Development and validation of a prediction model for long-term sickness absence based on occupational health survey variables

Corné Roelen, Sannie Thorsen, Martijn Heymans, Jos Twisk, Ute Bültmann & Jakob Bjørner

To cite this article: Corné Roelen, Sannie Thorsen, Martijn Heymans, Jos Twisk, Ute Bültmann & Jakob Bjørner (2018) Development and validation of a prediction model for long-term sickness absence based on occupational health survey variables, Disability and Rehabilitation, 40:2, 168-175, DOI: 10.1080/09638288.2016.1247471

To link to this article: https://doi.org/10.1080/09638288.2016.1247471

Published online: 10 Nov 2016.

Submit your article to this journal

Article views: 169

View related articles

View Crossmark data

Citing articles: 2 View citing articles
Development and validation of a prediction model for long-term sickness absence based on occupational health survey variables

Corné Roelen, Sannie Thorsen, Martijn Heymans, Jos Twisk, Ute Bültmann and Jakob Björner

Department of Health Sciences, Community and Occupational Medicine, University Medical Center Groningen, University of Groningen, Groningen, The Netherlands; Department of Epidemiology and Biostatistics, VU University Medical Center, VU University, Amsterdam, The Netherlands; National Research Center for the Working Environment, Copenhagen, Denmark; Department of Public Health, University of Copenhagen, Copenhagen, Denmark

ABSTRACT

Purpose: The purpose of this study is to develop and validate a prediction model for identifying employees at increased risk of long-term sickness absence (LTSA), by using variables commonly measured in occupational health surveys.

Materials and methods: Based on the literature, 15 predictor variables were retrieved from the DANish National working Environment Survey (DANES) and included in a model predicting incident LTSA (≥4 consecutive weeks) during 1-year follow-up in a sample of 4000 DANES participants. The 15-predictor model was reduced by backward stepwise statistical techniques and then validated in a sample of 2524 DANES participants, not included in the development sample. Identification of employees at increased LTSA risk was investigated by receiver operating characteristic (ROC) analysis; the area-under-the-ROC-curve (AUC) reflected discrimination between employees with and without LTSA during follow-up.

Results: The 15-predictor model was reduced to a 9-predictor model including age, gender, education, self-rated health, mental health, prior LTSA, work ability, emotional job demands, and recognition by the management. Discrimination by the 9-predictor model was significant (AUC = 0.68; 95% CI 0.61–0.76), but not practically useful.

Conclusions: A prediction model based on occupational health survey variables identified employees with an increased LTSA risk, but should be further developed into a practically useful tool to predict the risk of LTSA in the general working population.

IMPLICATIONS FOR REHABILITATION

- Long-term sickness absence risk predictions would enable healthcare providers to refer high-risk employees to rehabilitation programs aimed at preventing or reducing work disability.
- A prediction model based on health survey variables discriminates between employees at high and low risk of long-term sickness absence, but discrimination was not practically useful.
- Health survey variables provide insufficient information to determine long-term sickness absence risk profiles.
- There is a need for new variables, based on the knowledge and experience of rehabilitation professionals, to improve long-term sickness absence risk profiles.

Introduction

Western countries spend on an average of 10% of their social expenditures on sickness absence and disability benefits.[11] Sickness absence benefits account for most of these expenditures. Long-term sickness absence (LTSA) distances employees from the workplace and may ultimately end up in disability pension or unemployment, marginalizing employees from the labor market and reducing social participation.[2–5]

Previous studies have presented prediction models and rules for the outcomes of LTSA.[6–10] Most of these models and rules predict the duration of LTSA due to low back pain with specific variables, such as fear avoidance beliefs and pain behavior. For practical reasons, it would be more convenient to use variables which are readily available, for example, from public health registers. Pedersen et al. [11] developed a prediction model based on variables (age, gender, socioeconomic position, job type, chronic disease, prior records of sick-listing, and unemployment) obtained from the Danish register for sickness absence compensation benefits and social transfer payments (DREAM). The prediction model identified employees at risk of disability pension and unemployment, but not those at risk of LTSA. The authors concluded that other variables had to be included into a prediction model for LTSA.

Variables could be retrieved from health surveys. In Denmark, the health status, life style, and occupational exposures are surveyed every 5 years since 1990 in large samples of at least 5000 subjects randomly drawn from the general population.[12] The results have been presented in the Danish Work Environment Cohort Studies (DWECS). From the DWECS studies, it was reported...
that older employees, women, and employees of lower socioeconomic classes had a higher LTSA risk.\footnote{13–16} Smoking was also related to an increased LTSA risk, while alcohol consumption and leisure time physical activity were not.\footnote{17} With regard to occupational exposures, physical job demands \footnote{18,19} as well as psychosocial working conditions were found to predict LTSA.\footnote{18,20}

From their review of DWECS studies, Burr et al. \footnote{21} concluded that life style factors and the working environment played an equally important role in predicting LTSA. The objective of the present study was to investigate if the life style and work environment variables obtained from DANES can be used to identify employees with an increased risk of LTSA before they report sick. If prediction models identify employees at increased risk of LTSA, then healthcare professionals could use these models to refer high-risk employees to rehabilitation programs aimed at preventing work disability.

**Materials and methods**

Secondary analyses were conducted with data obtained from the Danish National working Environment Survey (DANES), an extra survey in 2008 halfway between DWECS 2005 and 2010. DANES was meant as a supplement to DWECS with a special focus on the psychosocial work environment. A total of 9913 Danish citizens aged 18–59 years were invited to participate in DANES late 2008 until early 2009.

A total of 6524 (66%) citizens participated in DANES. We drew a random sample of 4000 DANES participants for developing the prediction model; the data of the remaining 2524 participants were used to validate the prediction model. Figure 1 shows that 1508 (38%) employees of the development sample and 1014 (40%) employees of the validation sample were excluded predominantly because they were unemployed. The 102 participants of the development sample and 79 participants of the validation sample who were on sick leave or had had LTSA in the 12 weeks prior to baseline were also excluded from the analyses to ensure that the model predicted incident LTSA rather than a relapse of recent LTSA. The data of 2492 employees were used for developing the prediction model and the data of 1510 employees for validating the prediction model (Figure 1).

**Ethics approval**

DANES was approved by the Danish Data Protection Agency (reference 2007-54-0059). All procedures followed were in accordance with the Helsinki Declaration.

**Predictor variables**

Age and gender were retrieved from the Danish central person register. Education and occupation were obtained from DANES; occupation was used to assign employees to socioeconomic classes according to the European Socioeconomic Classification.\footnote{22}

**Health variables**

DANES measured self-rated health (SRH) with the question “In general, would you say your health is excellent (=5), very good (=4), good (=3), fair (=2), poor (=1)”. SRH reflects physical rather than mental health,\footnote{23} while mental health has been identified as an important LTSA predictor.\footnote{24} DANES measured mental health with the Mental Health Inventory (MHI) scale asking employees “How much of the time during the last 4 weeks, have you…” (1) been a very nervous person, (2) felt so down that nothing could cheer you up, (3) felt calm and peaceful, (4) felt downhearted and blue, and (5) been a happy person. The questions were answered on a 5-point frequency scale ranging from “all of the time” to

---

*Figure 1. Study sample flow chart.*
“none of the time” Scores were summed and divided by the number of MHI items to yield an average score, ranging from 1 (poor mental health) to 5 (excellent mental health). Smoking habits were assessed by the item “Do you smoke?” with responses “daily”, “on occasion”, “used to smoke but stopped”, and “have never smoked”. The first two responses were combined into the category “smoker”, so that there were three categories: smoker, ex-smoker, and non-smoker.

The interaction between health status and work is captured by the work ability concept,[25] which was investigated with the question “Is your ability to work reduced because of illness, accident, or being worn down?” with response categories “yes to a high degree” (=1), “yes to some degree” (=2), “no not very much” (=3), and “no not at all” (=4). The reliability of this item has been calculated in a test-retest design, yielding an intra-class coefficient of 0.82.[26]

Work variables
Physical job demands were investigated by a question on general physical activity in work: “How would you describe your physical activity at your main job?” with response categories “sedentary light work” (=1), “light work standing or walking” (=2), “work standing or walking with fairly heavy lifting or carrying” (=3), and “heavy or fast-moving work” (=4). This single item question has previously been shown to have very good reproducibility and good validity for assessing physical workloads in the general population.[27]

Psychosocial working conditions were assessed by items of the short-list version of the Copenhagen Psychosocial Questionnaire (COPSOQ).[28,29] The reliability of the COPSOQ items was tested by Cohens weighted kappa (κ) in a test–retest design. Cognitive job demands (Does your work require you to make difficult decisions?; κ = 0.58), role conflicts in work (Are contradictory demands placed on you at work?; κ = 0.59), and recognition by the management (Is your work recognized and appreciated by the management?; κ = 0.56) were assessed with single items. Emotional job demands (Does your work put you in emotionally disturbing situations? Do you have to relate to other people’s personal problems as part of your work?; κ = 0.61) and leadership quality (Would you say your immediate superior is good at solving problems? Would you say your immediate superior is good at work planning?; κ = 0.54) were measured with two items each. Scores on the items were standardized to range between 0 and 1, higher values representing higher levels of the condition measured.

Long-term sickness absence (LTSA)
Sickness absence was defined as a financially compensated leave due to any (i.e., work-related and non-work-related) illness. At the time of the study, sickness absence was recorded in DREAM when lasting four weeks or longer.[30] Consequently, we defined LTSA as sickness absence lasting ≥4 consecutive weeks. LTSA in the two years prior to baseline (no = 0, yes = 1) was used for the predictor variable “prior LTSA”. LTSA occurring in the year following baseline (no = 0, yes = 1) was the outcome variable.

Statistical analysis
The prediction models were developed and validated by using SAS, version 9.3 (SAS Inc., Cary, NC). Prediction models will better predict outcomes in the sample used to develop the model than in new samples of employees. This phenomenon, known as “over-optimism” restricts the external validity of prediction models.[31] Because of the large DANES sample size, it was possible to use part of the DANES sample for the development of the prediction model and part of the sample for the model’s external validation. Due to this split-sample design, the study provided LTSA predictions that can be generalized to the general population.

Development of the prediction model
Based on the DWECs studies, we identified 15 DANES variables as potential LTSA predictors. To calculate the sample size for developing the prediction model, we took the criterion of ≥10 LTSA events per variable and an estimated 6% LTSA incidence. We would need a sample of more than 2500 employees; bearing in mind the exclusion of unemployed citizens and those on (maternity or sick) leaves from work, we decided to draw a random sample of 4000 DANES participants for the development of the prediction model. Age, education, SRH, mental health, work ability, physical work, and psychosocial working conditions were included in logistic regression analysis as linear variables; prior LTSA, gender, socioeconomic class, and smoking were included as categorical variables. LTSA occurring (no = 0, yes = 1) during follow-up was the outcome of logistic regression analysis. Nagelkerke’s pseudór2 of logistic regression analysis was regarded as a measure of the model’s predictive performance. To reduce the model for practical use, variables were removed from the 15-predictor model by a backward stepwise procedure with Akaike’s Information Criterion as stopping rule.

Validation of the prediction models
Risk predictions of the 15-predictor model and the reduced model were calibrated by plotting predicted LTSA risks against observed LTSA frequencies in a calibration graph. If risk predictions are perfect, then the calibration graph is a straight line with slope = 1.[31]

Discrimination refers to the ability of the 15-predictor model and the reduced model to discriminate between employees with and without incident LTSA during 1-year follow-up. The discriminative ability was investigated with receiver operating characteristic (ROC) analysis, regarding the area-under-the-ROC curve (AUC) as discrimination measure. AUC = 0.50 indicates no discrimination above chance and AUC >0.50 represents significant discrimination; AUC ≥0.75 reflects practically useful discrimination.

Results
The baseline characteristics of the development sample (N = 2492) and the validation sample (N = 1510) are shown in Table 1.

Female gender, prior LTSA, smoking, and both physical and emotional job demands were associated with higher odds of LTSA during follow-up. Alternatively, education, SRH, mental health, work ability and recognition by the management were associated with lower odds of LTSA (Table 2).

Development of the prediction model
In the development sample, 146 (5.9%) employees had LTSA during follow-up. When all 15 predictor variables were included in the model (Nagelkerke’s R2 = 0.16), prior LTSA (odds ratio [OR] 2.18; 95% confidence interval [CI] 1.29–3.70) and self-rated work ability (OR = 0.73; 95% CI 0.56–0.94) were the strongest LTSA predictors. In six backward steps, role conflicts in work, physical job demands, socioeconomic class, leadership quality, cognitive job demands, and smoking were removed from the model.
The reduced 9-predictor model retained age, gender, education, SRH, mental health, prior LTSA, work ability, emotional job demands, and recognition by the management as predictor variables (Nagelkerke’s $R^2 = 0.14$).

**Validation of the prediction model**

In the validation sample, 82 (5.4%) employees had LTSA during follow-up. The calibration graph showed slopes of 0.58 and 0.63 for the 15- and 9-predictor model, respectively, indicating that both models over-predicted the LTSA risk (Figure 2). Figure 3 shows that discrimination by the 15-predictor model (AUC = 0.70; 95% CI 0.63–0.76) was comparable with discrimination by the 9-predictor model (AUC = 0.68; 95% CI 0.61–0.76).

**Discussion**

Fifteen variables were identified based on the Danish Work Environment Cohort Studies (DWECS) and included in a prediction model for the risk of long-term (≥4 weeks) sickness absence (LTSA) during 1-year follow-up. Backward stepwise model reduction revealed a 9-predictor model including age, gender, education, SRH, mental health, prior LTSA, work ability, emotional job demands, and recognition by the management. This 9-predictor model over-predicted the LTSA risk, but discriminated between employees with and without LTSA during follow-up.

In the process of backward stepwise model reduction, socioeconomic class was excluded as predictor variable. Probably socioeconomic class did not add predictive value to a model that
included education. When education was left out of the prediction model at start, socioeconomic class stayed in the reduced model [data not shown]. Thus, some indicator of socioeconomic position is required to predict the LTSA risk.[16,32–35] Manual work can also be regarded as a marker of low socioeconomic position, which might explain why the predictor variable physical job demands was discarded from the model. In our study, the bivariate association between physical job demands and LTSA was significant, but weakened and became non-significant when health and education were included in the model [data not shown].

It has been reported that the fraction of LTSA attributable to psychosocial working conditions is small.[18] However, if all psychosocial working conditions were excluded, the model’s predictive performance deteriorated significantly. Emotional job demands and recognition by the management remained in the 9-predictor model. The traditional psychosocial work factors quantitative job demands (i.e., workload and workspace) and decision latitude were not included in the prediction model at start, because the DWECs studies are inconclusive on their associations with LTSA. These factors, however, are commonly included in studies of psychosocial working conditions. Therefore, we investigated the performance of prediction models including quantitative job demands and decision latitude, but found that these models did not better identify employees with incident LTSA during follow-up than the 9-predictor model [data not shown].

Occupational health survey variables better than chance identified employees at increased risk of future LTSA. In ROC-analysis, the 9-predictor model had an AUC of 0.68, which implicates that the model correctly discriminated between employees with and without incident LTSA in 68% of the cases. Although statistically significant, discrimination of this magnitude is considered moderate and below the level recommended for practical use.[31]

Study strengths and weaknesses

The use of data representative of the general working population and the different data sources (DANES for predictor variables and DREAM register for LTSA data) were strengths of the study.

<table>
<thead>
<tr>
<th>Predictor variable</th>
<th>OR (95% CI)*</th>
</tr>
</thead>
<tbody>
<tr>
<td>Age (years)</td>
<td>0.99 (0.97–1.02)</td>
</tr>
<tr>
<td>Gender men</td>
<td>1</td>
</tr>
<tr>
<td>Education</td>
<td>1.75 (1.15–2.66)</td>
</tr>
<tr>
<td>Socioeconomic class higher professionals and managers</td>
<td>0.86 (0.79–0.95)</td>
</tr>
<tr>
<td>Lower professionals and managers</td>
<td>2.81 (1.03–7.62)</td>
</tr>
<tr>
<td>Higher clerical, services and sales workers</td>
<td>1.87 (0.68–5.17)</td>
</tr>
<tr>
<td>Lower supervisors and technicians</td>
<td>0.69 (0.08–6.02)</td>
</tr>
<tr>
<td>Lower clerical, services and sales workers</td>
<td>5.90 (2.23–15.63)</td>
</tr>
<tr>
<td>Skilled workers</td>
<td>1.95 (0.59–6.49)</td>
</tr>
<tr>
<td>Semi- and unskilled workers</td>
<td>4.28 (1.61–11.33)</td>
</tr>
<tr>
<td>Prior LTSA no</td>
<td>1</td>
</tr>
<tr>
<td>Yes</td>
<td>3.21 (1.98–5.18)</td>
</tr>
<tr>
<td>Self-rated health (range 1 = poor to 5 = excellent)</td>
<td>0.47 (0.37–0.61)</td>
</tr>
<tr>
<td>Mental health (range 1 = poor to 5 = excellent)</td>
<td>0.43 (0.32–0.56)</td>
</tr>
<tr>
<td>Smoking non-smoker</td>
<td>1</td>
</tr>
<tr>
<td>Ex-smoker</td>
<td>1.41 (0.84–2.38)</td>
</tr>
<tr>
<td>Smoker</td>
<td>1.95 (1.21–3.14)</td>
</tr>
<tr>
<td>Work ability</td>
<td>0.53 (0.43–0.64)</td>
</tr>
<tr>
<td>Physical job demands</td>
<td>1.31 (1.06–1.61)</td>
</tr>
<tr>
<td>Cognitive job demands (range 0 = low to 1 = high)</td>
<td>1.04 (0.29–2.20)</td>
</tr>
<tr>
<td>Emotional job demands (range 0 = low to 1 = high)</td>
<td>3.67 (1.72–7.84)</td>
</tr>
<tr>
<td>Role conflicts in work (range 0 = low to 1 = high)</td>
<td>1.31 (0.60–2.86)</td>
</tr>
<tr>
<td>Leadership quality (range 0 = low to 1 = high)</td>
<td>0.42 (0.17–1.01)</td>
</tr>
<tr>
<td>Recognition by management (range 0 = low to 1 = high)</td>
<td>0.26 (0.12–0.56)</td>
</tr>
</tbody>
</table>

*aOdds ratio (95% confidence interval).
Another asset is that the prediction model was developed in a random sample with sufficient statistical power in terms of the number of LTSA events to provide robust risk estimates. The split-sample design with validation of LTSA prediction models in employees other than those used to develop the prediction model, yielded risk predictions valid for the general working population.

The study presents the results of secondary analyses of data that had been collected without the idea of developing a prediction model in mind. Although it is generally preferred to define and measure variables in line with the objectives of a study, the aim of the current study was to investigate if variables commonly measured in occupational health surveys can identify employees at increased risk of future LTSA. Therefore, we did not regard the secondary analysis as a weakness of the study.

The definition of LTSA as sick leaves lasting $\geq 4$ consecutive weeks might be weaknesses of the study, because we could have missed emerging disability, presenting with sick leaves shorter than four weeks. The use of occupational health survey variables to predict sick leaves $< 4$ consecutive weeks as precursor of future LTSA and work disability remains to be investigated. It has been reported that LTSA and work disability develop over time. Therefore, future studies should consider a longitudinal design with repeated measurements during longer follow-up periods.

Another limitation of the study was the use of employee-reported predictor variables. It has been proposed that responses on self-reported work exposures are at least partly driven by personal dispositions, mood, expectations, and experiences. For prediction purposes, however, it is not necessary to measure work exposures objectively. On the contrary, subjectively experienced work exposures may be more predictive of LTSA than objectively measured work exposures. For example, an employee who experiences job demands as high is more likely to report sick and develop LTSA than an employee who does not experience job demands as high, even if job demands were objectively the same for both employees.

**Practical implications and directions for future research**

The first step in developing a prediction model for LTSA would be to investigate if such a model can be based on the variables which are commonly measured in occupational health surveys. Although these variables better than chance identified employees at increased risk of future LTSA, the present study showed that a prediction model based on occupational health survey variables moderately discriminated between employees with and without LTSA during 1-year follow-up. Apparently, occupational health survey variables do not provide sufficient information to adequately discriminate employees at high risk of LTSA from those at low risk. Interestingly, the predictive performance in terms of explained variance (Nagelkerke’s $R^2$ values) of the model based on occupational health survey variables was within the range of previously determined models based on specific variables, explaining 6–25% of the variance in the duration of LTSA in sicklisted employees. Nevertheless, healthcare providers cannot yet use the model presented in the current study to screen the general working population for risk of LTSA and refer high-risk employees to preventive rehabilitation programs.

The moderate discrimination may be explained by the way variables were measured. To prevent DANES from becoming too extensive, psychosocial working conditions were measured with single-items or brief scales. Although their (test–retest) reliability was good, such scales may not fully capture the physical and psychosocial exposures at the workplace and consequently not provide sufficient information for identifying employees at increased LTSA risk on a practically useful level. Future studies could search for additional predictor variables to improve the discriminative ability of the LTSA prediction model. Healthcare providers could...
then include these additional variables in occupational health surveys if they want to identify employees at risk of LTSA.

The moderate discrimination may also be due to the fact that we did not differentiate between LTSA diagnoses. It is well conceivable that predictors of LTSA due to mental disorders differ from predictors of LTSA due to musculoskeletal and other somatic disorders. It would be interesting to differentiate risk predictions by LTSA cause in future studies.

**Conclusion**

Prediction models based on occupational health survey variables discriminated between non-sicklisted employees with and without LTSA during 1-year follow-up, but have to be further developed into practically useful tools for predicting LTSA in the general working population.

**Disclosure statement**

The authors report that they have no conflicts of interest.

**Funding**

The present study was not funded.

**References**


