The history of registered sickness absence predicts future sickness absence

C. A. M. Roelen1,2, P. C. Koopmans3, J. A. H. Schreuder2, J. R. Anema1,4,5 and A. J. van der Beek1,4,5

1Department of Public and Occupational Health, EMGO Institute for Health and Care Research, VU University Medical Center, Van der Boechorststraat 7, 1081 BT Amsterdam, The Netherlands, 2ArboNed Corporate Accounts, PO Box 158, 8000 AD Zwolle, The Netherlands, 3ArboNed Statistics, PO Box 141, 9700 AC Groningen, The Netherlands, 4Body@Work, Research Center Physical Activity, Work and Health, TNO-VU/VUmc, Van der Boechorststraat 7, 1081 BT Amsterdam, The Netherlands, 5Research Center for Insurance Medicine AMC–UMCG–UWV–VU University Medical Center, Van der Boechorststraat 7, 1081 BT Amsterdam, The Netherlands.

Correspondence to: C. A. M. Roelen, ArboNed Corporate Accounts, PO Box 158, 8000 AD Zwolle, The Netherlands. Tel: +31 38 4554700; fax: +31 58 4537272; e-mail: corne.roelen@arboned.nl

Background The history of sickness absence has been found to predict future sickness absence. The individual number of days and episodes of sickness absence were ascertained for 762 hospital employees from 2004 to 2008 inclusive. Past sickness absence was included stepwise in ordinal regression models. The explained variance of the ordinal regression models reflected the extent to which future sickness absence could be predicted and was expressed in percentages calculated as Nagelkerke’s pseudo $R^2 \times 100\%$.

Methods

Aims To establish the review period of historical sickness absence data that is needed to predict future sickness absence.

Methods

The individual number of days and episodes of sickness absence were ascertained for 762 hospital employees from 2004 to 2008 inclusive. Past sickness absence was included stepwise in ordinal regression models. The explained variance of the ordinal regression models reflected the extent to which future sickness absence could be predicted and was expressed in percentages calculated as Nagelkerke’s pseudo $R^2 \times 100\%$.

Results

A total of 551 employees (72%) had complete data and were eligible for regression analysis. Days of sickness absence in the past year predicted up to 15% of future days of sickness absence. Adding the sickness absence data of the past 2 or 3 years did not further increase the predictability of days of sickness absence. Episodes of sickness absence in the past year predicted up to 25% of future episodes of sickness absence. The predictability of episodes of sickness absence increased to 30% when the past 2 years of sickness absence were included in the regression model, but did not further increase when sickness absence of the past 3 years was included.

Conclusions Employees who are more likely to have an above average sickness absence can be identified from their history of sickness absence in the past 2 years.

Key words Absenteeism; history of sickness absence; sick leave; sickness absence record.

Introduction

Sickness absence is an important socioeconomic problem because of lost productivity, staff replacement costs and both social insurance and medical costs of long-term work disability. Sickness absence is regarded as a public health risk indicator predicting future morbidity and mortality [1–3]. However, it is unlikely that there is a direct causal association between sickness absence and mortality [4]. Instead, health problems are probably related to both sickness absence and to future disease [3,4]. Therefore, sickness absence may provide general practitioners with a useful tool to identify patients with an increased risk of serious disease [4]. Taylor [5] was the first to report that sickness absence had a recurrent pattern and Poole [6] concluded that current and past sickness absence predicted future sickness absence. Dekkers-Sánchez et al. [7] found evidence in the literature for associations between an individual's history of sickness absence and future long-term sickness absence. Later, it was reported that the prior work attendance record was the best predictor of absenteeism among nurses [8] and that the risk of sickness absence increases with the number of prior episodes [9]. Sixty per cent of workers, who had four or more episodes of sickness absence in a baseline period of 1 year, repeated this number of episodes per year at least once during a follow-up period of 4 years [10].

Patterns of sickness absence may be due to recurring episodic disorders, such as migraine and asthma, or due to exacerbation of chronic disease [11]. On the other hand, repeating patterns of sickness absence may also
reflect a coping strategy [12], or may be due to excessive or uncertain responses to non-specific pain and other bodily sensations [13].

We expect that the pattern of sickness absence can be used to identify employees who are more likely to have above average sickness absence. The period of review of historical sickness absence records, however, has not been established. Therefore, we designed a sickness absence register study addressing the research questions:

(i) How far back in history do past days of sickness absence contribute to the predictability of future days of sickness absence?

(ii) How far back in history do past episodes of sickness absence contribute to the predictability of future episodes of sickness absence?

Methods

The study population consisted of employees working in a Dutch hospital. To ensure a homogeneous sample with regard to the type of work and working environment, employees working in the clinical wards or the outpatient clinic were involved in this study. Their sickness absence data were obtained from the personnel administration’s sickness absence register, which is updated daily with staff (new employees, retirements, dismissals and resignations) and sickness absence (sick reports and recovery dates) details. The hospital’s sickness absence register is accredited by the Netherlands Institute for Accreditation in Health Care according to the International Accreditation Program of the International Society for Quality in Health Care.

Employees with complete data in the period from 2004 to 2008 were eligible for analysis. Employees with incomplete data, for instance because they resigned, retired or received a disability pension after long-term sickness absence, were excluded from the analyses as regression models can only be compared when each model contains the same sample of subjects. Ethical clearance was not required for this observational study because the sickness absence data were retrieved without disclosing medical diagnoses or personal files. All employees gave informed consent to the use of their sickness absence data for scientific research.

Sickness absence was defined as not coming to work when scheduled because of illness. In the hospital, employees report sick to their manager who sends the sick report to the personnel administration of the hospital. The personnel administration of the hospital records sickness absence irrespective of its duration. Thus, even 1 day of absence from work due to illness is recorded in the sickness absence register. Episodes of sickness absence lasting up to 3 weeks remain self-certified. In the Netherlands, employees usually visit the occupational physician in the third or fourth week of sickness absence for a medical certification of sickness and socio-medical guidance supporting return to work. On the day that the employee resumes work, the manager sends a recovery report to the personnel administration of the hospital.

The calendar days between the first and last day of registered sickness absence were counted as days of sickness absence, irrespective of the actual work hours. Partial days off work were regarded as full days of sickness absence. The number of days of sickness absence was determined for each employee in each year from 2004 to 2008 inclusive. Likewise, the number of episodes of sickness absence was counted for each employee during this time. Episodes of sickness absence extending into the next calendar year were cut off at 31 December of each year and treated as a single episode, not as an episode of sickness absence in both years. The numbers of days of sickness absence in each year were correlated. The numbers of episodes of sickness absence in each year were correlated as well.

Based on the distribution of days of sickness absence, four ordinal duration categories were defined: no (0 days), low (1–7 days), medium (7–21 days) and high (>21 days) sickness absence in a year. It should be noted that these numbers are the cumulative days in a year and not necessarily consecutive. For instance, when an employee has an episode of sickness absence lasting 4 days and another episode lasting 6 days in the same year, this is cumulatively 10 days of sickness absence, which is categorized as medium sickness absence. The ordinal categories were used as an outcome for ordinal regression analysis, also known as polytomous universal model analysis [14,15]. This extension of logistic regression for ordered categorical data was recently used for predicting sickness absence [16]. The days of sickness absence in the past years were added stepwise to an ordinal regression model as independent variables, including age (≤35, 36–45, >45 years) and gender as categorical covariates. The extent to which past days of sickness absence contributed to the predictability of future sickness absence was assessed by the Nagelkerke’s pseudo $R^2$ [17], which was calculated as percentage (i.e. Nagelkerke’s pseudo $R^2 \times 100\%). A change was considered relevant when Nagelkerke’s pseudo $R^2$ increased >10% [14].

Based on the distribution of episodes of sickness absence, four ordinal frequency categories were defined: no (zero episodes), low (one episode), medium (two episodes) and high (greater than or equal to three episodes) sickness absence in a year. The ordinal groups were used as outcome for regression analysis, including the number of past episodes of sickness absence as independent variables and age categories and gender as covariates. The effects of stepwise addition past episodes of sickness absence to the regression model were assessed by the changes in Nagelkerke’s pseudo $R^2\%$, and a change was considered relevant when Nagelkerke’s pseudo $R^2$ increased >10%.
Because most employees had no sickness absence (range 31–40%) or low sickness absence (range 25–30%), meaning that the probability belonging to these categories was highest, a negative log–log link function was used in all ordinal regressions [15]. A total of eight stepwise regression analyses were performed, of which four assessed the associations between days of sickness absence and four examined the associations between episodes of sickness absence. Therefore, the level of significance of was set at 1% (i.e. \( \alpha = 0.01 \)).

**Results**

The hospital employed 1153 employees of whom 762 worked in the clinical wards or the outpatient clinic. Employees with incomplete data, for instance because they resigned, retired or received a disability pension after long-term sickness absence, were excluded from the analyses leaving 551 employees with complete sickness absence data from 2004 to 2008 eligible for inclusion.

The numbers of days of sickness absence were interrelated and the correlations were the strongest with the days in the preceding year and the following year. For example, Table 1 shows that the number of days of sickness absence in 2006 was correlated with the days of sickness absence in 2005 [Pearson correlation coefficient (r) = 0.52] and 2007 (r = 0.40), whereas the correlations with 2004 and 2008 were r = 0.14 and r = 0.22, respectively. The episodes of sickness absence were more strongly interrelated than days of sickness absence, with Pearson correlation coefficients ranging between r = 0.49 and r = 0.61 (Table 1).

The number of days of sickness absence in 2008 was positively associated with the days of sickness absence in 2007 (b = 0.39) and 2006 (b = 0.35). This means that the probability of being in a higher category for episodes of sickness absence in 2008 increased with the number of episodes of sickness absence in 2007 and 2006 (Table 3). The number of episodes in 2007 was positively associated with episodes of sickness absence in 2006 (b = 0.36), 2005 (b = 0.32) and 2004 (b = 0.26). The number of episodes of sickness absence in 2006 was associated with 2005 (b = 0.36) and 2004 (b = 0.16) and the number of episodes in 2005 with 2004 (b = 0.46).

The extent to which episodes of sickness absence were predicted by the past year’s history ranged from 16 to 25%, which was higher than for days of sickness absence. When the episodes of sickness absence of the past 2 years were included in the model, Nagelkerke’s pseudo \( R^2 \) increased to a range of 19–30% (Table 3). When the episodes of sickness absence in the past 3 years were included, the further increase in predictability was <10% and therefore not considered relevant.

**Discussion**

Sickness absence was associated with a registered history of sickness absence. The results of this study showed that the past year of registered sickness absence contributed to the predictability of future days of sickness absence and the past 2 years contributed to the predictability of future episodes of sickness absence.

The strength of the study lies in the fact that employer-registered sickness absence data were used, which excluded recall bias of self-reported sickness absence. It should be noted that it was necessary to select persons who had been working in the hospital for at least 5 years, to ensure that the different steps in the regression analyses contained the same population of subjects and the Nagelkerke’s pseudo \( R^2 \) percentages were comparable. This may have resulted in the selection of healthy workers and may have underestimated the association between the history of sickness absence and future sickness absence. Although the study population was relatively small, it was homogeneous in that it was limited to employees working in the clinical wards and outpatient clinic of one hospital. Hence,
employees were subject to the same type of work, work environment and company policies. The disadvantage of a homogeneous population is that the results of this study cannot be generalized to the total workforce, all the more because it has been reported that there are differences in sickness absence practices and cultures between companies and countries [18,19]. Therefore, more studies in different companies and across countries are necessary to develop practical guidelines.

A further limitation is that the ordinal groups were based on the distribution of sickness absence in the study population. It is likely that ordinal groups in other samples will differ from our study, although the categories of days of sickness absence in our study resembled those defined by Falkenberg et al. [16]. Besides, the problem in our study population was that most employees had no sickness absence or low sickness absence. Therefore, ordinal regression was performed with a negative log–log function that corrected for the higher probability of being in lower ordinal categories. Although the regression coefficient indicated the strength and direction of the association, the disadvantage of the negative log–log estimate was that the regression coefficient could not be interpreted in common terms of association such as rate ratios or odds ratios. However, the study concentrated on the contribution of the history of sickness absence to the predictability of future sickness absence instead of associations. The regression models predicted 4–17% of future days of sickness absence and 16–31% of future episodes of sickness absence.

These results confirm that sickness absence can be predicted from the history of sickness absence. Employees with a history of prolonged or frequent sickness absence in past years were more likely to have above average sickness absence. Dekkers-Sánchez et al. [7] reviewed cohort studies of workers, who were on sick leave for at least 6 weeks at baseline. They extracted 14 individual factors and 2 work-related factors predicting long-term sickness absence. However, the level of evidence was found to be sufficient only for older age and the history of sickness absence. Our study showed that the frequency of sickness absence was better predicted by the history of sickness absence than the duration in terms of days of sickness absence, which confirms earlier reports [9,10]. Duijts et al. [20] showed that the odds of a new episode of sickness absence due to psychosocial complaints were 1.95 times higher for employees who already had such an episode compared with employees without sickness absence. The odds of recurrence further increased to 3.32 and 4.43 when employees already had two and three prior episodes, respectively. We found that the number of episodes in the past 2 years contributed to the predictability of episodes of sickness absence. Therefore, general practitioners and occupational health providers can identify employees who are likely to take more than average sickness absence with the sickness absence data of the past 2 years.

It is important to identify employees with an increased risk of sickness absence, because preventive occupational health interventions were found to be more effective in

### Table 2. Predictability of days of sickness absence

<table>
<thead>
<tr>
<th>History</th>
<th>2008</th>
<th>2007</th>
<th>2006</th>
<th>2005</th>
</tr>
</thead>
<tbody>
<tr>
<td>Past year</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Coefficient (99% CI)</td>
<td>0.06 (0.04–0.08)**</td>
<td>0.11 (0.07–0.14)**</td>
<td>0.08 (0.05–0.10)**</td>
<td>0.05 (0.02–0.08)**</td>
</tr>
<tr>
<td>Nagelkerke’s pseudo $R^2%$</td>
<td>9%</td>
<td>15%</td>
<td>10%</td>
<td>4%</td>
</tr>
<tr>
<td>Past 2 years</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Coefficient (99% CI)</td>
<td>0.05 (0.03–0.08)**</td>
<td>0.10 (0.06–0.13)**</td>
<td>0.07 (0.04–0.10)**</td>
<td>n.a.</td>
</tr>
<tr>
<td>Coefficient (99% CI)</td>
<td>0.02 (–0.01–0.04)</td>
<td>0.02 (–0.01–0.04)</td>
<td>0.03 (–0.00–0.06)</td>
<td>n.a.</td>
</tr>
<tr>
<td>Nagelkerke’s pseudo $R^2%$</td>
<td>9%</td>
<td>16%</td>
<td>10%</td>
<td>n.a.</td>
</tr>
<tr>
<td>Past 3 years</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Coefficient (99% CI)</td>
<td>0.05 (0.03–0.08)**</td>
<td>0.09 (0.06–0.12)**</td>
<td>n.a.</td>
<td>n.a.</td>
</tr>
<tr>
<td>Coefficient (99% CI)</td>
<td>0.01 (–0.02–0.04)</td>
<td>0.01 (–0.02–0.04)</td>
<td>n.a.</td>
<td>n.a.</td>
</tr>
<tr>
<td>Coefficient (99% CI)</td>
<td>0.03 (0.00–0.06)**</td>
<td>0.03 (–0.01–0.06)</td>
<td>n.a.</td>
<td>n.a.</td>
</tr>
<tr>
<td>Nagelkerke’s pseudo $R^2%$</td>
<td>11%</td>
<td>17%</td>
<td>n.a.</td>
<td>n.a.</td>
</tr>
<tr>
<td>Past 4 years</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Coefficient (99% CI)</td>
<td>0.05 (0.02–0.07)**</td>
<td>n.a.</td>
<td>n.a.</td>
<td>n.a.</td>
</tr>
<tr>
<td>Coefficient (99% CI)</td>
<td>0.01 (–0.02–0.04)</td>
<td>n.a.</td>
<td>n.a.</td>
<td>n.a.</td>
</tr>
<tr>
<td>Coefficient (99% CI)</td>
<td>0.03 (0.00–0.06)**</td>
<td>n.a.</td>
<td>n.a.</td>
<td>n.a.</td>
</tr>
<tr>
<td>Coefficient (99% CI)</td>
<td>–0.00 (–0.04–0.03)</td>
<td>n.a.</td>
<td>n.a.</td>
<td>n.a.</td>
</tr>
<tr>
<td>Nagelkerke’s pseudo $R^2%$</td>
<td>10%</td>
<td>n.a.</td>
<td>n.a.</td>
<td>n.a.</td>
</tr>
</tbody>
</table>

The table shows the coefficients of ordinal regression reflecting the strength of the association with positive coefficients indicating that the probability of being in groups with higher sickness absence increases with the number of past days of sickness absence per 10 days, controlling for age and gender. The Nagelkerke’s pseudo $R^2\%$ reflects the extent to which days of sickness absence can be predicted. **$P < 0.01$; CI, confidence interval; n.a. = not available.
employees with a high risk of sickness absence [21–24]. The data of health checks and pre-placement assessments can be used to identify employees who have an increased risk of sickness absence [6,16,25]. However, health checks are time consuming and should assess an employee’s health risks in work, not the risk of sickness absence. Besides, it was reported that extended health checks did not predict sickness absence better than health rated by one single question [16].

Employees with an increased risk of sickness absence can be identified in a less time-consuming way by questionnaire [20,26–28]. However, questionnaire surveys frequently have moderate to low response rates [21]. Another disadvantage of questionnaire surveys is that healthy employees are more likely to respond than employees with health complaints, a phenomenon known as the ‘healthy volunteer effect’. These disadvantages are bypassed by using recorded sickness absence data to identify the employees who have an above average sickness absence. The sickness absence register data also provide information about the frequency of sickness absence, while health checks and questionnaires predominantly assess the risk of long-term sickness absence. Not only should employees with an increased risk of long-term sickness absence be considered for referral to occupational health interventions and programmes aimed at preventing sickness absence but also those who frequently take sickness absence [29,30].

In conclusion, sickness absence register data can be used to identify employees who are likely to have an above average sickness absence. The sickness absence data of the past 2 years predicted future sickness absence and it was not necessary to go further back in an employee’s history of sickness absence. More research is required to develop prognostic models that select employees for occupational health programmes based on the history of sickness absence in the past 2 years and validate these prognostic models in different populations.

### Key points
- Days of sickness absence in the past year predict future days of sickness absence.
- Episodes of sickness absence in the past 2 years predict future episodes of sickness absence.
- Sickness absence data of the past 2 years helps to identify employees who are likely to have an above average sickness absence.

### Conflicts of interest
None declared.

### References