The Stratifying Role of Job level for Sickness Absence and the Moderating Role of Gender and Occupational Gender Composition

by KRÖGER, Hannes

Online Supplementary material

Contents

S1 – Additional description of Sample and Variables

Figure A – Proportion of women of largest occupations in the sample by groups of gender composition

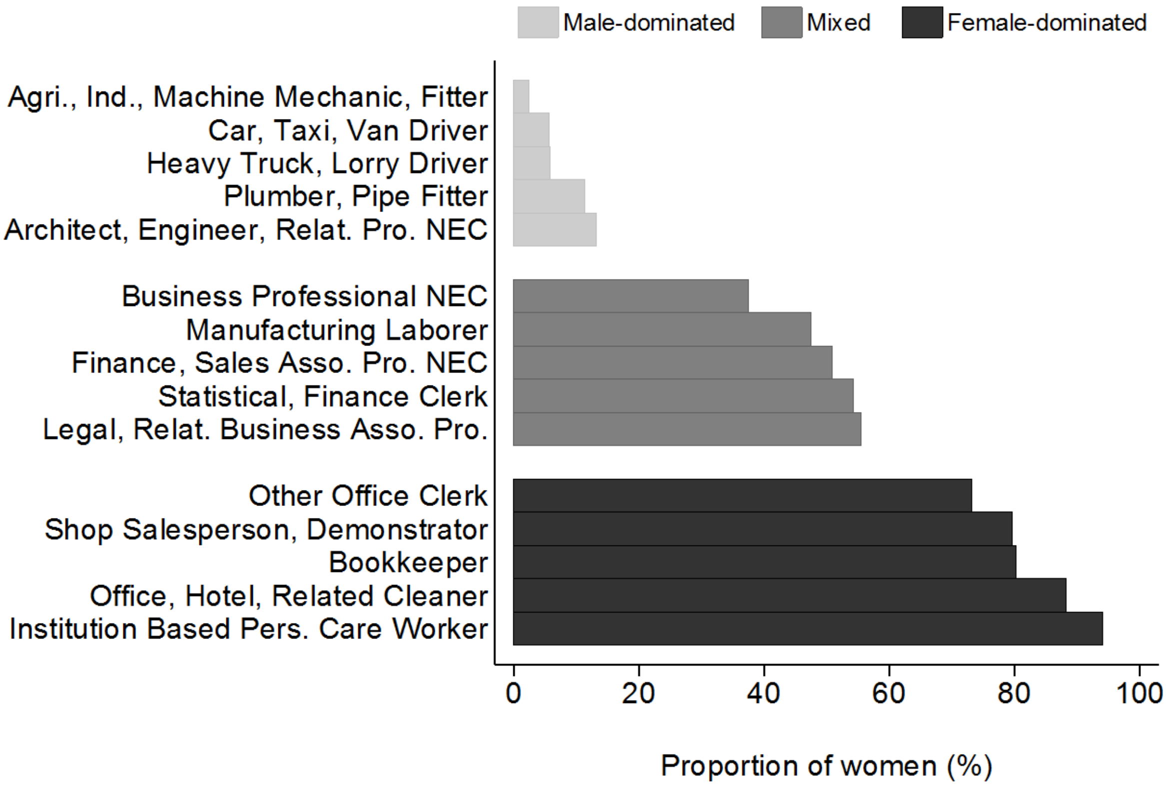


Table A – Definition of job level as a measure of social stratification

Job level hierarchy	Position
1	• Unskilled and semi-skilled workers (blue and white collar)
2	• Skilled blue-collar workers • White collar workers with basic duties
3	• White collar workers with qualified duties • (Assistant) foremen (blue collar)
4	• White collar workers with highly qualified and supervisory duties • Master craftsmen • (Senior) foremen (blue collar)
5	• Senior managers

Note: For the analyses category 5 and 4 are collapsed due to small number of observations in category 5

Table B – Description of variables used in the analyses

Variable	Description
Sickness absence	Self-reported days of sickness absence in the last year
Job level	Degree of authority and autonomy in the job, 4 categories: lowest; low; middle; higher
Education	Secondary and tertiary education (ref. primary) according to CASMIN
Income	Logarithmized monthly wage
Occupational class (EGP)	• higher service class (reference) • lower service class • non-manual routine • routine sales • skilled blue-collar • unskilled blue-collar
Age	Age in years
Age^2	Age in years squared
Physical component score (PCS)	Standardized value of the physical component of the SF-12 health questionnaire
Mental component score (MCS)	Standardized value of the mental component of the SF-12 health questionnaire

Table C - Summary statistics of the sample

	Analytic Sample		Excluded due to missing values	
	Mean	SD	Mean	SD
Age	42.1675	9.9710	40.1599	11.1335
Men	0.5601	0.4964	0.5166	0.4997
Days of sickness absence	6.5989	19.4476	7.9774	24.1431
Lowest job level	0.1672	0.3731	0.2193	0.4138
Lower job level	0.3097	0.4624	0.3264	0.4689
Medium job level	0.2944	0.4558	0.2686	0.4432
Higher job level	0.2288	0.4201	0.1857	0.3889
Women in male-dominated occupations	0.0541	0.2263	0.0529	0.2239
Women in mixed occupations	0.1290	0.3352	0.1374	0.3443
Women in female-dominated occupations	0.2568	0.4369	0.2945	0.4559
Men in male-dominated occupations	0.3981	0.4895	0.3568	0.4791
Men in mixed	0.1246	0.3303	0.1147	0.3186

occupations				
Men in female-dominated occupations	0.0374	0.1897	0.0436	0.2043
Low education	0.3348	0.4719	0.3437	0.4750
Middle education	0.4656	0.4988	0.4802	0.4996
High education	0.1996	0.3997	0.1761	0.3809
EGP: higher service class	0.1336	0.3402	0.1098	0.3126
EGP: lower service class	0.2173	0.4124	0.1908	0.3930
EGP: non-manual routine	0.0872	0.2821	0.0850	0.2789
EGP: routine sales	0.1579	0.3646	0.1609	0.3675
EGP: skilled blue-collar	0.1893	0.3918	0.1811	0.3851
EGP: unskilled blue-collar	0.2148	0.4107	0.2724	0.4452
Log. wage per month	7.5714	0.8010	7.3119	0.9504
Physical component score (PCS)	52.4645	7.8513	52.0364	8.5257
Mental component score (MCS)	50.2022	9.1334	49.5857	9.6514
Observations	87 573		12145	

Individuals	16 384	10922

S 2 -Bayesian Evaluation of Informative Hypotheses (BEIH)

The method is based on a Bayesian approach to statistical modeling and therefore deviates in certain respects from the common frequentist approach (Fennessey, 1977; Gelman et al., 2003). The hypotheses proposed in our study are so called informative hypotheses. Such hypotheses do not merely claim that there is an association between two variables (like sickness absence and job level), but they state that this association is larger for certain groups than for others. Consequently, they propose a ranking of the coefficients from a regression model. We use the *occupational minority hypothesis* ($\boldsymbol{H_{1c}}$) to illustrate the use of inequality constrained hypotheses. There are three times (for each level of job level compared to the references category) six parameters - representing stratification of sickness absence due to level in job level - that we want to estimate (see table 5). The hypothesis orders these parameters in the following way (see table 4):

$$\boldsymbol{H_{1c}}: \beta_{\mathrm{m,f}}, \beta_{w,m} > \beta_{\mathrm{m,mix}}, \beta_{\mathrm{m,m}}, \beta_{w,mix}, \beta_{w,f}$$

A comma indicates no ordering between the respective parameters. The inequality reflects the expectation that the stratification of job level – represented by the regression coefficients – is stronger for men in female-dominated and for women in male-dominated occupations than in each of the other occupational groups. Evaluating the support for such a complex inequality hypothesis is the core of a new method developed in psychology (*Bayesian evaluation of informative hypotheses*, 2008; van de Schoot et al., 2013). Originally, it is based on the Bayesian estimation of the model using Markov Chain Monte Carlo (MCMC) with a Gibbs sampler (Gelman et al., 2003; Lynch, 2007). The general modeling approach we use is called Integrated Nested Laplace Approximation (INLA) (Martins et al., 2013; Rue et al., 2009) implemented as a package for R (www.r-inla.org). It is an approach for latent Gaussian models which allows much flexibility in the specification of the model. Compared to more common Markov Chain Monte Carlo (MCMC) methods, it is very fast to compute.

To apply the Bayesian evaluation of informative hypotheses, we first have to specify the hypotheses as done in the previous section. In addition, we need to calculate how often we would expect support for a particular hypothesis if the coefficients were ordered randomly. This is necessary because the more specific a hypothesis is, the less support we will find just due to randomness in the data. For example, $\boldsymbol{H_2}$ makes no prediction about the ordering of the coefficients. It accepts all possible inequalities between the coefficients. Consequently, it will always find perfect support from the data, because no possible pattern in the data could contradict a statement that predicts no ordering at all. $\boldsymbol{H_{1c}}$, the occupational minority hypothesis, on the other hand, is much more specific than $\boldsymbol{H_2}$. It can receive little or a lot of support depending on the model estimates. The goal is to compare the *observed* support for the hypothesis from the estimated posterior distribution of the coefficients with the *expected* support for the hypothesis given random ordering of the coefficients ($\frac{1}{c_t}$). $\frac{1}{c_t}$ is therefore defined as the proportion of the prior distribution (expectation) that is in agreement with the hypothesis t.

After the posterior distribution has been approximated using INLA, it is possible to draw a random sample from this posterior distribution. Using a sample of e.g. 100 000 draws we can now ask the question: How often does e.g. hypothesis $\boldsymbol{H_{1c}}$ hold true in our sample from the posterior distribution? This gives us an idea as to the extent to which the hypothesis is supported by the data. The proportion $\frac{1}{d_t}$ formally denotes this degree of support, where *t* stands for the specific hypothesis. We can then go on and compare this to our alternative hypotheses, including the hypothesis that there is no ordering ($\boldsymbol{H_2}$).

Given the observed support and the expected support for the hypothesis we can calculate the so-called Bayes factor (Klugkist, 2008).

$$BF_{t2} = \frac{\frac{1}{d_t}}{\frac{1}{c_t}} = \frac{c_t}{d_t}$$

This tells us how much support there is for the hypothesis relative to H_2, which makes no predictions about the ordering. If the Bayes factor is 1, the hypothesis has no stronger predictive power than having no expectations at all and does not add anything of value. If it is below 1 it is actually worse in terms of prediction. If the Bayes factor is 3, for example, that means there is three times more support for the occupational minority hypothesis ($\boldsymbol{H_{1c}}$) than for $\boldsymbol{H_2}$.

Another way of displaying the relative explanatory power of the hypotheses is to calculate the posterior model probability (PMP):

$$PMP(\boldsymbol{H_t}) = \frac{BF_{t2}}{\sum_{t=1a}^{2} BF_{t2}}; t \in 1a, 1b, 1c, 1d, 2$$

The PMP states how much support one hypothesis receives compared to the overall support that all hypotheses under investigation receive. Possible values range from 0 to 100%. The higher the value, the stronger the support for the hypothesis in question compared to the competing hypotheses (van de Schoot et al., 2013). In contrast to null hypothesis significance testing (NHST), the approach does not follow a dichotomous decision criterion of rejecting or accepting a hypothesis, thus avoiding a great deal of the criticism directed at the strict adherence to the p<0.05 criterion of statistical significance (Gelman and Stern, 2006).

Bayesian evaluation of informative hypotheses, 2008. . Springer.

Fennessey, J., 1977. Improving inference for social research and social policy: The Bayesian paradigm. Soc. Sci. Res. 6, 309–327. doi:10.1016/0049-089X(77)90014-X

Gelman, A., Carlin, J.B., Stern, H.S., Rubin, D.B., 2003. Bayesian Data Analysis. CRC Press.

Gelman, A., Stern, H., 2006. The Difference Between “Significant” and “Not Significant” is not Itself Statistically Significant. Am. Stat. 60, 328–331.

Klugkist, I., 2008. Encompassing Prior Based Model Selection for Inequality Constrained Analysis of Variance, in: Hoijtink, H., Klugkist, I., Boelen, P.A. (Eds.), Bayesian Evaluation of Informative Hypotheses. Springer, New York, pp. 53–84.

Lynch, S.M., 2007. Introduction to Applied Bayesian Statistics and Estimation for Social Scientists, 2007 editi. ed. Springer, New York.

Martins, T.G., Simpson, D., Lindgren, F., Rue, H., 2013. Bayesian computing with INLA: new features. Comput. Stat. Data Anal. 67, 68–83.

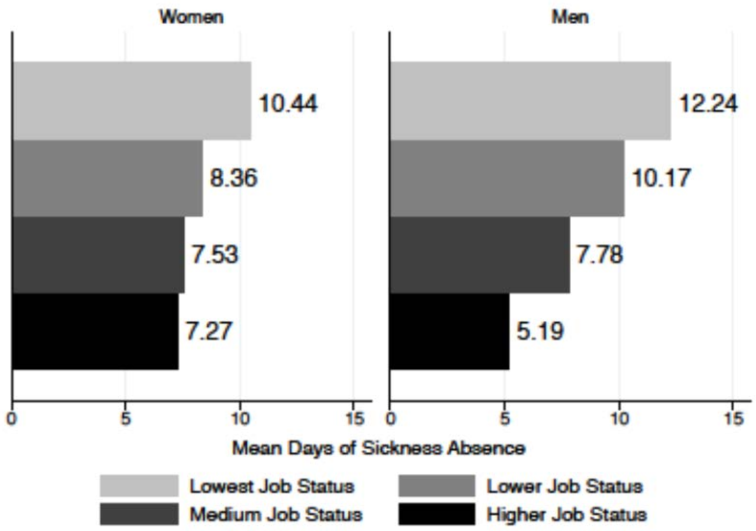
Rue, H., Martino, S., Chopin, N., 2009. Approximate Bayesian inference for latent Gaussian models by using integrated nested Laplace approximations. J. R. Stat. Soc. Ser. B (Statistical Methodol. 71, 319–392. doi:10.1111/j.1467-9868.2008.00700.x

van de Schoot, R., Verhoeven, M., Hoijtink, H., 2013. Bayesian evaluation of informative hypotheses in SEM using Mplus: A black bear story. Eur. J. Dev. Psychol. 10, 81–98.

S3 - Descriptive Evidence

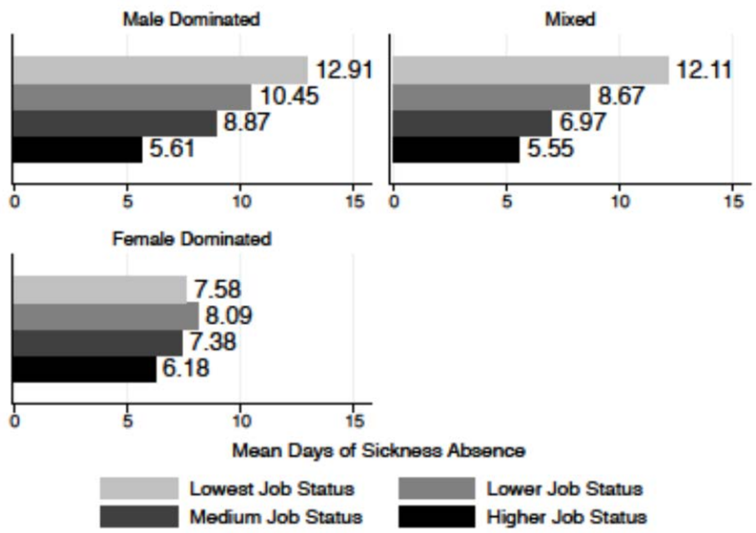
Figure B shows the number of the days of sickness absence by job level for men and women in the sample. We can see that the average number of days of sickness absence ranges between 5 and 12 days and is higher the higher the job level is. In the lowest job level group, men have a higher number of days of sickness absence while in the highest job level group women show higher rates of absence. Overall, the gradient in sickness absence in the sample is more pronounced for men than for women; however, this gender difference is not statistically significant for the lowest ($p< 0.116$) and low job level ($p= 0.734$), not the medium job level ($p= 0.443$). Figure C shows the days of sickness absence by job level for the three categories of occupational gender composition. The gradient is approximately the same for male dominated and mixed occupations. The rate of absence is lower in female dominated occupations in the lowest job level ($p< 0.001$). The gradient between levels of job level is also much flatter in female dominated occupations (except for the highest category), but it clearly visible in both mixed and male dominated occupations in which those from the lowest level have about twice the average number of days of sickness absence than those from higher job level.

Figure B – Stratification of sickness absence by job level for women and men



Note: Figure reports average number of sickness absence.

Figure C – Stratification of sickness absence by job level for three groups of occupational gender segregation



Note: Figure reports average number of sickness absence.

S4 - Documentation of full model results

These models were run in R 3.2.0 using the R-INLA package.

Negative Binomial (relative differences) model specification

Table 1: Full Model results - Women in male-dominated occupations

	Mean	LB CI	UB CI
Intercept	3.58	1.58	5.58
Low job status	0.10	-0.20	0.39
Middle job status	-0.30	-0.74	0.14
High job status	-0.70	-1.21	-0.19
Education: Middle	0.15	-0.11	0.41
Education: High	-0.09	-0.45	0.27
EGP: lower Service class	0.04	-0.24	0.33
EGP: non-manual routine	-0.40	-1.43	0.67
EGP: routine sales	0.25	-0.39	0.89
EGP: skilled blue-collar	0.00	-0.50	0.50
EGP:unskilled blue-collar	-0.05	-0.53	0.43
Age	0.04	-0.04	0.11
Age^2	-0.03	-0.12	0.06
Log. labor income	0.16	-0.01	0.33
PCS	-0.05	-0.06	-0.03
MCS	-0.02	-0.03	-0.01

Note: Coefficients reported in log. days of sickness absence.

Table 2: Full Model results - Women in mixed occupations

	Mean	LB CI	UB CI
Intercept	4.75	3.48	6.02
Low job status	-0.06	-0.28	0.16
Middle job status	-0.14	-0.38	0.10
High job status	-0.19	-0.48	0.10
Education: Middle	-0.01	-0.18	0.16
Education: High	-0.19	-0.40	0.02
EGP: lower Service class	-0.03	-0.23	0.17
EGP: non-manual routine	-0.12	-0.38	0.13
EGP: routine sales	-0.23	-0.51	0.05
EGP: skilled blue-collar	0.30	0.01	0.60
EGP:unskilled blue-collar	0.25	-0.03	0.54
Age	-0.05	-0.10	-0.01
Age^2	0.08	0.03	0.14
Log. labor income	0.23	0.13	0.34
PCS	-0.05	-0.06	-0.04
MCS	-0.02	-0.02	-0.01

Note: Coefficients reported in log. days of sickness absence.

Linear (absolute differences) model specification

Table 3: Full Model results - Women in female-dominated occupations

	Mean	LB CI	UB CI
Intercept	2.78	1.56	4.01
Low job status	0.00	-0.17	0.18
Middle job status	-0.11	-0.30	0.07
High job status	-0.35	-0.65	-0.05
Education: Middle	-0.16	-0.28	-0.04
Education: High	-0.19	-0.39	0.01
EGP: lower Service class	0.48	-0.30	1.22
EGP: non-manual routine	0.33	-0.44	1.08
EGP: routine sales	0.51	-0.26	1.26
EGP: skilled blue-collar	0.50	-0.33	1.29
EGP:unskilled blue-collar	0.53	-0.25	1.28
Age	0.01	-0.02	0.05
Age^2	-0.02	-0.06	0.03
Log. labor income	0.25	0.17	0.33
PCS	-0.04	-0.05	-0.04
MCS	-0.02	-0.02	-0.01

Note: Coefficients reported in log. days of sickness absence.

Table 4: Full Model results - Men in male-dominated occupations

	Mean	LB CI	UB CI
Intercept	4.97	4.05	5.90
Low job status	0.03	-0.08	0.13
Middle job status	-0.02	-0.15	0.12
High job status	-0.14	-0.30	0.02
Education: Middle	0.01	-0.08	0.09
Education: High	-0.30	-0.44	-0.16
EGP: lower Service class	0.08	-0.05	0.21
EGP: non-manual routine	-0.29	-1.18	0.63
EGP: routine sales	0.39	0.15	0.64
EGP: skilled blue-collar	0.25	0.10	0.41
EGP:unskilled blue-collar	0.31	0.14	0.48
Age	-0.02	-0.05	0.01
Age^2	0.04	0.01	0.07
Log. labor income	-0.06	-0.15	0.04
PCS	-0.03	-0.04	-0.03
MCS	-0.01	-0.02	-0.01

Note: Coefficients reported in log. days of sickness absence.

Table 5: Full Model results - Men in mixed occupations

	Mean	LB CI	UB CI
Intercept	4.96	3.32	6.60
Low job status	0.05	-0.22	0.31
Middle job status	-0.04	-0.32	0.25
High job status	0.11	-0.20	0.41
Education: Middle	-0.23	-0.43	-0.04
Education: High	-0.50	-0.74	-0.26
EGP: lower Service class	0.10	-0.11	0.32
EGP: non-manual routine	0.11	-0.19	0.42
EGP: routine sales	0.29	-0.01	0.59
EGP: skilled blue-collar	0.36	-0.02	0.74
EGP:unskilled blue-collar	0.46	0.16	0.76
Age	-0.04	-0.10	0.01
Age^2	0.07	0.00	0.13
Log. labor income	0.04	-0.12	0.19
PCS	-0.03	-0.04	-0.02
MCS	-0.02	-0.02	-0.01

Note: Coefficients reported in log. days of sickness absence.

Table 6: Full Model results - Men in female-dominated occupations

	Mean	LB CI	UB CI
Intercept	4.27	0.87	7.74
Low job status	-0.33	-0.86	0.19
Middle job status	-0.35	-0.91	0.20
High job status	-0.67	-1.32	-0.03
Education: Middle	-0.30	-0.64	0.03
Education: High	-0.45	-0.94	0.04
EGP: lower Service class	0.69	-1.29	2.53
EGP: non-manual routine	0.89	-1.08	2.72
EGP: routine sales	0.73	-1.24	2.57
EGP: skilled blue-collar	2.00	-0.25	4.19
EGP:unskilled blue-collar	0.71	-1.32	2.61
Age	0.04	-0.06	0.14
Age^2	-0.03	-0.14	0.09
Log. labor income	0.01	-0.23	0.25
PCS	-0.05	-0.07	-0.03
MCS	-0.02	-0.03	-0.00

Note: Coefficients reported in log. days of sickness absence.

Table 7: Full Model results - Women in male-dominated occupations

	Mean	LB CI	UB CI
Intercept	17.54	-6.88	41.94
Low job status	-2.13	-5.55	1.30
Middle job status	-3.06	-8.09	1.96
High job status	-7.08	-12.88	-1.29
Education: Middle	3.01	0.02	6.00
Education: High	0.98	-3.13	5.08
EGP: lower Service class	0.60	-2.84	4.05
EGP: non-manual routine	-7.12	-20.90	6.66
EGP: routine sales	1.61	-5.72	8.94
EGP: skilled blue-collar	0.39	-5.08	5.85
EGP:unskilled blue-collar	-1.16	-6.47	4.14
Age	0.48	-0.47	1.43
Age^2	-0.62	-1.74	0.50
Log. labor income	3.06	1.44	4.69
PCS	-0.58	-0.72	-0.43
MCS	-0.20	-0.32	-0.09

Note: Coefficients reported in log. days of sickness absence.

Table 8: Full Model results - Women in mixed occupations

	Mean	LB CI	UB CI
Intercept	37.79	22.75	52.81
Low job status	-3.09	-5.75	-0.43
Middle job status	-3.89	-6.82	-0.96
High job status	-3.42	-7.00	0.16
Education: Middle	0.10	-1.85	2.04
Education: High	-0.18	-2.68	2.32
EGP: lower Service class	1.18	-1.21	3.56
EGP: non-manual routine	1.60	-1.58	4.77
EGP: routine sales	-0.11	-3.36	3.14
EGP: skilled blue-collar	0.93	-2.62	4.47
EGP:unskilled blue-collar	2.72	-0.68	6.11
Age	-0.23	-0.78	0.31
Age^2	0.32	-0.35	0.99
Log. labor income	2.66	1.55	3.76
PCS	-0.64	-0.74	-0.55
MCS	-0.22	-0.30	-0.15

Note: Coefficients reported in log. days of sickness absence.

Table 9: Full Model results - Women in female-dominated occupations

	Mean	LB CI	UB CI
Intercept	17.55	6.59	28.50
Low job status	0.58	-0.87	2.03
Middle job status	-0.13	-1.70	1.43
High job status	-2.69	-5.31	-0.08
Education: Middle	-2.09	-3.12	-1.05
Education: High	-1.63	-3.27	0.02
EGP: lower Service class	2.59	-3.94	9.13
EGP: non-manual routine	1.66	-4.84	8.15
EGP: routine sales	2.91	-3.59	9.40
EGP: skilled blue-collar	2.10	-4.85	9.03
EGP:unskilled blue-collar	2.21	-4.34	8.75
Age	0.08	-0.24	0.39
Age^2	-0.21	-0.58	0.17
Log. labor income	2.57	1.94	3.21
PCS	-0.45	-0.51	-0.40
MCS	-0.13	-0.18	-0.09

Note: Coefficients reported in log. days of sickness absence.

Table 10: Full Model results - Men in male-dominated occupations

	Mean	LB CI	UB CI
Intercept	40.99	31.63	50.35
Low job status	-0.60	-1.77	0.58
Middle job status	-0.80	-2.31	0.70
High job status	-2.30	-4.01	-0.59
Education: Middle	-0.81	-1.70	0.08
Education: High	-1.18	-2.56	0.20
EGP: lower Service class	0.80	-0.54	2.14
EGP: non-manual routine	-3.81	-11.48	3.84
EGP: routine sales	1.80	-0.83	4.42
EGP: skilled blue-collar	1.37	-0.22	2.96
EGP:unskilled blue-collar	1.28	-0.47	3.02
Age	-0.43	-0.73	-0.13
Age^2	0.51	0.16	0.87
Log. labor income	1.36	0.45	2.27
PCS	-0.52	-0.57	-0.46
MCS	-0.17	-0.21	-0.12

Note: Coefficients reported in log. days of sickness absence.

Table 11: Full Model results - Men in mixed occupations

	Mean	LB CI	UB CI
Intercept	23.25	9.89	36.60
Low job status	0.41	-2.07	2.89
Middle job status	-0.25	-2.86	2.36
High job status	-0.59	-3.49	2.31
Education: Middle	-1.70	-3.33	-0.08
Education: High	-2.52	-4.49	-0.55
EGP: lower Service class	1.01	-0.75	2.77
EGP: non-manual routine	1.27	-1.31	3.84
EGP: routine sales	1.43	-1.06	3.92
EGP: skilled blue-collar	1.59	-1.42	4.59
EGP:unskilled blue-collar	3.86	1.19	6.53
Age	-0.13	-0.59	0.33
Age^2	0.21	-0.32	0.75
Log. labor income	1.06	-0.15	2.28
PCS	-0.28	-0.37	-0.20
MCS	-0.18	-0.25	-0.11

Note: Coefficients reported in log. days of sickness absence.

Table 12: Full Model results - Men in female-dominated occupations

	Mean	LB CI	UB CI
Intercept	52.77	19.23	86.28
Low job status	-8.30	-14.15	-2.45
Middle job status	-3.90	-10.08	2.27
High job status	-9.20	-16.18	-2.22
Education: Middle	-2.64	-6.32	1.03
Education: High	-5.51	-10.54	-0.48
EGP: lower Service class	-4.15	-20.17	11.85
EGP: non-manual routine	-2.93	-18.72	12.83
EGP: routine sales	-4.30	-20.12	11.51
EGP: skilled blue-collar	46.32	26.52	66.09
EGP:unskilled blue-collar	-4.55	-21.08	11.96
Age	0.08	-1.08	1.24
Age^2	-0.11	-1.45	1.24
Log. labor income	2.15	-0.35	4.64
PCS	-0.81	-1.02	-0.61
MCS	-0.18	-0.34	-0.01

Note: Coefficients reported in log. days of sickness absence.