherence with the negative binomial part below we switched the outcome to be membership of the susceptible subset B. This reparametrisation of the mathematically equivalent original model implies only a change of sign of the regression coefficients and the inversion of odds ratios from the original zero-inflation model.

It is further postulated that in the immune subpopulation A the probability of zero absence is simply 100%. In contrast to this, in the susceptible subpopulation B the number of days on sick leave is assumed to obey the negative binomial distribution. In this negative binomial part of the ZINB model the mean number of absence days is assigned to be dependent on the relevant covariates according to a log-linear model. Hence, in this part a given regression coefficient represents the natural logarithm of the ratio of mean values of the response variable associated with a unit change in the pertaining covariate. When presenting results, the estimated ratios of means (with 95% CI) are reported. See the Appendix for a more detailed description of the ZINB model.

The parameters of the ZINB model were estimated by maximum likelihood using the function `zeroinfl()` in the package `pscl`¹⁵ attached with the R environment for statistical computing and graphics (<http://www.r-project.org/>). The models were compared using the Akaike information criterion (AIC), and goodness-of-fit was evaluated by comparing the marginal observed frequencies to the expected frequencies, the latter being based on the fitted model in classes of categorised outcome.

RESULTS

We received 1507 responses (48.4%) of which 166 were excluded due to following reasons: inadequately filled questionnaire (n = 29), age-related pension granted (n = 1), part-time or disability pension granted (n = 24), or the subject did not provide consent to analyse sickness absence or pension records (n = 110). Additionally two subjects had missing absence data.

Table 3 The prevalence of self-reported health problems and characteristics of the distribution of the number of days on sick leave by gender, occupational grade and age

Gender	Age (years)	Subjects, n	Health problems (%)			Days on sick leave				
			One	Two or more	% with zero days	Median	Upper quartile	Maximum	Mean of all values	Mean of non-zero values
Male blue-collar	18-39	285	15	14	25	5	13	229	13	17
	40-49	266	18	15	34	3	13	180	11	17
	50-61	278	25	32	35	4	17	221	16	24
Male white-collar	18-39	98	12	4	71	0	2	40	2	8
	40-49	123	6	5	74	0	1	34	1	6
	50-61	124	15	14	73	0	1	197	5	17
Female white-collar	18-39	54	11	2	39	2	6	27	4	6
	40-49	58	12	3	55	0	2	135	6	13
	50-61	46	15	7	50	0	4	11	2	5
All eligible participants		1341	16	15	42	2	9	229	10	17
All respondents*		1366	17	14	44	2	9	347	11	18
Non-respondents†		1714	NA	NA	38	3	10	276	12	20

The data for blue-collar females are not shown due to small number of subjects (n=9), but their data are included in "All eligible participants".

*Includes (in addition to the eligible participants of the study) those who were excluded from the analyses due to granted pension.

†Includes all subjects who did not respond to the survey, or refused to participate in any part of the study. Their sickness absence data were gathered in an anonymous manner.

The final study population thus consisted of 1341 subjects. At the time of the questionnaire survey, the respondents were on average 44 years old (range 19-61 years). Of them 12% were females, and 61% were blue-collar workers.

The distribution of sickness days among non-respondents was very similar to that in respondents (table 3). Non-respondents were on average somewhat younger (mean 40 years) than respondents. Five per cent of non-respondents were females.

A total of 12 837 days of sickness absence were recorded in the study population during the 12 months. The distribution was heavily right-skewed in all age groups. Moreover, 42% had not been on sick leave at all, indicating a substantial zero-component in the response distribution (tables 3 and 4). The proportions of zero-absences were 31%, 73% and 47% in blue-collar males, white-collar males and white-collar females, respectively. The mean numbers of absence days among those with any sickness absence were 19, 11 and 8 days in these three groups, respectively. In blue-collar males and white-collar females the proportions with no sickness absence were lower in young employees than among those at least 40 years of age. An increasing trend of absence days by age was observed among those with any sick leave in the male groups. Thirty one

per cent of subjects reported health problems (table 3). Their share of the total number of days on sick leave was 61%.

Our first regression model, Model 1, included as covariates: the combination of gender and occupational grade (categories: male and blue-collar, male and white-collar, female and white-collar), age (seven groups), and self-reported health complaints (none, 1, ≥ 2). The AICs were 28 963, 20 441, 7124 and 7029, for the simple Poisson, the zero-inflated Poisson, the simple negative binomial and the zero-inflated binomial (ZINB) model, respectively. Based on these figures we chose the ZINB model for the subsequent analyses and presentation of results. The statistical appendix provides instructions on how the estimated model coefficients can be translated into predicted probabilities of susceptibility of sickness absence and of mean numbers of days on sick leave for any combination of prognostic factors. As the baseline odds for susceptibility to any sickness absence was more than 50%, the reported odds ratios are exaggerating the respective relative risks. Hence, we avoid direct quantitative interpretation of these odds ratios.

The results from fitting Model 1 are displayed in table 5. The high odds ratios for being susceptible to any sickness absence in male blue-collar and female white-collar workers, respectively, when compared to male white-collar employees were very consistent with the great contrasts observed in the proportions of workers with any sickness absence between these groups, as implied in table 3. The average number of sickness days among the susceptible to any sickness absence was about twice as high in male blue-collar workers as in male white-collar employees, but female white-collar subjects were not seen to differ from male white-collar employees in this regard. There was some evidence of an overall decreasing trend by age in the susceptibility to sickness absence by increasing age, but not for the average number of days on sick leave. The presence of health problems was associated with both the susceptibility to and the mean number of days on sick leave. Those who reported one health problem had on average almost twice the number of sickness absence days and those with two or more health complaints had both higher propensity for any sickness absence and 3.4 times higher total number of absence days than those who did not report any health problems, when adjusted for gender, occupational group and age (table 5).

In our second ZINB model, Model 2, we included as covariates gender, age, body mass index, alcohol consumption, depression score (DEPS score), stress and fatigue, shortage of

Table 4 The observed counts in 11 classes of the outcome variable and the expected frequencies predicted by the three fitted zero-inflated negative binomial regression (ZINB) models, including their values of the Akaike Information Criterion (AIC)

Days on sick leave	Observed	Model 1 (AIC 7029.4)	Model 2 (AIC 6976.3)	Model 3 (AIC 6961.4)
0	583	586	583	583
1-2	133	159	157	158
3-4	140	99	102	101
5-6	85	72	75	75
7-9	78	80	83	83
10-13	76	74	77	77
14-20	81	82	84	85
21-27	43	50	50	50
28-41	43	55	53	53
42-83	46	53	47	47
84-230	24	22	21	21

Table 5 Predicting the propensity to being susceptible versus immune to any sickness absence (zero-inflation part) and the duration of sickness absence, if susceptible (negative binomial part)

	Zero-inflated part (ZI)		Negative binomial part (NB)	
	Coefficient	OR (95% CI)	Coefficient	MR (95% CI)
Baseline (odds for ZI, mean for NB)	-0.29	0.75 (0.41 to 1.37)	1.60	4.93 (3.45 to 7.04)
Male blue collar	1.99	7.30 (4.72 to 11.30)	0.67	1.95 (1.45 to 2.62)
Male white collar (ref)	0	1	0	1
Female white collar	1.42	4.13 (2.20 to 7.76)	0.02	1.02 (0.68 to 1.54)
Age (years)				
18–29	0.16	1.17 (0.48 to 2.86)	0.05	1.05 (0.72 to 1.52)
30–34	0.21	1.24 (0.52 to 2.93)	-0.04	0.96 (0.65 to 1.40)
35–39 (ref)	0	1	0	1
40–44	-0.22	0.81 (0.40 to 1.63)	-0.15	0.86 (0.62 to 1.20)
45–49	-0.68	0.51 (0.26 to 1.01)	0.01	1.01 (0.71 to 1.43)
50–54	-0.64	0.53 (0.26 to 1.06)	-0.09	0.91 (0.65 to 1.27)
55–61	-0.68	0.51 (0.26 to 1.01)	0.16	1.18 (0.84 to 1.66)
Health problems				
None (ref)	0	1	0	1
One	0.24	1.27 (0.76 to 2.12)	0.62	1.87 (1.44 to 2.42)
Two or more	0.92	2.51 (1.35 to 4.68)	1.23	3.41 (2.64 to 4.40)

Estimated model coefficients, odds ratios (OR) and mean ratios (MR) with 95% CI from fitting a zero-inflated negative binomial regression Model 1 including age, gender, occupational grade and the presence of self-reported health problems as covariates.

The estimate of the dispersion parameter was $\phi = 0.56$.

“Zero-inflated part” refers to the model component for predicting membership to the subpopulation A with high propensity to zero absence, and “Negative binomial part” to the component predicting the days on sick leave among the susceptible subpopulation B. To facilitate interpretation, for the zero-inflation part we have shown the odds ratios associated with the complementary propensity to having any sickness absence—that is, inclusion in subpopulation B.

sleep (in hours), daytime alertness (ESS score), pain, impairment due to musculoskeletal problems at work (scale 0–10), and self-predicted future work ability (categories: able to work, uncertain, unable to work). The goodness-of-fit improved from Model 1 (table 4). However, apart from age, occupational grade and gender, only musculoskeletal problems, insufficient sleep and predicted future work ability appeared to have any major effect on the outcome (data not shown). As it also became apparent that the independent effect of age was essentially similar within the broad age classes 19–39 years and 45–61 years, respectively, we pooled the age factor into three levels only.

We then fitted a third model, Model 3, with these covariates: combination of gender and occupational grade, age, musculoskeletal impairment at work, insufficient sleep and predicted work ability. The AIC was clearly smaller than in the previous models, and the expected counts were very similar to those of Model 2 (table 4). The results on age, gender and occupational grade were very similar to those from Model 1 (table 5) apart from some changes in the mean ratios across the subgroups defined by gender and occupational grade. In this model both the self-predicted future working ability and the score for musculoskeletal impairment were strong predictors for the number of sickness absence days (table 6). Among the susceptible, the estimated mean number of absence days increased by 14% for each rise of 1 unit of the impairment score. Those susceptible to any sickness absence and whose prediction of their future working ability was “uncertain” or “not able” had twice or three times as high mean number of days on sick leave, respectively, when compared to those whose own prediction on working ability was positive. In addition, insufficient sleep predicted a somewhat increased propensity for any sickness absence, but not the total number of absence days.

DISCUSSION

Main findings

The prevalence of health problems increased with age, and blue-collar workers had far more sickness absence days than

white-collar employees. When self-reported health problems and occupational grade were accounted for, age was not associated with the total number of absence days, and older workers were less likely to stay out of work than the young. Self-reported health problems predicted sickness absence in a dose-related manner. Of the individual items of self-reported health problems, self-rating of future working ability and impairment due to musculoskeletal problems showed strongest associations with sickness absence.

Strengths and weaknesses of the study

Sickness absences serve as a measure of health in the working population when health is understood as a mixture of social, psychological and physiological functioning.^{17, 18} Recorded sickness absence data have several advantages: the quality of the data in terms of coverage, accuracy and consistency over time is superior to that achievable via self-reports.¹⁹ However, their analysis is difficult with traditional statistical methods because a substantial fraction is clustered at value zero, and this proportion is greater than predicted by any basic probability model for count data. Also, the residual variability in the non-zero part of the distribution exceeds that predicted by a Poisson model for counts. For these reasons we chose the zero-inflated negative binomial (ZINB) regression model^{15, 16} as our analysis tool, which provided a reasonably acceptable fit. Although it was perhaps not able to deal with all the complexity associated with this type of response variable, among computationally feasible approaches it is clearly more appropriate than the common simpler alternative models in dealing with both the extra-zero component and the overdispersion. However, the observed counts in response classes 1–2 and 21–42 absence days were systematically lower than the expected counts predicted by the ZINB models, whereas in classes 3–6 absence days the situation was vice versa (table 4). This pattern suggests that the fit of the ZINB model was not as good as desired, although it was the best of the realistically available models. The relative peak at 3–6 days could be interpreted that the outcome distribution may in reality have more than two

Table 6 Predicting the propensity to being susceptible versus immune to any sickness absence (zero-inflation part) and the duration of sickness absence, if susceptible (negative binomial part)

	Zero-inflated part (ZI)		Negative binomial part (NB)	
	Coefficient	OR (95% CI)	Coefficient	MR (95% CI)
Baseline (odds for ZI, mean for NB)	-0.54	0.58 (0.37 to 0.91)	1.73	5.67 (4.12 to 7.80)
Male blue-collar	2.02	7.53 (4.76 to 11.90)	0.34	1.40 (1.04 to 1.88)
Male white-collar (reference)	0	1	0	1
Female white-collar	1.44	4.24 (2.24 to 8.00)	-0.26	0.77 (0.52 to 1.15)
Age (years)				
18-39 (reference)	0	1	0	1
40-44	-0.37	0.69 (0.40 to 1.21)	-0.11	0.90 (0.68 to 1.17)
45-61	-0.72	0.49 (0.31 to 0.75)	-0.12	0.89 (0.72 to 1.10)
Musculoskeletal impairment due to work (per 1 unit; scale 0-10)	0.13	1.13 (1.02 to 1.26)	0.13	1.14 (1.09 to 1.19)
Insufficient sleep (per hour)	0.32	1.38 (1.14 to 1.66)	-0.09	0.92 (0.86 to 0.98)
Predicted work ability				
"Able"	0	1	0	1
"Uncertain"	0.11	1.12 (0.63 to 1.98)	0.69	2.00 (1.53 to 2.62)
"Not able"	0.63	1.87 (0.50 to 6.99)	1.13	3.09 (1.89 to 5.05)

Estimated model coefficients, odds ratios (OR) and mean ratios (MR) with 95% CI from fitting a zero-inflated negative binomial regression Model 3 including age, occupational grade, gender, musculoskeletal impairment, insufficient sleep and self-rated future working ability as covariates.

The estimate of the dispersion parameter was $\varphi = 0.62$.

"Zero-inflated part" refers to the model component for predicting membership to the subpopulation A with high propensity to zero absence, and "negative binomial part" to the component predicting the days on sick leave among the susceptible subpopulation B. To facilitate interpretation, for the zero-inflation part we have provided the ORs associated with the complementary propensity to having any sickness absence—that is, inclusion in subpopulation B.

components: the excess zero part, a component centred around small values (3-6) of absence days, and a third component centred around a relatively high mean level, perhaps more than 84 days. It is difficult to evaluate what the quantitative implications are of this observed deficiency of our model to the validity and precision of the estimates based on it. One likely consequence is, however, that the confidence intervals reported here underestimate to some extent the true uncertainty associated with our estimation.

A "healthy worker effect" might be present if employees with worse health level (long-term absence and disability states) had not responded. This potential bias would underestimate the associations as the respondents would be healthier, and possibly have had less sickness absence than non-respondents. The participation rate was in line with other studies in occupational populations in many countries.²⁰ In our study, the non-respondents were slightly younger than respondents. When comparing the distribution of absence days between respondents and non-respondents, there was no relevant difference in mean absence. Therefore we think that the study population is reasonably representative of the original target population in this respect.

As our study is based on cross-sectional data, there is a possibility of reverse causality. That is, sickness absence due to any reason could potentially modify the reporting of health problems. Although this may partly explain the results, especially because those on sick leave at the time of responding to the survey were also included, we believe that experienced health problems determine sickness absence, and not vice versa.

Some differences in comparison to previous studies

Besides age, gender and occupational grade, the assessment of future working ability and the score for musculoskeletal impairment were strong determinants of sickness absence, in line with our hypothesis and previous studies.²¹⁻²² Contrary to our expectations and earlier findings,²³⁻²⁵ the prevalence of depression, fatigue or stress was fairly low and was not significantly associated with sickness absence in this cohort.

Although greater decision authority predicts low sickness absence,²⁶⁻²⁷ it may increase the risk of psychological distress and fatigue,²⁸⁻²⁹ especially if the employees are exposed to high job demands. Our cohort mainly included blue-collar workers with low decision authority concerning *which* job tasks to perform, but good job-related autonomy concerning *how* to perform the task. This may partly explain our results that the prevalence of psychological distress or fatigue was low (table 2) and not associated with sickness absence, and that the most frequently reported health problem was physical impairment from musculoskeletal problems. Neither alcohol consumption nor smoking explained the associations of self-reported health problems or age to sickness absence.

Many previous studies have reported that females have more sickness absence than males, but this was not the case in our study. Female white-collar workers had higher propensity for any sickness absence, if susceptible, but similar numbers of absence days as their male counterparts.

Meaning of the study

Construction workers are apparently at a greater risk of developing certain health disorders and sickness absence than workers in many other industries.³⁰⁻³¹ Physically demanding job tasks and occupational injuries are likely determinants for the high prevalence. Subjects exposed to challenging tasks more likely report underlying health problems than subjects in sedentary tasks. However, this does not explain the inverse association between age and propensity to sickness absence.

The "healthy worker survivor effect" describes a continuing selection process: those who remain employed in a specific profession tend to be healthier than those who leave employment. This phenomenon is particularly true in the construction industry³² as well as in other physically demanding jobs. Maybe this partly explains the inverse association between age and propensity to absence, which was contrary to some previous reports.⁴⁻⁵ However, all employees participating in the present study were paid a full salary during their sick leave from the first day and there was no diversity in this respect due to age. We think that there may also be psychosocial and behavioural

Main messages

- Higher total counts of absence days were found among subjects reporting certain health problems and weakened working ability, regardless of age, sex and occupational grade.
- One third of the subjects reported named health problems, but their share of the total number of days on sick leave was over 60%.
- When self-reported health problems, gender and occupational grade were accounted for, age was not associated with the total number of absence days, and older workers were less likely to stay out of work than younger employees.
- A zero-inflated negative binomial regression model provided a reasonably acceptable fit to sickness absence data characterised by skewness, overdispersion and heavy clumping at zero value.

Policy implications

- It is possible to identify individuals at a high risk of sickness absence with a simple health questionnaire among employees predominantly engaged in physical work.
- Irrespective of age, the healthcare system needs to pay more attention to the health problems and working ability experienced by employees, as these are strongly related to sickness absence.
- Psychosocial and behavioural differences between younger and older workers should be taken into account in the prevention of work absence among the young.

differences between the younger and older workers: perhaps their attitudes and values towards work are different. This may have implications for the prevention of work absence among young construction workers. In addition, irrespective of age, the healthcare system needs to address health and working ability, which are strongly related to sickness absence.

Unanswered questions and future research

It remains to be seen whether similar associations between age, self-reported health problems and sickness absence exist also in, for example, knowledge-intensive sedentary occupations. The order of the causality—that is, that age and self-reported health problems determine sickness absence—must also be confirmed in prospective studies. Further research is needed to find out the medical, psychosocial and behavioural determinants of sickness absence in the young.

Authors' affiliations

S Taimela, J Tiekso, S Justén, Evalua International, Vantaa, Finland
E Läärä, University of Oulu, Department of Mathematical Sciences, Oulu, Finland
H Sintonen, University of Helsinki, Department of Public Health, Helsinki, Finland
A Malmivaara, Finnish Office for Health Technology Assessment, FinOHTA/Stakes, Helsinki, Finland
T Aro, Mutual Pension Insurance Company Ilmarinen, Helsinki, Finland

Funding: Finnish Funding Agency for Technology and Innovation (TEKES); The Finnish National Fund for Research and Development (SITRA); Pfizer Oy. The authors' work was independent of the funders.

Competing interests: ST and JT are shareholders of and SJ employed by Evalua International. EL, AM, HS and TA have no competing interests to declare.

Ethics: The Helsinki University Research Ethics Board for the Occupational Health reviewed the study plan and gave their approval in advance. Record number (Dnro): 28/E2/04 (23 April 2004). All subjects received written information regarding the study according to the principles of the Declaration of Helsinki. Only subjects who gave their signed informed consent were included in the study. The consent letters are stored with other study material.

REFERENCES

- 1 **Briner RB**. ABC of work related disorders. Absence from work. *BMJ* 1996;**313**:874–7.
- 2 **Reiso H, Nygard JF, Brage S, et al**. Work ability and duration of certified sickness absence. *Scand J Public Health* 2001;**29**:218–25.
- 3 **Virtanen M, Kivimäki M, Elovainio M, et al**. From insecure to secure employment: changes in work, health, health related behaviours, and sickness absence. *Occup Environ Med* 2003;**60**:948–53.
- 4 **Brenner H, Ahern W**. Sickness absence and early retirement on health grounds in the construction industry in Ireland. *Occup Environ Med* 2000;**57**:615–20.
- 5 **de Zwart BC, Frings-Dresen MH, van Duivenbooden JC**. Senior workers in the Dutch construction industry: a search for age-related work and health issues. *Exp Aging Res* 1999;**25**:385–91.
- 6 **Vahtera J, Kivimäki M, Pentti J**. The role of extended weekends in sickness absenteeism. *Occup Environ Med* 2001;**58**:818–22.
- 7 **Laatikainen T, Tapanainen H, Alftan G, et al**. FINRISKI 2002: Implementation of the Study and Results 1. (Tutkimuksen toteutus ja tulokset 1. Perusraportti). Helsinki: National Public Health Institute, Finland, 2003.
- 8 **Simpura J**. Development of common instrument for alcohol consumption. In: Nasikov A, Guder C, eds. *Eurohis*. WHO Regional Office for Europe: IOS Press, 2003.
- 9 **National Public Health Institute**. *Implementation and methods of the Health 2000 Survey*. (Menetelmäraportti. Terveys 2000-tutkimuksen toteutus, aineisto ja menetelmät). Helsinki: National Public Health Institute, 2005.
- 10 **Salokangas RK, Poutanen O, Stengard E**. Screening for depression in primary care. Development and validation of the Depression Scale, a screening instrument for depression. *Acta Psychiatr Scand* 1995;**92**:10–16.
- 11 **Kauppinen T, Hanhela R, Heikkilä P, et al**. *Work and health in Finland in 2003 (Työ ja terveys Suomessa 2003)*. Helsinki: Finnish Institute of Occupational Health, 2003.
- 12 **Partinen M, Gislason T**. Basic Nordic Sleep Questionnaire (BNSQ): a quantitated measure of subjective sleep complaints. *J Sleep Res* 1995;**4**:150–5.
- 13 **Johns MW**. A new method for measuring daytime sleepiness: the Epworth sleepiness scale. *Sleep* 1991;**14**:540–5.
- 14 **Tuomi K, Ilmarinen J, Jahkola A, et al**. *Work Ability Index (Työkykyindeksi)*. Helsinki: Finnish Institute of Occupational Health, 1992.
- 15 **Cheung YB**. Zero-inflated models for regression analysis of count data: a study of growth and development. *Stat Med* 2002;**21**:1461–9.
- 16 **Moon S, Shin J**. Health care utilization among Medicare-Medicaid dual eligibles: a count data analysis. *BMC Public Health* 2006;**6**:88.
- 17 **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.
- 18 **Kivimäki M, Head J, Ferrie JE, et al**. Sickness absence as a global measure of health: evidence from mortality in the Whitehall II prospective cohort study. *BMJ* 2003;**327**:364.
- 19 **Ferrie JE, Kivimäki M, Head J, et al**. A comparison of self-reported sickness absence with absences recorded in employers' registers: evidence from the Whitehall II study. *Occup Environ Med* 2005;**62**:74–9.
- 20 **Gimeno D, Benavides FG, Amick BC III, et al**. Psychosocial factors and work related sickness absence among permanent and non-permanent employees. *J Epidemiol Community Health* 2004;**58**:870–6.
- 21 **Nakata A, Haratani T, Takahashi M, et al**. Association of sickness absence with poor sleep and depressive symptoms in shift workers. *Chronobiol Int* 2004;**21**:899–912.
- 22 **Kaaria S, Kaila-Kangas L, Kirjonen J, et al**. Low back pain, work absenteeism, chronic back disorders, and clinical findings in the low back as predictors of hospitalization due to low back disorders: a 28-year follow-up of industrial employees. *Spine* 2005;**30**:1211–18.
- 23 **Michie S, Williams S**. Reducing work related psychological ill health and sickness absence: a systematic literature review. *Occup Environ Med* 2003;**60**:3–9.
- 24 **Marshall NL, Barnett RC, Sayer A**. The changing workforce, job stress, and psychological distress. *J Occup Health Psychol* 1997;**2**:99–107.
- 25 **Zavala SK, French MT, Zarkin GA, et al**. Decision latitude and workload demand: implications for full and partial absenteeism. *J Public Health Policy* 2002;**23**:344–61.

- 26 Christensen KB, Nielsen ML, Rugulies R, *et al.* Workplace levels of psychosocial factors as prospective predictors of registered sickness absence. *J Occup Environ Med* 2005;**47**:933–40.
- 27 Labriola M, Lund T, Burr H. Prospective study of physical and psychosocial risk factors for sickness absence. *Occup Med (Lond)* 2006.
- 28 Marchand A, Demers A, Durand P. Do occupation and work conditions really matter? A longitudinal analysis of psychological distress experiences among Canadian workers. *Social Health Illn* 2005;**27**:602–27.
- 29 Bulmann U, Huibers MJ, van Amelsvoort LP, *et al.* Psychological distress, fatigue and long-term sickness absence: prospective results from the Maastricht Cohort Study. *J Occup Environ Med* 2005;**47**:941–7.
- 30 Snashall D. Safety and health in the construction industry. *BMJ* 1990;**301**:563–4.
- 31 Burkhart G, Schulte PA, Robinson C, *et al.* Job tasks, potential exposures, and health risks of laborers employed in the construction industry. *Am J Ind Med* 1993;**24**:413–25.
- 32 Siebert U, Rothenbacher D, Daniel U, *et al.* Demonstration of the healthy worker survivor effect in a cohort of workers in the construction industry. *Occup Environ Med* 2001;**58**:774–9.
- 33 Kauermann G, Ortlieb R. Temporal pattern in number of staff on sick leave: the effect of downsizing. *Applied Statistics* 2004;**53**:355–67.
- 34 Wang K, Yau KK, Lee AH. A hierarchical Poisson mixture regression model to analyse maternity length of hospital stay. *Stat Med* 2002;**21**:3639–54.

STATISTICAL APPENDIX

We provide here a detailed technical description of the zero-inflated negative binomial (ZINB) model (see also^{15 16}). The outcome or response variable is denoted by Y = number of days on sick leave during the one-year observation period, and it can obtain non-negative integer values. The ZINB model is a mixture of (a) the zero-inflation (ZI) part, and (b) the negative binomial (NB) part.

THE ZERO-INFLATION PART

We postulate that the study population is latently divided into two subsets:

A = subjects with a very high propensity to have zero days on sick leave

B = subjects with substantial probability of at least one absence day.

Let p_B be the probability that an individual is susceptible—that is, he/she belongs to subset B, and $p_A = 1 - p_B$ is the probability of being immune—that is, the subject belongs to subset A. It is assumed that these probabilities depend on the individual values of the model terms X_1, X_2, \dots, X_m that are appropriately constructed from the relevant explanatory variables or covariates, according to the common *logistic* regression model:

$$\text{logit}(p_A/p_B) = a_0 + a_1X_1 + \dots + a_mX_m,$$

in which “log” stands for the natural logarithm function. Each coefficient a_j ($j = 1, \dots, m$) is interpreted as the change of the log-odds of the subject belonging to subset A rather than to subset B corresponding to a unit change in the value of covariate term X_j when all the other covariates are kept unchanged. Thus, $\text{OR}_j = \exp(a_j)$, the antilog of a_j , is the odds ratio describing the effect of a unit change in X_j on the chances of being immune rather than susceptible, adjusted for the other covariates.

Equally, this logistic model can be specified in terms of contrasting the odds for B versus A, in which case the regression coefficients will only have their signs changed, and the odds ratios will be inverted. In fact, when presenting our results (tables 5 and 6), we chose to display the relative odds in this way to describe the covariate effects on the probability of being susceptible rather than immune.

If the subject belongs to the immune subset A, the distribution of the response is assumed to be degenerate such that the probability of zero days on sick leave is 1.

THE NEGATIVE BINOMIAL PART

When a subject belongs to the susceptible subset B, the response variable Y may get either a zero or any positive integer

value. Let q_y be the *conditional* probability of being exactly y days on sick leave, given membership in this subpopulation. This probability is assumed to come from the negative binomial (NB) distribution, obeying the following formula for any $y = 0, 1, 2, \dots$

$$q_y = (\varphi\mu)^y (1 + \varphi\mu)^{-y-1/\varphi} \Gamma(y+1/\varphi) [\Gamma(y+1) \Gamma(1/\varphi)]^{-1}$$

in which μ is the expected value or theoretical mean and $\varphi > 0$ is the dispersion parameter of the NB distribution, and $\Gamma(u)$ refers to the gamma function evaluated at real number value u . Actually after some manipulation this probability can also be expressed in a simplified form as

$$q_y = \mu^y \exp(-\mu)/y! \times R(y, \mu, \varphi)$$

which is a product of the simple and familiar Poisson probability formula and the more complicated function $R(y, \mu, \varphi)$ describing the relative deviation of the NB distribution from the Poisson one at each value of y . The NB variance is $\mu(1 + \varphi\mu)$, being obviously greater than μ which is the Poisson variance.

In the NB part of the ZINB model the mean number of days on sick leave in the susceptible is postulated to depend on the covariates according to a *log-linear* structure:

$$\log(\mu) = b_0 + b_1X_1 + \dots + b_mX_m.$$

Here a regression coefficient b_j refers to the change in the logarithm of the expected value μ per unit change in covariate term X_j keeping the other covariates constant. Accordingly, $\text{MR}_j = \exp(b_j)$ is the ratio of mean responses—that is, the multiplicative effect of a unit change in covariate X_j on the expected response among the susceptible and adjusted for the other covariates.

Note that we have the same set of covariate terms X_1, X_2, \dots, X_m to predict both the probability p_A of being immune and the mean response among the susceptible. However, it may well be that certain covariates X_j have no effect on predicting p_A , in which case the parameters a_j associated with these covariates in the ZI part are zero-valued, whereas in the NB component some other covariate terms X_k may have no effect on the mean response μ in the susceptible.

ZINB MODEL: MIXTURE OF THE TWO PARTS

Finally, the total or *marginal* probability Q_y for a subject being exactly y days on sick leave during the one-year period is combined from the above probabilities as follows:

$$Q_0 = p_A + p_Bq_0 = \text{probability of zero days,}$$

$$Q_y = p_Bq_y = \text{probability of } y \text{ days for } y = 1, 2, \dots$$

Hence the marginal probability distribution is a *mixture* of the degenerate distribution (concentrated at zero) pertinent to the immune subjects and the NB distribution, which is presumed to hold for the susceptible individuals, such that the mixing proportions are p_A and p_B , respectively. The marginal expected value $E(Y)$ of the response is a weighted average of the conditional means, which simplifies into $E(Y) = p_B\mu$. The variance of this mixture distribution is $\text{var}(Y) = p_B\mu [1 + (p_A + \varphi)\mu]$.

This general specification of the ZINB model contains the following special cases: the zero-inflated Poisson (ZIP) model is obtained when the dispersion parameter φ is put to approach 0. On the other hand, keeping φ positive but putting $p_A = 0$, we get the non-inflated NB model. When both $\varphi \rightarrow 0$ and $p_A = 0$, the model reduces to the simple Poisson model.

The likelihood function is created straightforwardly from the definitions of the probabilities Q_y expressed as functions of the $2m+2$ regression coefficients $a_0, a_1, \dots, a_m, b_0, b_1, \dots, b_m$, and the dispersion parameter φ (see Moon and Shim¹⁶). Estimation of the parameters and assessment of their precision (by standard

errors and confidence intervals) applying the principle of maximum likelihood can be computationally effected in some statistical programmes such as R, Stata, Limdep and S-Plus.

PREDICTING SICKNESS ABSENCE BY THE MODEL

We illustrate how the fitted model can be used for individual predictions on sickness absence days given any covariate profile. From the results of Model 3 reported in table 6 we find the following:

Case 1

Male, white-collar, age 30 years, no musculoskeletal impairment, sufficient sleep, and self-predicted working ability rated "able". The baseline log-odds of -0.54 converts to baseline odds of $\exp(-0.54) = 0.58$ and estimated probability $0.58 / (1+0.58) = 37\%$ of being susceptible to sick leave. Given susceptibility, the conditional baseline mean number of days on sick leave is 5.67. Hence, the marginal expected value for this type of worker is $0.37 \times 5.67 = 2.1$ days.

Case 2

Male, blue-collar, age 50 years, musculoskeletal impairment score 7, insufficiency of sleep 2 h/night, predicted working ability "not able". The log-odds for belonging to subset B is computed as

$$-0.54 + 2.02 - 0.72 + 7 \times 0.13 + 2 \times 0.32 + 0.63 = 2.94$$

from which the estimated probability of susceptibility is $\exp(2.94) / [1 + \exp(2.94)] = 95\%$. The mean number of sickness days for susceptible workers like him is obtained as

$$\exp(1.73 + 0.34 - 0.12 + 7 \times 0.13 + 2 \times (-0.09) + 1.13) = \exp(3.81) = 45.2 \text{ days}$$

from which the marginal expected value is $0.95 \times 45.2 = 42.9$ days.

Case 3

Female, white-collar, age 42 years, musculoskeletal impairment score 3, insufficiency of sleep 1 h/night, predicted working ability "uncertain". The log-odds for belonging to subset B is

$$-0.54 + 1.44 - 0.37 + 3 \times 0.13 + 1 \times 0.32 + 0.11 = 1.35$$

from which the probability of susceptibility is estimated as $\exp(1.35) / [1 + \exp(1.35)] = 79\%$. The mean number of sickness days for susceptible workers like her is obtained as

$$\exp(1.73 - 0.26 - 0.11 + 3 \times 0.13 + 1 \times (-0.09) + 0.69) = \exp(2.35) = 10.5 \text{ days}$$

from which the marginal expected value is

$$0.79 \times 10.5 = 8.3 \text{ days.}$$

ADEQUACY OF THE ZINB MODEL

In our application, the ZINB model proved to be a more suitable approach to analyse sickness absence data as compared with some popular but simpler models for discrete counts. It was certainly more appropriate than common procedures for continuous outcome variables, like normal-theory linear modelling or non-parametric testing. However, assuming complete "immunity" is obviously an oversimplification of having very low propensity of being on sick leave. On the other hand, inspection of observed and expected frequencies (table 4) suggested that the probability distribution of the response variable may actually be composed of three components: one with nearly zero mean, another with low mean, and a third with high mean value for the number of days on sick leave. Applications of finite mixture models with a low and a high mean component in analogous contexts have been reported,^{33 34} but these were based on the simpler Poisson distribution for the separate components. Fitting complicated mixture models would also require tailoring of special computing solution. In our case, it is difficult to say how essential the impact was of the shortcomings in our model specification. One likely consequence is that the reported standard errors and confidence intervals are apparently, to some extent, underestimating the uncertainty associated with the estimation on the interesting quantities. Nevertheless, we believe that allowance of a nearly zero mean and a high mean component in the ZINB model enabled us to capture essential features of the response distribution in order to obtain reasonably realistic estimates of the effects of relevant covariates and adequate predictions on the overall mean levels and variability of the number of days on sick leave.

Let us assist you in teaching the next generation

Figures from all articles on our website can be downloaded as a PowerPoint slide. This feature is ideal for teaching and saves you valuable time. Just click on the image you need and choose the "PowerPoint Slide for Teaching" option. Save the slide to your hard drive and it is ready to go. This innovative function is an important aid to any clinician, and is completely free to subscribers. (Usual copyright conditions apply.)