

Risk reclassification analysis investigating the added value of fatigue to sickness absence predictions

Corné A. M. Roelen · Ute Bültmann ·
Johan W. Groothoff · Jos W. R. Twisk ·
Martijn W. Heymans

Received: 4 October 2014 / Accepted: 4 February 2015 / Published online: 22 February 2015
© The Author(s) 2015. This article is published with open access at Springerlink.com

Abstract

Background Prognostic models including age, self-rated health and prior sickness absence (SA) have been found to predict high (≥ 30) SA days and high (≥ 3) SA episodes during 1-year follow-up. More predictors of high SA are needed to improve these SA prognostic models. The purpose of this study was to investigate fatigue as new predictor in SA prognostic models by using risk reclassification methods and measures.

Methods This was a prospective cohort study with 1-year follow-up of 1,137 office workers. Fatigue was measured at baseline with the 20-item checklist individual strength and added to the existing SA prognostic models. SA days and episodes during 1-year follow-up were retrieved from an occupational health service register. The added value of fatigue was investigated with Net Reclassification Index (NRI) and integrated discrimination improvement (IDI) measures.

Results In total, 579 (51 %) office workers had complete data for analysis. Fatigue was prospectively associated with both high SA days and episodes. The NRI revealed that adding fatigue to the SA days model correctly reclassified workers with high SA days, but incorrectly reclassified workers without high SA days. The IDI indicated

no improvement in risk discrimination by the SA days model. Both NRI and IDI showed that the prognostic model predicting high SA episodes did not improve when fatigue was added as predictor variable.

Conclusion In the present study, fatigue increased false-positive rates which may reduce the cost-effectiveness of interventions for preventing SA.

Keywords Absenteeism · Prediction models · Reclassification table · Risk distribution · Risk stratification · Sick leave

Introduction

Sickness absence (SA) is an increasing problem in developed economies. The Organization for Economic Cooperation and Development reported that countries spend on average 2 % of their gross domestic product (GDP) and 10 % of their social expenditures on SA and disability benefits (OECD 2011). When off work due to sickness, the probability of resuming work decreases with increasing SA duration (Labriola 2008; Lund et al. 2008). Eventually, long-term SA leads to disability pension which excludes workers from the workplace and marginalizes them from the labor market. Therefore, it is important to identify high-risk workers before they report sick, so that they can be invited for interventions aimed at preventing SA (Taimela et al. 2008a; Kant et al. 2008).

Recently, two prognostic models for identifying workers at risk of high SA were developed in a sample of Dutch hospital workers (Roelen et al. 2013a) and validated in Dutch office workers (Roelen et al. 2013b) and Danish eldercare workers (Roelen et al. 2014a). The prognostic model predicting high SA days (i.e., ≥ 30 cumulated days

C. A. M. Roelen · J. W. R. Twisk · M. W. Heymans
Department of Epidemiology and Biostatistics, VU University
Medical Center, VU University, Amsterdam, The Netherlands

C. A. M. Roelen · U. Bültmann · J. W. Groothoff
Department of Health Sciences, University Medical Center
Groningen, University of Groningen, Groningen,
The Netherlands

C. A. M. Roelen (✉)
ArboNed, PO Box 158, 8000 AD Zwolle, The Netherlands
e-mail: corne.roelen@arbo ned.nl

during 1-year follow-up) showed fair performance in hospital workers, but poor performance at external validation. The prognostic model predicting high SA episodes (i.e., ≥ 3 episodes during 1-year follow-up) showed good performance in the development setting and maintained fair performance at external validation. It was concluded that more predictors of SA are needed to improve the SA prognostic models, particularly the model predicting high SA days.

Fatigue is a common symptom of ill-health, ranking third in prevalence after back pain and muscular aches in working populations (Parent-Thirion et al. 2012). Wikman et al. (2005) reported that fatigue was the most prevalent indicator of morbidity in the Swedish workforce. Several studies have already shown that fatigue is associated with future SA (Janssen et al. 2003; Bültmann et al. 2005, 2013; Åkerstedt et al. 2007; Roelen et al. 2014b). However, we need more research to investigate whether or not fatigue should be added as predictor to the SA prognostic models.

Traditionally, the added value of a new predictor to existing prognostic models is investigated by changes in the area under the receiver operating characteristic curve or *c*-statistic (Steyerberg et al. 2010). However, only very strong predictors can increase the performance of well-predicting prognostic models (Pepe et al. 2004; Janes et al. 2008; Cook 2007, 2008). The Net Reclassification Index (NRI) has been introduced as novel measure to assess the added value of predictor variables (Pencina et al. 2008; Steyerberg et al. 2010; Cook and Paynter 2011; Sundström et al. 2011). The NRI summarizes reclassifications of subjects when the new predictor is added. Subjects who develop the outcome are correctly reclassified when they move up into a higher risk category, and subjects who do not develop the outcome are correctly reclassified when they move down into a lower risk category. A disadvantage is that the NRI heavily depends on the thresholds of risk stratifications (Sundström et al. 2011). Pencina et al. (2011) defined an NRI without using risk categories, but this category-free NRI variant has been criticized for its high rates of false-positive conclusions (Pepe et al. 2014). The integrated discrimination improvement (IDI) is an alternative category-free measure to quantify risk discrimination improvement (Pencina et al. 2008; Steyerberg et al. 2010; Sundström et al. 2011).

Empirical evaluations of the literature showed that reclassification methods and measures are often applied inappropriately, for example, to assess the performance of new prognostic models that differ from the previously established models (Tzoulaki et al. 2011; Bouwmeester et al. 2012). In addition, the reclassification measures NRI and IDI are frequently misinterpreted (Kerr et al. 2014; Leening et al. 2014a). The objective of the present study was to introduce risk reclassification and related NRI and IDI measures in an occupational health context by

investigating the added value of fatigue to the existing SA prognostic models. Reclassification analysis can be a key method for guiding decisions in occupational health care, because it presents the distribution of risks in the population and classifies subjects into relevant risk categories.

Methods

To illustrate risk reclassification, we used the health check data from a previously described prospective cohort study of 1,137 office workers (Roelen et al. 2013b, 2014b). Predictor variables were measured by health check questionnaires administered in November 2006. SA in 2007 was retrieved from an occupational health service (OHS) register. The Medical Ethics Committee of the University Medical Center Groningen granted ethical clearance for the study.

Baseline predictor variables

The health check questionnaire asked workers to rate their general health in categories ‘excellent’ = 4, ‘good’ = 3, ‘fair’ = 2, and ‘poor’ = 1 (Ware et al. 2002). Self-rated health (SRH) is widely used in health research and has been associated with various morbidity and mortality measures (Halford et al. 2012). The health check questionnaire measured fatigue with the checklist individual strength (CIS) consisting of 20 statements (Cronbach’s $\alpha = 0.92$) on the severity of fatigue and its functional impact. The statements were scored on a seven-point scale ranging from 1 ‘fully agree’ to 7 ‘fully disagree’ amounting to a sum score ranging between 20 and 140, with higher scores reflecting higher levels of fatigue. The CIS was chosen because it has been validated for measuring fatigue in working populations (Beurskens et al. 2000; De Vries et al. 2003). Previous research has shown that the CIS has high internal consistency and good content validity for measuring fatigue (Beurskens et al. 2000; Dittner et al. 2004; Shahid et al. 2010). Convergent validity was satisfactory as reflected in correlations between the CIS and the Fatigue Assessment Scale (Pearson’s correlation coefficient $r = 0.90$; Michielsen et al. 2002), Fatigue Scale ($r = 0.87$; Chalder et al. 1993), and Maslach Burnout Inventory ($r = 0.71$; Maslach and Jackson 1986).

Age at baseline and SA in the 2 years prior to baseline were retrieved from the OHS register. SA was defined as temporary paid leave from work due to any (i.e., work-related and non-work-related) injury or illness and was recorded in the OHS register from the first SA day until return to work. The calendar days between the first and last SA day were counted as SA days, regardless of whether these were work days. SA days and episodes in the 2 years

prior to baseline were accumulated for the predictor variables prior SA days and prior SA episodes, respectively (Roelen et al. 2013a, b). In agreement with Dutch SA insurance policies, SA episodes with less than 28 days worked between them were regarded as one episode.

Outcome at 1-year follow-up

SA days and episodes in 2007 were retrieved from the OHS register. In previous SA prognostic studies, high SA days was defined as ≥ 30 cumulated (i.e., not necessarily consecutive) SA days and high SA episodes as ≥ 3 SA episodes during 1-year follow-up (Roelen et al. 2013a, b, 2014a). We adopted the same definitions for high SA days and episodes to ensure that the prognostic models in the present study did not differ from the established SA prognostic models.

Statistical analysis

Descriptive statistics and logistic regression analysis were done with IBM SPSS Statistics for Windows, version 20.0 (IBM Corp. Armonk, NY, released 2011). Reclassification analyses were performed in R (Project for Statistical Computing) using the regression modeling strategies (rms) package (Harrell 2013) for calculating NRI and the predictABEL package (Kundu et al. 2011) for calculating IDI. Reclassification after adding fatigue was presented in reclassification tables (Cook 2008; Janes et al. 2008). It is recommended to distinguish between subjects with and without events in reclassification analysis (Pencina et al. 2008; Pepe 2011; Kerr et al. 2014; Leening et al. 2014a). Therefore, we summarized the reclassification of workers with high SA in the NRI for events (NRIe) and reclassification of workers without high SA in the NRI for nonevents (NRIne). $NRIe = P(\text{up}|\text{event}) - P(\text{down}|\text{event})$ representing the net proportion of subjects with events assigned to a higher risk category. $NRIne = P(\text{down}|\text{nonevent}) - P(\text{up}|\text{nonevent})$ representing the net proportion of subjects without events assigned to a lower risk category. The NRI varies between -100 and 100 %; positive values reflect improved classification, and negative values worsened classification.

Risk reclassification analysis requires calibrated models (Pepe and Janes 2011; Leening et al. 2014b). Calibration of the SA prognostic models with and without fatigue was addressed with the Hosmer–Lemeshow (H–L) goodness-of-fit test; adequate calibration was concluded for H–L $p \geq 0.05$ (Steyerberg et al. 2010). After investigating the calibration of the SA prognostic models, we considered the models' ability to stratify the population into risk categories. The risk categories should be clinically relevant in the sense that changing categories implies that subjects receive different treatments or interventions. As far as we know, relevant risk categories or risk thresholds have not

yet been defined in SA research. Therefore, we chose two data-driven risk thresholds (10 and 20 % risk of high SA) based on previously reported risk distributions (Roelen et al. 2013b). In addition, we calculated the IDI to evaluate the improvement of the models' ability to discriminate between high-risk and low-risk subjects without using categories. The IDI reflects the change in discrimination slope (i.e., difference between the mean estimated risk for cases and non-cases) of the model with the new predictor compared to the model with only the established predictors. IDI (range -100 to 100 %) represents overall risk discrimination improvement, but its magnitude is hard to interpret.

Results

A total of 633 (56 %) office workers participated in the health checks. Non-participant analysis showed that participants were older (mean age = 44.5, standard deviation [SD] = 9.3 years) than non-participants (39.0, SD = 9.4 years; t test $p < 0.01$). Sixty-two percent of participants were men as compared to 68 % of non-participants (Chi-square $p = 0.04$). Fifteen percent of participants had high SA episodes compared to 22 % of non-participants (Chi-square $p < 0.01$). The proportions of office workers with high SA days did not differ (Chi-square $p = 0.45$) between participants and non-participants.

Fifty-four participants (8 %) had missing data on predictor and/or outcome variables, leaving 579 office workers with complete data for reclassification analysis (Table 1). Baseline fatigue was moderately correlated (Pearson $r = 0.44$; $p < 0.01$), but not collinear with SRH.

Prognostic model for high SA days

Fifty-nine (10 %) office workers had high SA days during 1-year follow-up. Prior SA days, SRH and fatigue were significantly associated with high SA days, while age was not (Table 2). Calibration was adequate for the model without fatigue (H–L $\chi^2 = 13.8$, $df = 8$; $p = 0.09$) and the model with fatigue (H–L $\chi^2 = 4.7$, $df = 8$; $p = 0.79$). The lower H–L model Chi-square indicated that the risks predicted by the prognostic model with fatigue were more in agreement with the observed frequencies of high SA days.

At 10 % risk threshold, 84 workers (15 %) were reclassified when fatigue was added to the model. Seven workers with high SA days correctly moved up to the high-risk category, but four incorrectly moved down to the low-risk category. Table 3 shows that NRIe was 5.09 % and not significant ($p = 0.91$). Of the workers without high SA days, 73 were reclassified of whom 49 incorrectly. NRIne was

Table 1 Study population characteristics ($N = 579$)

Variable	Mean	SD	<i>N</i>	%
Age (years)	44.4	9.3		
Gender				
Men			361	62
Women			218	38
Work (h/week)	34.5	8.0		
Duration of employment (years)	3.2	1.3		
Self-rated health				
Excellent			143	25
Good			330	57
Fair			103	18
Poor			3	1
Prior sickness absence days				
0			118	20
1–7			143	25
8–14			104	18
15–30			94	16
31–60			60	10
>60			60	10
Prior sickness absence episodes				
0			118	20
1			130	22
2			115	20
3			74	13
4			64	11
>4			78	13
Fatigue (range 2–140)	51.2	21.0		

SD standard deviation

–4.81 % reflecting significant ($p < 0.01$) worsening of the classification of workers without high SA days when fatigue was added (Table 3). At a more specific threshold risk of 20 %, NRIe was –1.69 % ($p = 0.32$) and NRIne was –1.15 % ($p = 0.04$). IDI was 0.25 % (95 % CI –0.25 to 0.75 %) indicating that fatigue did not significantly ($p = 0.33$) improve risk discrimination.

Table 2 Prospective associations with sickness absence (SA) days and episodes

Predictor	High (≥ 30) SA days			High (≥ 3) SA episodes		
	OR	95 % CI	<i>p</i>	OR	95 % CI	<i>p</i>
Age	0.99 ^a	0.74–1.32	$p = 0.92$	0.92 ^a	0.70–1.21	$p = 0.56$
Prior SA	1.02 ^a	1.00–1.08	$p = 0.02$	1.60	1.41–1.82	$p < 0.01$
Self-rated health	0.59	0.39–0.90	$p = 0.01$	0.56	0.37–0.84	$p < 0.01$
Fatigue	1.16 ^a	1.03–1.31	$p = 0.02$	1.14 ^a	1.01–1.28	$p = 0.03$

The table shows odds ratios (OR) and 95 % confidence intervals (CI) for univariate associations between predictor variables and sickness absence

^a Per ten-point increase

Prognostic model for high SA episodes

Sixty-five (11 %) workers had high SA episodes at follow-up. Prior SA episodes, SRH and fatigue were significantly associated with high SA episodes, whereas age was not (Table 2). The prognostic models with fatigue (H–L $\chi^2 = 3.0$, $df = 8$; $p = 0.94$) and without fatigue (H–L $\chi^2 = 2.1$, $df = 8$; $p = 0.98$) were both well calibrated.

At a risk threshold of 10 %, seven workers (1 %) were reclassified when fatigue was added to the model. None of the workers with high SA episodes was reclassified (Table 4). Consequently, NRIe was not available. Of workers without high SA episodes, five were correctly reclassified as low risk and two were incorrectly reclassified as high risk; NRIne was 0.58 % and non-significant ($p = 0.26$). At a more specific 20 % risk threshold, four workers were reclassified after adding fatigue, although all incorrectly moved up to high risk. NRIne was –0.78 % and marginally significant ($p = 0.05$). IDI = 0.11 % (95 % CI –0.18 to 0.41 %) indicated that fatigue did not significantly ($p = 0.45$) improve risk discrimination.

Discussion

Fatigue is a core symptom of many medical conditions and an important indicator of morbidity in working populations. The present study confirmed that fatigue was prospectively associated with high sickness absence (SA) days and episodes during 1-year follow-up. However, reclassification analysis showed that fatigue did not improve risk predictions for office workers with high SA and worsened risk predictions for office workers without high SA.

Although reclassification analysis has become very popular, the NRI is often misinterpreted as ‘percentage of the population reclassified’ (Kerr et al. 2014; Leening et al. 2014a). In addition, the NRI equally values the consequences of false-negative and false-positive misclassifications. These problems were overcome in the current study by presenting NRI for events (NRIe) and nonevents

Table 3 Reclassification table for sickness absence (SA) days

Risk without fatigue	Risk with fatigue		Risk without fatigue	Risk with fatigue	
Events	≤10 %	>10 %	Events	≤20 %	>20 %
≤10 %	21	7	≤20 %	57	0
>10 %	4	27	>20 %	1	1
<i>NR</i> Ie (95 % CI)	5.09 % (−5.93 to 16.10 %)		<i>NR</i> Ie (95 % CI)	−1.69 % (−5.02 to 1.63 %)	
Nonevents	≤10 %	>10 %	Nonevents	≤20 %	>20 %
≤10 %	318	49	≤20 %	501	7
>10 %	24	129	>20 %	1	11
<i>NR</i> Ine (95 % CI)	−4.81 % (−8.03 to −1.59 %)		<i>NR</i> Ine (95 % CI)	−1.15 % (−2.22 to −0.09 %)	

*NR*Ie Net Reclassification Index for events, *NR*Ine Net Reclassification Index for nonevents, *CI* confidence interval

Table 4 Reclassification table for sickness absence (SA) episodes

Risk without fatigue	Risk with fatigue		Risk without fatigue	Risk with fatigue	
Events	≤10 %	>10 %	Events	≤20 %	>20 %
≤10 %	16	0	≤20 %	33	0
>10 %	0	49	>20 %	0	32
<i>NR</i> Ie (95 % CI)	n.a.		<i>NR</i> Ie (95 % CI)	n.a.	
Nonevents	≤10 %	>10 %	Nonevents	≤20 %	>20 %
≤10 %	355	2	≤20 %	462	4
>10 %	5	152	>20 %	0	48
<i>NR</i> Ine (95 % CI)	0.58 % (−0.43 to 1.59 %)		<i>NR</i> Ine (95 % CI)	−0.78 % (−1.54 to −0.00 %)	

*NR*Ie Net Reclassification Index for events, *NR*Ine Net Reclassification Index for nonevents, *CI* confidence interval, *n.a.* not available

(*NR*Ine). Another problem is that data-driven risk thresholds may spuriously inflate the *NRI* (Sundström et al. 2011; Hilden and Gerds 2014). As there are no clinically relevant risk thresholds for SA, we dealt with this problem by analyzing reclassification at different (10 % and 20 %) risk thresholds. For the SA days model, *NR*Ie was not significant, whereas *NR*Ine was significant. For the SA episodes model, *NR*Ine was on the verge of significance at the 20 % risk threshold.

However, the fact that *NR*Ine was negative indicated worsened rather than improved risk classification of office workers without high SA. Workers without high SA days incorrectly moved up from the low-risk category to the high-risk category when fatigue was added as predictor variable. In other words, adding fatigue increased false-positive predictions of high SA. High false-positive rates can be problematic when resources

are limited or when the burden and costs of interventions are high. Furthermore, Taimela et al. (2008b) reported that preventive consultations cost-effectively reduced SA only in high-risk workers and not in workers with a moderate or low SA risk. Hence, using fatigue as predictor not only increases unnecessary utilization of preventive interventions, but may also reduce cost-effectiveness.

Nowadays, there are many fatigue instruments because fatigue is recognized as a major symptom of clinical conditions. All instruments are self-report measures, and answers on self-report measures are driven by the respondent's interpretation as well as factors such as personal dispositions, mood, expectations, and previous experiences. Furthermore, an instrument developed to measure fatigue in one patient group may not apply to other groups when fatigue experiences depend on the clinical condition. The *CIS* was previously shown to have good psychometric properties for measuring fatigue in working populations (Beurskens et al. 2000; Bültmann et al. 2002; De Vries et al. 2003). However, associations between fatigue and high SA days or episodes may differ when fatigue is measured with other instruments. Before drawing definite conclusions of the added value of fatigue to SA predictions, we need more studies investigating fatigue with different instruments or even a combination of instruments, because it is doubtful whether one instrument can capture fatigue, given the wide range of mechanisms underlying fatigue and the differing manifestations of fatigue (Dittner et al. 2004).

Study strengths and limitations

The study population was limited to 1,137 office workers employed in one company, which is an advantage because it excludes SA variability due to differences in work conditions, work environment, and organizational policies and practices (Virtanen et al. 2008). The SA percentage (3.6 %) was slightly lower than in the Dutch workforce (4.2 %) in 2007, whereas the mean SA frequency in 2007 was 1.1 episodes in both the study population and the Dutch workforce (Statistics Netherlands 2014). However, this does not implicate that the results can be generalized to other working populations.

Another limitation of the study is that 56 % of the workers participated in the health checks. Janssen et al. (2003) reported a total *CIS* score of 53.4 in a heterogeneous sample ($N = 7,495$) of non-sicklisted workers participating in the Dutch Maastricht Cohort Study. The mean *CIS* score in the present study was 51.2 reflecting lower levels of fatigue (one sample t test $p = 0.02$). This might indicate that the office workers in our study were healthier. Such a 'healthy volunteer effect' (Etter and Perneger 1997; Fromm et al.

1999) may have weakened associations between fatigue and high SA, consequently reducing the added value of fatigue to SA predictions.

Conclusion

In the present study, fatigue increased false-positive rates which may reduce the cost-effectiveness of interventions aimed at preventing SA. The findings illustrate that we have to carefully consider the added value of potential predictors rather than strive for the most comprehensive prognostic models, even when we want to predict a complex multifactorial outcome such as SA. When applied and interpreted appropriately, risk reclassification can be used to gauge the added value of new predictors to established prognostic models.

Conflict of interest The authors declare that they have no conflict of interest.

Open Access This article is distributed under the terms of the Creative Commons Attribution License which permits any use, distribution, and reproduction in any medium, provided the original author(s) and the source are credited.

References

- Åkerstedt T, Kecklund G, Alfredsson L, Selén J (2007) Predicting long-term sickness absence from sleep and fatigue. *J Sleep Res* 16:341–345
- Beurskens AJ, Bültmann U, Kant IJ, Vercoulen JH, Bleijenberg G, Swaen GM (2000) Fatigue among working people: validity of a questionnaire measure. *Occup Environ Med* 57:353–357
- Bouwmeester W, Zuithoff NPA, Mallett S, Geerlings MI, Vergouwe Y, Steyerberg EW et al (2012) Reporting and methods in clinical prediction research: a systematic review. *Plos Med* 9:e1001221
- Bültmann U, Kant I, Kasl SV, Beurskens AJ, van den Brandt PA (2002) Fatigue and psychological distress in the working population: psychometrics, prevalence and correlates. *J Psychosom Res* 52:445–452
- Bültmann U, Huibers MJ, van Amelsvoort LP, Janssen PPM, Schröer CAP (2005) Psychological distress, fatigue and long-term sickness absence: prospective results from the Maastricht Cohort Study. *J Occup Environ Med* 47:941–947
- Bültmann U, Nielsen MB, Madsen IEH, Burr H, Rugulies R (2013) Sleep disturbances and fatigue: independent predictors of sickness absence? A prospective study among 6,538 employees. *Eur J Public Health* 23:123–128
- Chalder T, Berelowitz G, Pawlikowska T, Watts L, Wessely S, Wright D et al (1993) Development of a fatigue scale. *J Psychosom Res* 37:147–153
- Cook NR (2007) Use and misuse of the receiver operating characteristic curve in risk prediction. *Circulation* 115:928–935
- Cook NR (2008) Statistical evaluation of prognostic versus diagnostic models: beyond the ROC-curve. *Clin Chem* 54:17–23
- Cook NR, Paynter NP (2011) Performance of reclassification statistics in comparing risk prediction models. *Biom J* 53:237–258
- De Vries J, Michielsen HJ, van Heck GL (2003) Assessment of fatigue among working people: a comparison of six questionnaires. *Occup Environ Med* 60(suppl 1):i10–i15
- Dittner AJ, Wessely SC, Brown RG (2004) The assessment of fatigue. A practical guide for clinicians and researchers. *J Psychosom Res* 56:157–170
- Etter JF, Perneger TV (1997) Analysis of non-response bias in a mailed health survey. *J Clin Epidemiol* 50:1123–1128
- From P, Melamed S, Kristahl-Boneh E, Benbassat J, Ribak J (1999) Healthy volunteer effect in industrial workers. *J Clin Epidemiol* 52:731–735
- Halford C, Wallman T, Welin L, Rosengren A, Bardel A, Johansson S et al (2012) Effects of self-rated health on sick leave, disability pension, hospital admissions and mortality. A population-based longitudinal study of nearly 15,000 observations among Swedish women and men. *BMC Public Health* 12:1103
- Harrell FE (2013) Regression modeling strategies. <http://cran.r-project.org/web/packages/rms>. Accessed 30 Nov 2014
- Hilden J, Gerds TA (2014) A note on the evaluation of novel biomarkers: do not rely on integrated discrimination improvement and Net Reclassification Index. *Stat Med*. doi:10.1002/sim.5804
- Janes H, Pepe MS, Gu W (2008) Assessing the value of risk predictions by using risk stratification tables. *Ann Intern Med* 149:751–760
- Janssen N, Kant IJ, Swaen GHM, Janssen PP, Schröer CA (2003) Fatigue as predictor of sickness absence: results from the Maastricht Cohort Study. *Occup Environ Med* 60(Suppl1):i71–i76
- Kant IJ, Jansen NW, van Amelsvoort LG, van Leusden R, Berkouwer A (2008) Structured early consultation with the occupational physician reduces sickness absence among office workers at high risk for long-term sickness absence: a randomized controlled trial. *J Occup Rehabil* 18:79–86
- Kerr KF, Wang Z, Janes H, McClelland RL, Psaty BM, Pepe MS (2014) Net reclassification indices for evaluation risk prediction instruments: a critical review. *Epidemiology* 25:114–121
- Kundu S, Aulchenko YS, van Duijn CM, Janssens AC (2011) PredictABEL: an R package for the assessment of risk prediction models. *Eur J Epidemiol* 26:261–264
- Labriola M (2008) Conceptual framework of sickness absence and return to work, focusing on both the individual and contextual level. *Work* 30:377–387
- Leening MJ, Vedder MM, Witteman JC, Pencina MJ, Steyerberg EW (2014a) Net reclassification improvement: computation, interpretation, and controversies: a literature review and clinician's guide. *Ann Intern Med* 160:122–131
- Leening MJ, Steyerberg EW, van Calster B, D'Agostinho RB, Pencina MJ (2014b) Net reclassification improvement and integrated discrimination improvement require calibrated models: relevance from a marker and model perspective. *Stat Med* 33:3415–3418
- Lund T, Kivimäki M, Labriola M, Villadsen E, Christensen KB (2008) Using administrative sickness absence data as a marker of future disability pension: the prospective DREAM study of Danish private sector employees. *Occup Environ Med* 65:28–31
- Maslach C, Jackson SE (1986) The Maslach Burnout Inventory. Manual, 2nd edn. Sage, Palo Alto
- Michielsen HJ, De Vries J, van Heck GL (2002) Psychometric qualities of a brief self-rated fatigue measure: the fatigue assessment scale (FAS). *J Psychosom Res* 54:45–52
- OECD (2011) Society at a glance 2011: OECD social indicators. http://www.oecd-ilibrary.org/social-issues-migration-health/society-at-a-glance-2011_soc_glance-2011-en. Accessed 30 Nov 2014
- Parent-Thirion A, Vermeylen G, van Houten G, Lyly-Yrjänäinen M, Bileta I, Cabrera J (2012) Fifth European survey on working conditions. <http://www.eurofound.europa.eu/publications/htmlfiles/efl182.htm>. Accessed 30 Nov 2014

- Pencina MJ, d'Agostinho RB Sr, d'Agostinho RB Jr, Vasan RS (2008) Evaluating the added predictive ability of a new marker: from area under the ROC curve to reclassification and beyond. *Stat Med* 27:157–172
- Pencina MJ, d'Agostinho RB Sr, Steyerberg EW (2011) Extensions of net reclassification improvement calculations to measure usefulness of new biomarkers. *Stat Med* 30:11–21
- Pepe MS (2011) Problems with risk reclassification methods for evaluating prediction models. *Am J Epidemiol* 173:1327–1335
- Pepe MS, Janes H (2011) Reporting standards are needed for evaluations of risk reclassification. *Int J Epidemiol* 40:1106–1108
- Pepe MS, Janes H, Longton G, Leisenring W, Newcomb P (2004) Limitations of the odds ratio in gauging the performance of a diagnostic, prognostic, or screening marker. *Am J Epidemiol* 159:882–890
- Pepe, Janse H, Li CI (2014) Net risk reclassification *p* values: valid or misleading? *J Natl Cancer Inst* 106:dju041
- Roelen CA, van Rhenen W, Bültmann U, Groothoff JW, van der Klink JJ, Heymans MW (2013a) The development and validation of two prediction models to identify employees with high sickness absence. *Eur J Public Health* 23:128–133
- Roelen CA, Bültmann U, van Rhenen W, van der Klink JJ, Twisk JWR, Heymans MW (2013b) External validation of two prediction models identifying employees at risk of high sickness absence: cohort study with 1-year follow-up. *BMC Public Health* 13:105
- Roelen CA, Stapelfeldt CM, Heymans MW, van Rhenen W, Labriola M, Nielsen CV, et al (2014a) Cross-national validation of prognostic models predicting sickness absence and the added value of work environment variables. *J Occup Rehabil*. doi:10.1007/s10926-014-9536-3
- Roelen CA, van Rhenen W, Groothoff JW, van der Klink JJJ, Bültmann U (2014b) Prolonged fatigue is associated with sickness absence in men but not in women: prospective study with 1-year follow-up of white collar employees. *Int Arch Occup Environ Health* 24:101–105
- Shahid A, Shen J, Shapiro CM (2010) Measurement of sleepiness and fatigue. *J Psychosom Res* 69:81–89
- Statistics Netherlands (2014) New and revised StatLine tables. <http://www.cbs.nl/en-GB/menu/themas/gezondheid-welzijn/cijfers/nieuw/default.htm>. Accessed 29 Nov 2014
- Steyerberg EW, Vickers AJ, Cook NR, Gerds T, Gonen M, Obuchowski M et al (2010) Assessing the performance of prediction models: a framework for traditional and novel measures. *Epidemiology* 21:128–138
- Sundström J, Byberg L, Gedeberg R, Michaëlsson K, Berglund L (2011) Useful tests of usefulness of new risk factors: tools for assessing reclassification and discrimination. *Scand J Public Health* 39:439–441
- Taimela S, Malmivaara A, Justén S, Läärä E, Sintonen H, Tiekso J et al (2008a) The effectiveness of two occupational health intervention programmes in reducing sickness absence among employees at risk. Two randomised controlled trials. *Occup Environ Med* 65:236–241
- Taimela S, Justén S, Aronen P, Sintonen H, Läärä E, Malmivaara A et al (2008b) An occupational health intervention programme for workers at high risk for sickness absence. Cost effectiveness analysis based on a randomised controlled trial. *Occup Environ Med* 65:242–248
- Tzoulaki I, Liberopoulos G, Ioannidis JPA (2011) Use of reclassification for assessment of improved prediction: an empirical evaluation. *Int J Epidemiol* 40:1094–1105
- Virtanen P, Siukola A, Luukkaala T, Savinainen M, Arola H, Nygård CH et al (2008) Sick leaves in four factories—do characteristics of employees and work conditions explain differences in sickness absence between workplaces? *Scand J Work Environ Health* 34:260–266
- Ware JE, Kosinski MA, Turner-Bowker DM, Gandek B (2002) User's manual for the SF-12v2® health survey. Quality Metric Incorporated, Lincoln
- Wikman A, Marklund S, Alexanderson K (2005) Illness, disease, and sickness absence: an empirical test of differences between concepts of ill health. *J Epidemiol Community Health* 59:450–455