The development and validation of two prediction models to identify employees at risk of high sickness absence

Corné A. Roelen1,2, Willem van Rhenen2,3, Johan W. Groothoff1, Jac J. van der Klink1, Ute Bültmann1, Martijn W. Heymans4,5

1 Department of Health Sciences, Community and Occupational Medicine, University Medical Center Groningen, University of Groningen, Groningen, The Netherlands
2 365 Occupational Health Service, Utrecht, The Netherlands
3 Center for Human Resource, Organization and Management Effectiveness, Business University Nyenrode, Utrecht, The Netherlands
4 Department of Epidemiology and Biostatistics, EMGO+ Institute, VU University Medical Center, Amsterdam, The Netherlands
5 Department of Methodology and Applied Biostatistics, Faculty of Earth and Life Sciences, VU University, Amsterdam, The Netherlands

Correspondence: Corné A. Roelen, 365ArboNed, PO Box 158, 8000AD Zwolle, the Netherlands, tel: +31 38 4554700, fax: +31 58 4537272, e-mail: corne.roelen@365.nl

Background: Sickness absence (SA) is a public health risk marker for morbidity and mortality. The aim of this study was to develop and validate prediction models to identify employees at risk of high SA. Methods: Two prediction models were developed using self-rated health (SRH) and prior SA as predictors. SRH was measured by the categories excellent, good, fair and poor in a convenience sample of 535 hospital employees. Prior SA was retrieved from the employer’s register. The predictive performance of the models was assessed by logistic regression analysis with high (>90th percentile) vs. non-high (<90th percentile) SA days and SA episodes as outcome variables and by using bootstrapping techniques to validate the models. Results: The overall performance as reflected in the Nagelkerke’s pseudo R² was 11.7% for the model identifying employees with high SA days and 31.8% for the model identifying employees with high SA episodes. The discriminative ability, represented by the area (AUC) under the receiver operating characteristic (ROC), was 0.729 (95% CI 0.667–0.809) for the model identifying employees with high SA days and 0.831 (95% CI 0.784–0.877) for the model identifying employees with high SA episodes. The Hosmer–Lemeshow test showed acceptable calibration for both models. Conclusions: The prediction models identified employees at risk of high SA, but need further external validation in other settings and working populations before applying them in public and occupational health research and care.

Introduction

Sickness absence (SA) is an economic risk marker for disability pensioning1,2 and a public health risk marker for morbidity and mortality.3,4 In a Finnish 10-town study, the overall mortality rate in municipal employees who had more than one long-term (>3 days) SA episode per year was 4.3 times higher in men and 3.3 times higher in women as compared with employees without long-term absences.5 From the French Gazel cohort, it was reported that employees with long-term (>7 days) SA episodes over a 3-year period had a 60% excess risk of early death.6 Structured early consultations with occupational health providers were found to identify employees with unrecognized clinical disorders. Of 142 employees who attended preventive occupational health consultations, 64 (45%) were referred to specialists for further diagnosis and treatment.7 It was shown that preventive consultations were cost-effective in reducing SA in employees with a high SA risk, but not in those with moderate or low SA risks.6 Hence, it is important to identify employees with high SA. Although questionnaires have been developed to detect employees at risk of high SA,5–9 questionnaire surveys frequently have moderate to low response rates and healthy employees may be more likely to respond than employees with health problems.10

Prediction models and rules are alternatives to identify high-risk employees. In public health, various prediction models have been developed to predict the future occurrence of disease and...
target preventive interventions at high-risk subjects, for example, the well-known Framingham models for predicting the risk of cardiovascular disease. The few prediction models in occupational health predict the risk of shoulder pain related SA\textsuperscript{11,12} and the durations of SA due to low back pain\textsuperscript{13} and common mental disorders.\textsuperscript{14} To date, there are no validated models that predict high SA. Moreover, the aforementioned occupational health prediction models were developed by selecting predictors from a set of variables with stepwise statistical techniques and based on the strength of the associations with SA. It is preferred to select predictors independently of their relationship with the outcome variable, because of the instability of selection, biased estimation of regression coefficients and misspecification of variability.\textsuperscript{15,16} Therefore, this study pre-defined prediction models to identify employees at risk of high SA by selecting strong predictors of SA from the literature.

In 1999, Poole proposed a model to predict all-cause SA, including factors such as prior SA, body mass index, smoking and depression or ischaemic heart disease.\textsuperscript{17} In a study of 400 British employees, Poole’s model predicted more SA hours in the high-risk group as compared with the low-risk group.\textsuperscript{18} However, the model was not validated and the variables were obtained by health check-ups. For daily practice, we need prediction models consisting of variables that are readily available or easy to obtain by physicians. Falkenberg et al.\textsuperscript{19} reported that asking employees to rate their own health had as good a quality in predicting SA as extended health check-up models. Self-rated health (SRH) was found to be strongly related to SA in the British Whitehall II studies\textsuperscript{20,21} and a recent Finnish cohort study.\textsuperscript{22} Many other factors have been identified as predictors of SA, but they usually explain 10–15% of the variance in SA. Recently, it was reported that the number of prior SA episodes explained up to 30% of the variance in future SA episodes,\textsuperscript{23} confirming that prior SA is a strong predictor of future SA.\textsuperscript{24–26}

The variables SRH, which is easy to obtain by physicians, and prior SA, which is readily available from employers’ registers or social benefit registries, were used as predictors in models to identify employees at risk of high SA. The purpose of this study was to validate the performance of these pre-defined prediction models.

**Methods**

**Study population**

The study population was a convenience sample of 535 employees working in a Dutch hospital. In October 2008, the hospital employees answered the question ‘In general, would you say your health is: ‘excellent’ (=4), ‘good’ (=3), ‘fair’ (=2) or ‘poor’ (=1).’\textsuperscript{22} This question was derived from the RAND-36 and has been used as a health measure in surveys worldwide.\textsuperscript{27} Ethical approval for this study was not required as the Dutch Act on Medical Research involving Human Subjects does not apply to brief surveys. All participants agreed to link the SRH score to their SA data.

**SA data**

SA was defined as absence from work due to work-related or non-work-related injuries and illnesses. The Human Resources department of the hospital recorded all SA even if employees were only absent from work due to illness for 1 day. The calendar days between the first and the last day of SA were regarded as SA days, irrespective of the actual working hours and partial days off work were considered as full SA days. Roelen et al.\textsuperscript{25} showed that SA in 2 years preceding a baseline survey contributed to the predictability of SA during a 1-year follow-up. Therefore, the total number of SA days in 2007 and 2008 was tallied for each employee as a measure for prior SA days. Likewise, the total number of SA episodes in 2007 and 2008 was accumulated for each employee as a measure for prior SA episodes.

The numbers of SA days and episodes were prospectively recorded for each employee during a 1-year follow-up in 2009. Based on the distribution of SA in 2009, high SA was defined as SA in the upper decile of the distribution of SA. The 90th percentile of the number of SA days in 2009 was 30 days; hence, high SA days was defined as ≥30 days, though it should be acknowledged that these were accumulated SA days and not necessarily consecutive SA days. The 90th percentile of the number of SA episodes in 2009 was three episodes and therefore high SA episodes were defined as ≥3 episodes during follow-up.

**Missing data**

Due to administrative shortcomings, SA data were incomplete in 16% of employees. To fill in these missing data, multiple imputation was applied in R by using the Multivariate Imputation by Chained Equations (MICE) package version 2.5.\textsuperscript{28} This type of imputation uses regression models to estimate the missing values conditional on the information of other variables in the data set. A series of 10 imputed data sets was generated for prediction modelling.\textsuperscript{29} The results of these 10 imputed data sets were summarized by calculating pooled regression coefficients.\textsuperscript{30}

**Prediction modelling**

The prediction models were developed by including the readily available factors SRH and prior SA, which were recognized as strong predictors of SA. Age was added to the prediction models as the effect of age is important in many medical prediction problems.\textsuperscript{16} Together with age, gender is an obvious demographic factor to consider in the prediction of outcome. Gender, however, was not added to the prediction models because of the low number of men (n = 22) in the study population.

Two logistic regression models were developed: one to identify employees with high SA days and one to identify employees with high SA episodes. Age, SRH and prior SA were included as continuous variables. Although SRH is a categorical variable, a linear coding is preferred because dummy coding ignores the ordering of SRH and herewith causes substantial loss of predictive information.\textsuperscript{16} Interaction between the predictors was tested by adding the interaction terms age×SRH, age×prior SA and SRH×prior SA to the logistic regression models, and no significant interactions were found. Linearity was checked for age, prior SA and SRH by adding 3-knot restricted cubic spline functions, which did not improve the regression models. Hence, it was reasonable to assume linearity for the association between the independent variables and SA.\textsuperscript{15,16}

**Apparent performance of the prediction models**

Nagelkerke’s pseudo $R^2$ was used to provide insight in the predictability of high SA by the covariates fitted in the prediction models.\textsuperscript{31} The performance of prediction models comprises two specific characteristics: discrimination and calibration.\textsuperscript{15,16,32} In this study, discrimination referred to the ability of the prediction models to distinguish employees with high SA from those without high SA. The area (AUC) under the receiver operating characteristic (ROC) curve was regarded as a measure for discrimination. An AUC of 0.5 indicates no discrimination above chance and an AUC of 1.0 indicates perfect discrimination. A rough guide for classifying the discriminative ability is AUC 0.9–1.0 excellent, AUC 0.8–0.9 good, AUC 0.7–0.8 fair, AUC 0.6–0.7 poor and AUC <0.6 fail. Calibration referred to the agreement between observed and predicted high SA, and was performed with the Hosmer–Lemeshow goodness-of-fit test. The Hosmer–Lemeshow test compares expected probabilities of high SA with the observed probabilities by deciles of predictions,
based on the null-hypothesis that expected and observed probabilities are equal.

**Internal validation of the prediction models**

A prediction model will perform better in the subjects used to develop the model than in new subjects. In general, this means that the regression coefficients and performance measures are estimated too high, a phenomenon known as overfitting. Overfitted models are too optimistic in predicting outcomes in new subjects. Internal validation corrects for this over optimism. The prediction models were internally validated by bootstrapping, which is a data simulation technique in which subjects are drawn at random with replacement. Thus, a ‘new’ data set is created, called a bootstrap sample, which is of equal size as the original sample, but has a different data structure. Although bootstrapping partially solves the problem of overfitting because it is based on the same individuals, it is the preferred method for internal validation of prediction models in small samples. For each prediction model, 200 bootstrap samples were drawn. The performance of the prediction models was evaluated in the bootstrap samples and compared with the performance in the original sample to calculate the over optimism. The models’ performance after internal validation is the performance that can be expected in new subjects.

**Software**

The prediction models were validated in R by using Harrell’s Regression Modeling Strategies (rms) package, version 3.2-0.34

**Results**

The characteristics of the 535 hospital employees are shown in table 1. A total of 170 (32%) employees rated their health as excellent, 318 (59%) as good, 43 (8%) as fair and 4 (1%) as poor. The distribution of prior SA and SA during follow-up is presented in table 2.

**Performance of the SA days model**

A total of 65 employees had high SA days. The validated prediction model \( \text{ln}(\text{odds}_{\text{SA}}) = 0.601 - 0.016 \times \text{age} + 0.007 \times \text{prior SA} - 0.718 \times \text{SRH} \) had an explained variance of 11.7%, indicating that other factors than those included in the prediction model were also important for predicting SA days. Figure 1 shows the ROC curve visualizing the discriminative ability of the model with an internally validated AUC of 0.729 [95% CI 0.667–0.809], which reflected a fair ability to discriminate employees with high SA days from those without high SA days. A total of 223 employees (42%) had a predicted probability \( \leq 10\% \) and 403 (75%) had a predicted probability \( \leq 20\% \). The sensitivity and specificity were 0.57 and 0.88, respectively, at a predicted probability of 25%, 0.66 and 0.79 at a predicted probability of 20%, 0.77 and 0.71 at a predicted probability of 15%, and 0.91 and 0.66 at a predicted probability of 10%. The Hosmer–Lemeshow test \( P = 0.41 \) reflected acceptable calibration. Table 3 presents the probability of high SA days for each predictor.

**Table 2 SA distribution in the study population \( n = 535 \)**

<table>
<thead>
<tr>
<th>Days</th>
<th>Prior SA in 2007 + 2008, n (%)</th>
<th>SA during follow-up in 2009, n (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>118 (23)</td>
<td>216 (42)</td>
</tr>
<tr>
<td>1–10</td>
<td>163 (32)</td>
<td>191 (37)</td>
</tr>
<tr>
<td>11–29</td>
<td>79 (16)</td>
<td>48 (9)</td>
</tr>
<tr>
<td>30–60</td>
<td>63 (13)</td>
<td>31 (6)</td>
</tr>
<tr>
<td>&gt;60</td>
<td>80 (16)</td>
<td>34 (6)</td>
</tr>
<tr>
<td>Missing</td>
<td>32</td>
<td>15</td>
</tr>
</tbody>
</table>

**Discriminative ability of the prediction models.** The figure shows the ROC curves of the model identifying employees with high SA days (grey line) with an AUC of 0.729 [95% CI 0.667–0.809] and the model identifying employees with high SA episodes (black line) with an AUC of 0.831 [95% CI 0.784–0.877]; an AUC of 0.5 indicates no discrimination above chance and an AUC of 1.0 indicates perfect discrimination.
of symptoms of chronic disease resulting in a pattern of repeated SA. Apart from underlying disorders, repeated SA may also represent a strategy to cope with adverse work conditions. SA is not only determined by medical impairments, but also by the way complaints are perceived and acted upon. Illness behaviour is defined as the varying ways in which individuals respond to bodily sensations, monitor internal states, define and interpret symptoms, make attributions, take remedial actions and utilize sources of health care. Thus, both illness itself and illness behaviour may result in a habitual pattern of repeated SA, which may explain the good performance of the model identifying employees with high SA episodes. The number of SA days shows less of a pattern and therefore the inclusion of prior SA days may not contribute much to the predictability of SA days. Furthermore, the low Nagelkerke’s pseudo $R^2$ of the model identifying employees with high SA days indicated underfitting, which means that important predictors of SA days were missing from the model.

**What about the predictors age and SRH?**

It was unexpected that age was negatively associated with the risk of future SA days, because Dutch national statistics show higher SA in older age groups. The negative association may indicate that only the healthiest employees sustain their work with increasing age. This so-called ‘healthy worker effect’ is conceivable because most employees were working in physically and mentally demanding nursing care. The negative association between age and the risk of future SA episodes is supported by the Dutch national SA statistics. SRH was negatively associated with the risk of high SA days and episodes, which is in agreement with the findings of previous studies. Previously, SRH was found to be associated with various morbidity measures, such as specific health problems, use of health services, changes in functional status and recovery from episodes of ill health. The association of SRH with high SA days (regression coefficient $-0.718$) was of equal strength as the association with high SA episodes (regression coefficient $-0.715$). This supports the finding from the French Gazel cohort that both high SA and low SRH relate to suboptimal health. If the relationship between SA and SRH is driven by rare or severe disorders, then one would expect a stronger association of SRH with SA days than with SA episodes.

**Strengths and weaknesses of the study**

The strength of the study is the use of employer-registered SA data, which restricted recall bias. Furthermore, pre-specified models have the asset that it was not necessary to select predictors based on their relationship with the outcome variable. The prediction models were parsimonious despite the fact that epidemiological studies have provided us with numerous factors that are associated with SA. Parsimonious models are useful in daily practice and do not require health check-ups or questionnaire surveys.

Another strength was that the study population was homogeneous and limited to employees working in one hospital. Hence, it was reasonable to assume a similar influence of organizational policies and practices on employees’ SA. The disadvantage of such a homogeneous population is that the results may not apply to other settings, because SA practices and cultures differ between organizations.

**Practical implications and future directions**

Although internal validity is a prerequisite for external validity, bootstrapping techniques sampled different sets of the same hospital employees. Prediction models that only predict outcomes in the sample in which they were developed are useless. Hence, it is of great importance to assess the external validity, which is the ability of prediction models to provide accurate predictions in other working populations. Justice et al. proposed a five-level
hierarchy in the external validation of prediction models, ranging from prospective validation in different time frames, via independent validation by other researchers to multisite validation. The more numerous and diverse the settings in which prediction models are tested and found accurate in identifying employees with high SA, the more confidently they can be used in untested working populations to select high-risk employees for interventions to prevent or reduce SA.

To select high-risk employees with high specific, cut-off probabilities of 20% (SA days model) and 25% (SA episodes model) could be advised, but the sensitivity at these cut-off probabilities was low due to high false-negative rates. Missing employees at risk of high SA will usually not be problematic, though lower cut-off probabilities, for example 10%, can be chosen to identify as much employees at risk of high SA as possible, at the expense of specificity. The issue of cut-off points will be further addressed in future studies of larger and more heterogeneous working populations.

Conflicts of interest: None declared

Key points
- SA is not only an occupational health problem, but also a public health risk marker for morbidity and mortality.
- Preventive consultations have been reported to be effective in employees at high risk of SA, but not in employees with a moderate or low risk of SA.
- The present study tested two parsimonious prediction models, including age, SRH and prior SA, to identify employees with high SA.
- The prediction models discriminated high-risk employees from low-risk employees and showed acceptable calibration.
- The prediction models need to be externally validated in other settings and working populations before they can be used as public health prognostic tools to select high-risk employees for preventive consultations.

References
34. Harrell FE. Regression modelling strategies. Available at: http://cran.r-project.org/web/packages/mrms (23 September 2011, date last accessed).
Changes in COPD mortality rate after amendments to the Preventive Vaccination Law in Japan

Kosuke Kiyohara, Noriko Kojimahara, Yasuto Sato, Naohito Yamaguchi

Department of Public Health, Tokyo Women's Medical University, Tokyo, Japan

Correspondence: Kosuke Kiyohara, Department of Public Health, Tokyo Women's Medical University, 8-1 Kawadacho, Shinjuku-ku, Tokyo, Japan, tel: +81-3-3353-8111 ext. 22122, fax: +81-3-5269-7420, e-mail: kiyosuke0817@hotmail.com

Background: The Japanese government amended the Preventive Vaccination Law in November 2001 to specify elderly people aged ≥65 years as the target population for influenza vaccinations. The vaccine coverage among this age group rapidly increased thereafter. Our goal was to evaluate how this amendment affected the nationwide mortality rate of chronic obstructive pulmonary disease (COPD). Methods: The number of monthly COPD deaths by gender and age was obtained from the Monthly Vital Statistics Reports of the Ministry of Health, Labour and Welfare. Data between January 1995 and December 2009 were used for analyses. The COPD mortality rate for each month was calculated separately for the two age groups: age <65 years and age ≥65 years. Changes in the COPD mortality rates after amendments were evaluated each month using the Poisson regression analysis to calculate risk ratios (RRs) and to compute 95% confidence intervals (95% CIs) adjusting for gender, age, trend and seasonal variations. Results: After amendments to the law, a statistically significant reduction in the COPD mortality rates was observed in January (RR 0.84; 95% CI 0.81–0.88), February (RR 0.85; 95% CI 0.81–0.89) and March (RR 0.92; 95% CI 0.88–0.96) among the population aged ≥65 years. However, in the population aged <65 years, statistically no significant changes in the COPD mortality rate were found in any month after the amendments were made. Conclusion: A legal approach to improving influenza vaccine coverage for the elderly population would contribute to the risk reduction of COPD deaths during the influenza season.

Methods

Data source

The number of monthly COPD deaths by gender and age was obtained from the Monthly Vital Statistics Reports of the Ministry of Health, Labour and Welfare, targeting all Japanese living in Japan and released 5 months after the survey month. COPD is defined in terms of codes J40–44 in the International Statistical Classification of Diseases and Related Health Problems 10th version (ICD-10). Data between January 1995 and December 2009 were analysed. In addition, Japan’s total population by gender and age was obtained from reports by the Population Estimation in the Statistics Bureau, Ministry of Internal Affairs and Communications for each month.