Scand J Work Environ Health. 2015;41(4):338–346. doi:10.5271/sjweh.3499

Explaining educational differences in sickness absence: a population-based follow-up study [1]

by Risto Kaikkonen, MSSc,[2] Tommi Härkänen, PhD, Ossi Rahkonen, PhD, Raija Gould, PhD, Seppo Koskinen, PhD

1. Appendix
2. Correspondence to: Risto Kaikkonen, National Institute for Health and Welfare, Department of Health, Functional Capacity and Welfare, P.O. Box 30, 00271 Helsinki, Finland. [E-mail: risto.kaikkonen@thl.fi]

In the predicted margins (PM) method, the number and the lengths of the sickness absence episodes were estimated in the education groups. The PM is based on the estimates of the regression parameters denoted by vector $\hat{b}$ and the covariate values in vector X_i of individual $i = 1, 2, ..., n$. The prediction is the expected value of the outcome Y_i, which is in the cases of Poisson and gamma regression models $\exp(\hat{b}^{\mathrm{T}} X_i)$. The PM is defined as the weighted mean of the individual predictions

(1) $$\mathrm{PM} := \frac{\sum_{i=1}^{n} w_i \exp(b^T X_i)}{\sum_{i=1}^{n} w_i}$$

where w_i is the weight for the individual i. In this work the postratification weights accounted for the effect of non-response. The covariate vector X_i can be modified in order to produce adjusted estimates, for example, for the education groups. In this case the value of the education group variable of the modified covariate vector X_i^* was set to a fixed value such as the tertiary education for all individuals i.

We have applied the idea of PM for calculating the sickness absence DWY. Let $\hat{b}^{\mathrm{P}}$ and $\hat{b}^{\mathrm{G}}$ denote the regression coefficient estimates of the Poisson and gamma regression models, respectively. The predicted number of sickness absence days for individual i is then the product of the expected number of sickness absence episodes and the expected length of an episode

(2) $$\exp\left\{\left(\hat{b}^P\right)^T X_i\right\} \exp\left\{\left(\hat{b}^G\right)^T X_i\right\}.$$

Note that here the covariates are the same for both regression models, but a generalisation is straightforward. The DWY is then defined as the average of the predictions (2) similarly to PM:

(3) $$\mathrm{PM} := \frac{\sum_{i=1}^{n} w_i \exp\left\{\left(\hat{b}^P\right)^T X_i\right\} \exp\left\{\left(\hat{b}^G\right)^T X_i\right\}}{\sum_{i=1}^{n} w_i}.$$

The standard errors of the population-level estimates were calculated using the delta method.