Aalborg Universitet

AALBORG UNIVERSITY

Optimal Cut-Off Points for the Short-Negative Act Questionnaire and Their Association with Depressive Symptoms and Diagnosis of Depression

Conway, Paul Maurice; Høgh, Annie; Nabe-Nielsen, Kirsten; Brødsgaard Grynderup, Mathias; Mikkelsen, Eva Gemzøe; Persson, Roger; Rugulies, Reiner; Bonde, Jens Peter Ellekilde; Francioli, Laura; Hansen, Åse Marie

BOHS
The Chartered Society for
Worker Health Protection

OXFORD

Original Article

Optimal Cut-Off Points for the Short-Negative Act Questionnaire and Their Association with Depressive Symptoms and Diagnosis of Depression

Paul Maurice Conway[1],*, **Annie Høgh[1]**, **Kirsten Nabe-Nielsen[2]**,
Matias Brødsgaard Grynderup[2], **Eva Gemzøe Mikkelsen[3]**, **Roger Persson[4,5]**,
Reiner Rugulies[1,2,6], **Jens Peter Ellekilde Bonde[2,7]**, **Laura Francioli[8]** and
Åse Marie Hansen[2,6]

[1]Department of Psychology, University of Copenhagen, Øster Farimagsgade 2A, 1353 Copenhagen K, Denmark; [2]Department of Public Health, University of Copenhagen, Øster Farimagsgade 5, 1014 Copenhagen K, Denmark; [3]Department of Communication and Psychology, Aalborg University, Nyhavnsgade 14, 9000 Aalborg, Denmark; [4]Department of Psychology, Lund University, Box 213, 221 00, Lund, Sweden; [5]Department of Laboratory Medicine, Division of Occupational and Environmental Medicine, Lund University, Box 117, 221 00, Lund, Sweden; [6]National Research Centre for the Working Environment, Lersø Parkallé 105, 2100, Copenhagen Ø, Denmark; [7]Department of Occupational and Environmental Medicine, Bispebjerg-Frederiksberg University Hospital, Bispebjerg Bakke 23, 2400 Copenhagen NV, Denmark; [8]Independent researcher, H.C. Ørstedsvej 58 ST TV, 1879 Frederiksberg C, Denmark

*Author to whom correspondence should be addressed. Tel: +45-35324935; e-mail: paul.conway@psy.ku.dk

Abstract

Objectives: The behavioural experience method has been extensively used in the literature for the measurement of potential bullying behaviours at work. However, this approach presents limitations when used to classify respondents as targets or non-targets of workplace bullying. Therefore, the present study aimed to: (i) identify optimal cut-off points, reflecting a possible subjectively experienced exposure to occasional and frequent workplace bullying, for the 9-item Short Negative Act Questionnaire (S-NAQ), and (ii) examine the criterion validity of these cut-off points in relation to depressive symptoms and diagnosis of depression.
Methods: The study was based on a sample of 4882 participants from the Danish MODENA cohort study (year 2011), which included both the S-NAQ (score range 9–45) and a one-item measure applying the self-labelling method with a definition to assess occasional and frequent workplace bullying. We employed receiver operating characteristic (ROC) curve analyses to derive the cut-off points for

the S-NAQ. Based on these cut-off points, we created a new S-NAQ variable with three levels of exposure (i.e. 'not exposed', 'first threshold', and 'second threshold') and tested its criterion validity in relation to depressive symptoms (N = 4071) and diagnosis of depression (N = 4844).

Results: The S-NAQ cut-off points obtained were ≥12 and ≥16 when using occasional and frequent bullying as reference standards, respectively. Both cut-off points showed high classification accuracy (area under the curve = 0.89 and 0.93) as well as good sensitivity (84.8% and 88.0%) and specificity (77.4% and 94.7%). In the adjusted linear regression analyses, both the first (B = 0.78, 95% confidence interval [CI] = 0.66–0.90) and the second threshold of exposure (B = 1.65, 95% CI = 1.44–1.86) were significantly associated with depressive symptoms. In the adjusted logistic regression analyses, both the first (odds ratio [OR] = 3.55, 95% CI = 1.98–6.38) and the second threshold of exposure (OR = 5.90, 95% CI = 2.93–11.88) were significantly associated with diagnosis of depression.

Conclusions: The two cut-off points for the S-NAQ identified in this study showed a significant association with both depressive symptoms and diagnosis of depression. However, future prospective studies are needed to establish the predictive validity of the proposed cut-off points.

Keywords: behavioural experience method; depressive symptoms; diagnosis of depression; ROC curve analysis; self-labelling method with a definition; Short Negative Act Questionnaire; workplace bullying

Introduction

There is consistent empirical evidence that being a target of workplace bullying is a risk factor for reduced mental health (Nielsen *et al.*, 2014; Theorell *et al.*, 2015; Verkuil *et al.*, 2015). From the perspective of risk assessment and prevention, such evidence implies the need of adopting valid measures that enable researchers and practitioners to correctly identify targets of bullying. This is especially crucial for accurately estimating the risk to mental health associated with the experience of bullying at work.

To contribute in this regard, in the present study we aim to improve the usability of the Short Negative Act Questionnaire (S-NAQ; Notelaers and Einarsen, 2008), a tool based on the behavioural experience method that measures self-perceived exposure to potential bullying behaviours on a continuous scale. By determining and providing empirically determined cut-off points for the S-NAQ, which reflect the level of exposure to negative acts that corresponds to a possible self-perceived exposure to workplace bullying, the versatility of the S-NAQ will increase as will the possibilities to make decisions regarding the severity of the exposure to the negative acts.

Available approaches to the measurement of workplace bullying

In the literature, there are two main approaches for measuring workplace bullying, i.e. the behavioural experience method and the self-labelling method (Einarsen and Skogstad, 1996; Nielsen *et al.*, 2009, 2010). When using the behavioural experience method, participants are asked to report whether, and how frequently, they have been exposed to a set of negative behaviours over a period

of at least 6 months. There are several behavioural inventories available in the literature (see Nielsen *et al.*, 2011, for a review), each including a different set and number of negative acts, for instance acts that target the individual personally (e.g. slander and social isolation) and/or his/her work situation (e.g. assigning too many or too few work tasks or persistent criticisms of one's work). The most frequently used inventories are the 22-item Negative Act Questionnaire Revised (NAQ-R; Einarsen *et al.*, 2009) and its reduced 9-item version, i.e. the S-NAQ. Contrary to the behavioural experience method, the self-labelling method measures workplace bullying by means of a single-item question asking participants to report how frequently (e.g. 'never', 'now and then', 'monthly', 'weekly', and 'daily') they have experienced bullying over a period of at least 6 months. When using the self-labelling method with a definition, prior to answering the question on bullying, the respondents are presented with a definition that covers aspects that are generally considered as key elements of the phenomenon, including the persistence and frequency of the negative acts and the difficulty to defend oneself against them (Einarsen *et al.*, 2011).

Limitations of the behavioural experience method in identifying targets of workplace bullying

The behavioural experience method is believed to present some advantages compared to the self-labelling method (Nielsen *et al.*, 2010, 2011). One is that the negative acts listed in the inventories are not explicitly referred to as acts of bullying, which may reduce problems of under-reporting potentially occurring when using the self-labelling method. Indeed, with the latter approach, some respondents may

refrain from reporting their exposure to bullying because they may find it difficult to admit to their victimization (Mikkelsen and Einarsen, 2001; Nielsen *et al.*, 2010, 2011). In addition, when presented with a series of negative behaviours, respondents may be reminded about potential acts of bullying that they would not have thought about if administered a single item that does not mention any specific act, as it occurs with the self-labelling method. Finally, the behavioural experience method provides information, not contributed by the self-labelling method, about the specific nature of the negative acts (Nielsen *et al.*, 2010), which might be of use when devising targeted measures against bullying behaviours in work organizations.

Despite these advantages, the behavioural experience method presents two major limitations that may affect its utility in classifying respondents as targets or non-targets of workplace bullying. A first limitation is that the negative acts listed in the behavioural inventories may not necessarily be perceived as bullying by the respondents (Mikkelsen and Einarsen, 2001; Nielsen *et al.*, 2010; Neall and Tuckey, 2014; Burton, 2015; Hewett *et al.*, 2016; Hogh *et al.*, 2016). However, in line with the transactional theory of stress (Lazarus and Folkman, 1984), the appraisal of an encounter as a source of threat is a key element that plays a major role in the aversive nature of such encounter. Capturing a target's perception of being subjected to bullying is thus critical in determining his/her affective response to this experience and the consequent effects on health. Such perception is explicitly acknowledged when using the self-labelling method with a definition, which enquires into whether an employee labels him/herself as a target of bullying. Supporting this perspective, Vie *et al.* (2011) found that self-labelling partially mediated the association between self-reported exposure to negative behaviours and poor health status. More recently, Hewett *et al.* (2016) found that the association between self-reported exposure to frequent negative acts and subjective well-being was stronger among those who labelled themselves as bullied. These findings point to the importance of attributing one's exposure to negative acts to a situation of bullying when estimating the impact that these acts may have on health status.

Contrary to the self-labelling method, the instruments based on the behavioural experience method do not refer to the subjective perception of being a target of bullying. Hence, one needs to make arbitrary choices when deciding which thresholds of exposure to negative behaviours have to be applied in order to distinguish between individual perceiving themselves as targets or non-targets. In the research domain, a common approach when adopting the behavioural experience method is the operational criterion method, which distinguishes between targets and non-targets by applying cut-off points corresponding to a minimum number of frequent negative behaviours reported by the respondents (e.g., Leymann, 1990; Mikkelsen and Einarsen, 2001; Agervold, 2007). However, this approach has been criticized because these cut-off points are decided arbitrarily by the researchers (Notelaers and Einarsen, 2013), making it difficult to establish whether individuals scoring above the applied threshold would perceive the reported behaviours as acts of bullying. Another approach adopted in research is to rely on the continuous scale scores calculated on the items included in behavioural inventories and examine their association with mental health symptoms, e.g. symptoms of depression and anxiety (e.g. Hauge *et al.*, 2010) and of Post-Traumatic Stress Disorder (e.g. Balducci *et al.*, 2011). However, finding a significant association with this approach would only suggest a relationship between higher levels of exposure to negative behaviours and lower mental health; yet, whether these scores reflect exposure to bullying, rather than to other types of exposures (e.g. job stressors), cannot be established. Besides research, the arbitrariness in classifying targets of workplace bullying when employing behavioural inventories may also affect the validity of assessment and intervention measures in the realm of practice, where behavioural instruments are often used to measure bullying within specific organizational contexts.

A second limitation of the behavioural experience method is that it does not capture whether the target has experienced difficulties in defending him/herself against the negative behaviours (Nielsen *et al.*, 2010, 2011). Yet, the inability to defend oneself, which in the literature is often attributed to an imbalance of power between the parties involved (Einarsen *et al.*, 2011), is regarded by many authors as a defining element of workplace bullying (Salin, 2001, 2003, 2015; Nielsen *et al.*, 2010, 2011).

The present study

Given these limitations, there is a need to determine non-arbitrary cut-off points to identify thresholds of exposure to negative behaviours that reflect a possible self-perceived exposure to workplace bullying, while also capturing a key element in the definition of bullying such as the inability of the individuals to defend themselves during the adverse situation. Such thresholds would enable researchers and practitioners to harness the information provided by the instruments based on the behavioural experience method, while at the same time equipping them with empirical cut-off points that they can apply to better classify targets of workplace bullying.

To contribute in this regard, in the present study we aim to provide empirically derived cut-off points for the

S-NAQ, a widely use instrument based on the behavioural experience method. Previously, Notelaers and Einarsen (2013) used receiver operator characteristic (ROC) curve analysis to determine empirically based, as opposed to arbitrary, cut-off points for the NAQ-R (Einarsen *et al.*, 2009). We set out to employ a similar ROC curve analysis approach to identify optimal cut-off points for the S-NAQ. For identifying these cut-off points, we adopted, as reference standards, two levels of exposure to workplace bullying, i.e. occasional and frequent bullying, as measured with the self-labelling method with a definition. The choice of this reference standard is justified by the fact that the self-labelling method, when accompanied by a definition, captures the aforementioned two crucial features of bullying, i.e. the perception of being bullied and the inability to defend oneself against the negative behaviours. We then examined the criterion validity of the new empirically derived S-NAQ cut-off points by investigating their association with depressive symptoms and diagnosis of depression.

Materials and methods

Participants
Participants included in the ROC curve analysis
This sample is composed of participants in the third wave (year 2011) of a cohort study (MODENA; Gullander *et al.*, 2014) that combines two separate Danish cohorts, i.e. the Workplace Bullying and Harassment (WBH) cohort (years 2006–2008) and the PRISME cohort (years 2007–2009). The WBH cohort is a workplace-based study including employees from 60 workplaces in both the public and the private sector in Denmark (see, e.g. Hogh *et al.*, 2012, for more information about the WBH cohort). The PRISME cohort includes hospital and civil-service employees from 468 work units in several public workplaces in the Central Denmark Region (see, e.g. Kolstad *et al.*, 2011 for more information on the PRISME cohort). In the beginning of 2011, all those who participated in 2006 in the WBH cohort and in 2007 in the PRISME cohort were invited to participate in the third-wave MODENA study. A total of 8128 participants were approached in 2011, and a total of 5489 responded (response rate = 67.5%). In all, valid values for both the S-NAQ sum score and the one-item question based on the self-labelling method with a definition were available for a sample of 4882 participants. The main sample socio-demographic characteristics at baseline and the baseline prevalence of workplace bullying according to the self-labelling method are shown in Table 1. Table 2 shows the S-NAQ sum score and the nine S-NAQ items, along with their mean and standard deviation.

Table 1. Distribution of baseline characteristics among participants in the ROC curve analysis (*N* = 4882[a]).

Baseline characteristics	*N*	Column % (valid values)	Mean (SD)
Gender			
Women	3668	75.1	
Men	1214	24.9	
Age (in years)	4882		49.0 (9.6)
Education			
<3 years	1192	26.4	
3–4 years	2675	59.3	
>4 years	641	14.2	
Cohabitation			
Cohabiting	3978	83.4	
Living alone	794	16.6	
Smoking status			
Non-smoker	4215	87.1	
Current smoker	626	12.9	
Sense of coherence	4853		83.2 (11.6)
Exposure to workplace bullying (self-labelling method)			
Never exposed	4629	94.8	
Occasionally exposed	224	4.6	
Frequently exposed	29	0.6	

[a]The total number of participants is lower than 4882 for education, cohabitation, smoking status, and sense of coherence due to the presence of missing values.

Table 2. Mean (standard deviation) of the S-NAQ sum score and the component items in the sample included in the ROC curve analysis (*N* = 4882).

S-NAQ items	Mean (SD)
S-NAQ sum score (continuous) (9–45)	10.89 (2.80)
Someone withholding information which affects your performance (1–5)	1.56 (0.73)
Spreading of gossip and rumours about you (1–5)	1.23 (0.52)
Being ignored or excluded (1–5)	1.15 (0.46)
Having insulting or offensive remarks made about your person, attitudes or your private life (1–5)	1.12 (0.41)
Being shouted at or being target of spontaneous anger (1–5)	1.23 (0.53)
Repeated reminders of your errors or mistakes (1–5)	1.20 (0.47)
Being ignored or facing a hostile reaction when you approach (1–5)	1.19 (0.48)
Persistent criticism of your work and effort (1–5)	1.11 (0.38)
Practical jokes carried out by people you don't get along with (1–5)	1.09 (0.35)

At all study waves, employees were informed that participation was on a voluntary basis and that all data would be treated confidentially. The WBH and MODENA study protocols were approved by the local ethics committee (KF 01 302955 and H-2-2010-119, respectively). The PRISME study was approved by the local scientific ethics committee of Aarhus County, Denmark.

Participants included in the criterion validity analysis
The criterion validity of the S-NAQ cut-off points obtained through ROC curve analysis was tested in the MODENA sample, using both depressive symptoms and diagnosis of depression as outcome. For the analysis involving depressive symptoms as outcome, from the 5037 participants with valid S-NAQ values, 966 were excluded because information was missing on the outcome and/or the confounders considered, resulting in a final sample of *N* = 4071 participants. With regard to the analysis involving diagnosis of depression (assessed via the Schedules for Clinical Assessment in Neuropsychiatry (SCAN) interviews, which took place 3 to 5 months after the questionnaires were completed), from the 5037 participants we excluded: (i) those with missing information on the confounders entered in the logistic regression analyses (*N* = 187) and (ii) those who were selected only based on their self-reported exposure to workplace bullying and/or negative acts in the questionnaires (*N* = 116), which was one of the four criteria employed to select participants for the SCAN interviews (see below for details about the selection criteria). This exclusion was decided to address differential misclassification that may result from the fact that those who were selected for the SCAN interviews only based on their self-reported exposure to bullying and/or negative acts in the questionnaires, presented a higher probability of being selected for the diagnostic interviews, thus

potentially inflating the association between the newly formed S-NAQ categories and diagnosis of depression. Overall, this exclusion procedure resulted in 193 participants excluded (note that this number does not correspond to the sum of participants excluded according to criterion (i) and (ii) because the two exclusion groups were partly overlapping); therefore, the final sample included in the criterion validity analysis involving diagnosis of depression as outcome was composed of 4844 participants.

Measures of workplace bullying
S-NAQ
We used a Danish version of the S-NAQ scale, which is a translation of the S-NAQ validated by Notelaers and Einarsen (2008). These authors selected nine items from the 22-item NAQ-R (Einarsen *et al.*, 2009) based on the best fitting model out of a set of confirmatory factor analyses. A high convergence (*r* = 0.85) between the S-NAQ and the NAQ-R was observed in the study by Notelaers and Einarsen (2008). The Danish version of the S-NAQ asks the respondents to report, while referring to the previous 12 months, the frequency of exposure to each of the nine acts included, choosing one among five possible options (1 = 'never', 2 = 'now and then', 3 = 'monthly', 4 = 'weekly', 5 = 'daily'). The S-NAQ scale was calculated by summing the scores obtained on the nine items (Table 2), resulting in a theoretical range from 9 to 45, with higher scores indicating a higher perceived exposure to negative acts. The scale was not computed for those 452 participants with one or more missing values over the nine items.

Self-labelling method with a definition
Prior to asking the question on bullying, the respondents were presented with the following definition of the

phenomenon: 'Bullying at work occurs when one or several persons are exposed repeatedly and over a long period of time to unpleasant or negative acts or behaviours at work, which are difficult to defend oneself against'. This was followed by a commonly used one-item measure of workplace bullying (Nielsen *et al.*, 2011): 'Have you been exposed to bullying at your current workplace within the last 12 months?' The five response categories were 'never', 'now and then', 'monthly', 'weekly', and 'daily'. A three-level exposure variable with the following categories was created: 'never bullied', 'occasional bullying' ('now and then' and 'monthly'), and 'frequent bullying' ('weekly' and 'daily').

Depressive symptoms

Depressive symptoms were measured using the SCL-DEP6 subscale included in the Common Mental Disorders Questionnaire (CMDQ; Christensen *et al.*, 2005). SCL-DEP6 consists of six items, to be answered using a response scale with five options, i.e. 'Not at all', 'A little', 'Moderately', 'Quite a bit', 'Extremely'. Each item is coded 0 if the participant answers 'Not at all', while is coded 1 when any other response option is endorsed. A scale sum score is then computed by adding up the individual responses across the six items. The SCL-DEP6 scale score may thus range from 0 to 6, with higher scores indicating higher levels of depressive symptoms.

Diagnosis of depression

The diagnosis of clinical depression was made according to the International Classification of Diseases, 10th Revision, Diagnostic Criteria for Research, using SCAN interviews, version 2.1, Part I. We used the following sections of Part I: 3 (worrying and tension), 4 (panic, anxiety, and phobias), and 6 to 8 (depression). All interviews referred to the previous 3 months and were all computer-aided and semi-structured. Each interview took about one hour to complete and was conducted by a trained psychology or medical student. Out of the 5037 participants with valid values on the S-NAQ, 286 participated in the SCAN interview among the 597 invited. The selection of participants to invite in the SCAN interviews was based on the following four inclusion criteria (Gullander *et al.*, 2014; Hogh *et al.*, 2016): (i) being exposed, weekly or daily, to at least one of the negative acts included in the S-NAQ, or self-labelling as bullied at work (monthly or more often) during the last 12 months (N = 195); (ii) reporting depressive symptoms (point score of at least three on three or more of the six depressive symptoms items on the CMDQ, N = 221); (iii) reporting anxiety symptoms (point score of at least two of the three

anxiety symptoms items on the CMDQ, N = 166). One last group was selected randomly (N = 197). The total number of participants selected for the SCAN interviews does not add up to 597 because some of the participants selected fulfilled more than one criterion.

Statistical analysis
ROC curve analyses

We calculated the descriptive statistics for the sample and for both the sum score and the single items of the S-NAQ. We conducted two separate ROC analyses to determine two cut-off points for the S-NAQ, using the 'occasional bullying' ('now and then' and 'monthly') and 'frequent bullying' ('weekly' and 'daily') categories of the self-labelling method as reference standards (Table 3). ROC curve analysis is a statistical technique enabling the identification of optimal cut-off points that correspond to the best possible combinations of sensitivity (i.e. the ability of a test to correctly identify subjects with a condition, i.e. the true positives) and specificity (i.e. the ability of a test to correctly identify subjects without a condition, i.e. the true negatives) for a given screening tool. To determine the diagnostic accuracy of a ROC curve, the non-parametric statistic 'area under the curve' (AUC) is used, with values closer to 1 indicating that an instrument reliably distinguishes between those exposed (i.e. targets of bullying) and those not exposed (i.e. non-targets of workplace bullying), whereas values around 0.50 indicate no better than chance prediction (Zhou *et al.*, 2002). According to Swets *et al.* (2000), AUC values ≥0.9 are 'excellent', ≥0.80 'good,' ≥0.70 'fair', and <0.70 'poor'. The optimal cut-off points for the S-NAQ, using the 'occasional bullying' and 'frequent bullying' categories as reference standards, were determined by means of the Youden index (Youden, 1950), which ranges from 0 (complete overlap between those exposed and not-exposed to bullying) to 1 (complete separation between those exposed and non-exposed to bullying). The ROC curve analyses were performed using IBM SPSS for Windows, version 24.0 and the free version of Analyse-It, which is a recommended program for conducting ROC curve analyses (Stephan *et al.*, 2003).

Criterion validity analysis

We created a three-level categorical exposure variable for the S-NAQ, i.e. 'Non-exposed', 'First threshold' and 'Second threshold', based on the two cut-off points obtained through the ROC curve analyses. The association between the newly formed three-level exposure and continuous depressive symptoms score was tested

Table 3. Results of the receiver operating characteristic curve analysis for identifying optimal cut-off points for the S-NAQ using occasional bullying and frequent bullying as reference standards (*N* = 4882).

	Self-labelling method with a definition	
	Occasional bullying	Frequent bullying
Area under the ROC curve (AUC)	0.89 (95% CI 0.88 to 0.90)	0.93 (95% CI 0.93 to 0.94)
z statistic of AUC (*P* level)	33.13 (<0.001)	12.22 (<0.001)
Youden's index	0.62 (95% CI 0.57 to 0.67)	0.83 (95% CI 0.65 to 0.92)
S-NAQ score associated with Youden's index	>11 (95% CI >11 to >12)	>15 (95% CI >15 to >16)
Sensitivity	84.8	88.0
Specificity	77.4	94.7

using multivariate linear regression analysis, estimating unstandardized B coefficients and their 95% confidence intervals (CIs). The association between the newly formed three-level exposure and diagnosed depression was tested using multivariate logistic regression analysis, estimating odds ratios (ORs) and 95% CI. For both outcomes, we calculated both crude and adjusted coefficients (Model 1 and Model 2 in Tables 6 and 7, respectively). We considered the following set of potential confounders, based on previous evidence: gender, age, cohabitation, education, smoking status, personal history of mental disorders, family history of mental disorders, psychotropic drug prescriptions in the past week (Kessler, 1997; Hasin *et al.*, 2005; Burcusa *et al.*, 2007; Andersen *et al.*, 2009; Boden *et al.*, 2010), and sense of coherence (Antonovsky, 1993). Except for gender (*P* = 0.087), all confounders were significantly (*P* < 0.05) associated with the S-NAQ three-level exposure variable. Because all of the confounders were also significantly (*P* < 0.05) associated with depressive symptoms (Table 4), all of them were entered in the adjusted linear regression model. With specific regard to the analysis including diagnosis of depression, given the few cases available presenting a diagnosis of depression and the presence of missing values on these potential confounders, to avoid loss of participants in the multivariate models we decided to adjust for gender, age and, among the other confounders considered, only for those that showed significant bivariate associations with diagnosed depression and/or did change the OR for the association between the S-NAQ three-level exposure variable and depression by 10% or more (see Table 5 for the bivariate associations between each potential confounder and diagnosed depression). Using these criteria, in the analysis including diagnosis of depression as outcome, we adjusted only for gender, age, cohabitation, and sense of coherence. The multivariate linear and logistic analyses were performed using IBM SPSS for Windows, version 24.0.

Results

ROC curve analyses

The observed empirical range of the S-NAQ sum score scale was between 9 and 39. The overall mean of the S-NAQ scale was 10.9 (SD = 2.8).

The results of the two ROC curve analyses are shown in Table 3. When using the 'occasional bullying' category of the self-labelling question as reference standard, we obtained an AUC of 0.89 (95 % CI = 0.88–0.90; *P* < 0.001), which is very close to the threshold of 0.90 indicating excellent classification accuracy. The optimal cut-off point for the S-NAQ was ≥12, which was associated with a Youden's index of 0.62 (95% CI = 0.57–0.67). The cut-off point of ≥12 corresponded to a sensitivity of 84.8% (true positives) and a specificity of 77.4% (true negatives). The ROC curve is displayed in Fig. 1.

When using the 'frequent bullying' category of the self-labelling question as reference standard, the AUC was 0.93 (95% CI = 0.93–0.94; *P* < 0.001), which is above the threshold of 0.90 indicating excellent classification accuracy. The optimal cut-off point for the S-NAQ was ≥16, which was associated with a Youden's index of 0.83 (95% CI = 0.65–0.92). The cut-off point of ≥16 corresponded to a sensitivity of 88% (true positives) and a specificity of 94.7% (true negatives). The ROC curve is displayed in Fig. 2.

When applying these cut-off scores on the whole sample (*n* = 4882), 74.1% had a S-NAQ score below or equal to the first threshold (<12); 20.1% had a score between 12 and 15 (i.e. equal to the first threshold and below the second threshold); 5.8% had a score equal to or above the second threshold (≥16).

Criterion validity analysis
Association with depressive symptoms
The association between the newly formed S-NAQ three-level categorical exposure and depressive symptoms is

Table 4. Bivariate associations between sample characteristics and depressive symptoms (*N* = 4071[a])

Sample characteristics	N	Column %	Mean (SD)	Depressive symptoms
				B (95% CI)
Gender				
Women	3060	75.2		0.24 (0.13–0.36)***
Men (ref.)	1011	24.8		
Age (in years)			49.3 (9.5)	−0.01 (−0.013 to −0.003)**
Cohabitation				
Cohabiting (ref.)	3402	83.6		–
Living alone	669	16.4		0.54 (0.40–0.67)***
Education				
<3 years	1068	26.2		0.06 (−0.11 to 0.22)
3–4 years	2406	59.1		0.18 (0.03–0.32)*
>4 years (ref.)	597	14.7		–
Smoking status				
Non-smoker (ref.)	3564	87.5		–
Current smoker	507	12.5		0.15 (0.001–0.30)*
Personal history of depression				
No (ref.)	4004	98.4		–
Yes	67	1.6		1.03 (0.64–1.42)***
Family history of depression				
No (ref.)	3030	74.4		–
Yes	361	8.9		0.52 (0.34–0.69)***
Do not know	680	16.7		0.44 (0.30–0.57)***
Psychotropic drug prescriptions				
No (ref.)	4051	99.5		–
Yes	20	0.5		1.39 (0.68–2.09)***
Sense of coherence			83.6 (11.4)	−0.07 (−0.073 to −0.065)***

[a]Final sample included in the multivariate linear regression analysis (Table 6) that results after listwise excluding those participants with missing values on the S-NAQ, the confounders and/or depressive symptoms.

*$P < 0.05$; **$P < 0.01$; ***$P < 0.001$.

shown in Table 6. In the unadjusted model (Model 1), both the first and the second thresholds of exposure were significantly associated with higher scores on the depressive symptoms scale (*B* = 0.81, 95% CI = 0.69–0.93 and *B* = 1.77, 95% CI = 1.56–1.98, respectively). In the fully adjusted model, the association remained significant for both thresholds (*B* = 0.78, 95% CI = 0.66–0.90 and *B* = 1.65, 95% CI = 1.44–1.86, respectively).

Association with diagnosis of depression

The association between the newly formed S-NAQ three-level categorical exposure and diagnosis of depression is shown in Table 7. In all, 63 participants were diagnosed with depression. In the unadjusted model (Model 1), both the first and the second thresholds of exposure were significantly associated with a higher probability of being diagnosed with a depression (OR = 4.31, 95% CI = 2.42–7.68 and OR = 9.78, 95% CI = 5.07–18.82,

respectively). In the fully adjusted model, the associations remained significant for both thresholds (OR = 3.55, 95% CI = 1.98–6.38 and OR = 5.90, 95% CI = 2.93–11.88, respectively).

Discussion

In the present study, we identified two optimal cut-off points (≥12 and ≥16) for the S-NAQ reflecting a possible experienced exposure to occasional and frequent bullying behaviours, respectively. Based on these cut-off points, we created a new three-level categorical exposure for the S-NAQ that demonstrated criterion validity in relation to both depressive symptoms and diagnosed depression.

The two cut-off points identified lay close to the lower end of the S-NAQ scale. This indicates that employees responding to the S-NAQ may perceive

Table 5. Bivariate associations between sample characteristics and diagnosis of depression (*N* = 4844[a]).

Sample characteristics	N	Column % (valid values)	Mean (SD)	Depression, N (row %)	OR (95 % CI)
Gender					
Women	3639	75.1		51 (1.4)	1.43 (0.75–2.66)
Men (ref.)	1205	24.9		12 (1.0)	1
Age (in years)			49.1 (9.6)	63 (1.3)	0.99 (0.97–1.02)
Cohabitation					
Cohabiting (ref.)	4034	83.3		46 (1.1)	1
Living alone	810	16.7		17 (2.1)	1.86 (1.06–3.26)*
Education					
<3 years	1150	23.7		13 (1.1)	0.54 (0.25–1.17)
3–4 years	2572	53.1		34 (1.3)	0.63 (0.33–1.20)
>4 years(ref.)	626	12.9		13 (2.1)	1
Missing	496				
Smoking status					
Non-smoker (ref.)	642	13.3		50 (1.2)	1
Current smoker	4202	86.7		13 (2.0)	1.72 (0.93–3.18)
Personal history of depression					
No (ref.)	4612	95.2		56 (1.2)	1
Yes	85	1.8		3 (3.5)	2.98 (0.92–9.71)
Missing	147				
Family history of depression					
No (ref.)	3474	71.7		41 (1.2)	1
Yes	422	8.7		6 (1.4)	1.21 (0.51–2.86)
Do not know	810	16.7		14 (1.7)	1.7 (0.80–2.72)
Missing	138				
Psychotropic drug prescriptions					
No (ref.)	4725	97.5		59 (1.2)	1
Yes	22	0.5		1 (4.5)	3.77 (0.50–28.46)
Missing	97				
Sense of coherence			83.1 (11.8)	63 (1.3)	0.95 (0.93–0.96)***

[a]Final sample included in the multivariate logistic regression analysis (Table 6) that results after listwise excluding those participants with missing values on the S-NAQ, the confounders entered in the adjusted model (gender, age, cohabitation, and sense of coherence) and/or diagnosis of depression.

*$P < 0.05$; ***$P < 0.001$.

themselves as occasionally or frequently bullied already at low thresholds of exposure to negative acts. It must be noted that our cut-off points were lower, in proportion to the scale range, than those derived by Notelaers and Einarsen (2013) for the 22-item NAQ-R (their cut-off points were, respectively, ≥33 and ≥45 on a scale ranging from 22 to 110). One possible explanation for this discrepancy is that in our study we used, as reference standard, the self-labelling method only, whereas in Notelaers and Einarsen (2013) the reference standard did include both the self-labelling method and self-reported symptoms of depression and anxiety. Our choice might have resulted in S-NAQ cut-off points being more liberal in classifying respondents as targets of bullying than those

obtained by Notelaers and Einarsen (2013). However, an advantage of using a reference standard consisting of the self-labelling method only is that the resulting cut-off points may be useful to identify individuals who, independent of their current mental health status, may be at a higher risk of developing future mental health problems as a result of their perceived exposure to bullying behaviours. Another possible explanation for our lower cut-off points could be that, as noted previously by other authors (Nielsen *et al.*, 2011; Hogh *et al.*, 2016), the NAQ-R contains items, not included in the S-NAQ, that can reflect job stressors caused by poor organizational conditions rather than potential bullying behaviours (e.g. 'Being given tasks with unreasonable deadlines' or 'Being
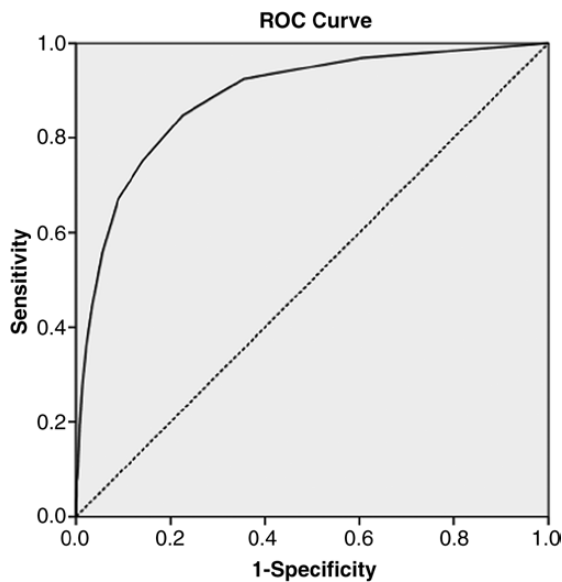
Figure 1. ROC curve for determining optimal cut-off points on the continuous Short Negative Acts Questionnaire (S-NAQ) sum scores, using occasional bullying as reference standard.
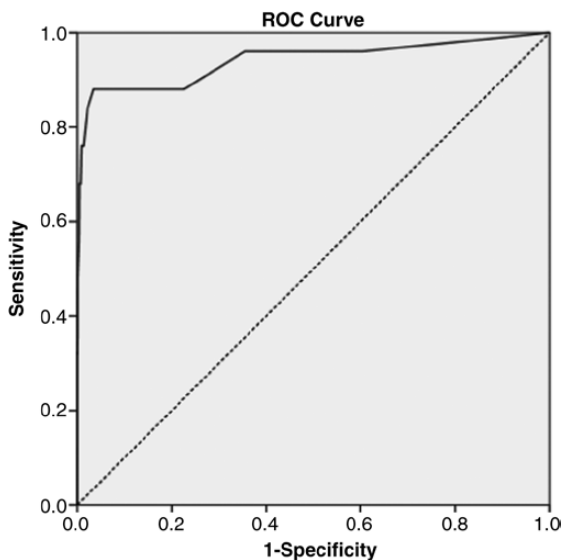

Figure 2. ROC curve for determining optimal cut-off points on the continuous Short Negative Acts Questionnaire (S-NAQ) sum scores, using frequent bullying as reference standard.

exposed to unmanageable workload'). These items may inflate some of the scores obtained on the NAQ-R, thus increasing the threshold for perceiving oneself as bullied.

A major benefit of adopting a ROC-based approach is that the identification of meaningful cut-off points

for instruments based on the behavioural experience method is performed on an empirical basis. Moreover, this method allows identifying cut-off points that reflect different levels of exposure to bullying behaviours. Contrary to this, with the operational criterion approach, frequently employed when using the behavioural experience method, the cut-off points are chosen arbitrarily and return a coarse, dichotomous classification that simply distinguishes between targets and non-targets of bullying (e.g. Leymann, 1990; Mikkelsen and Einarsen, 2001; Agervold, 2007). However, providing a more nuanced classification of targets aligns with the notion, put forth previously (e.g. Notelaers *et al.*, 2006), that bullying should not be considered as an 'either-or' phenomenon. In particular, identifying employees experiencing occasional bullying, in addition to those experiencing frequent bullying, is crucial since earlier studies showed that a perceived exposure to occasional bullying is already associated with reduced mental health (e.g. Gullander *et al.*, 2014); moreover, occasional bullying, if not managed appropriately, may evolve into more severe forms of bullying (Einarsen, 1999; Einarsen *et al.*, 2011).

The proposed S-NAQ cut-off points showed criterion validity as demonstrated by their significant association with both depressive symptoms and diagnosed depression. With specific regard to diagnosis of depression, although the OR for the second threshold was higher than the OR for the first threshold, the overlapping 95% CI do not allow us to indicate a dose–response association.

Previously, significant associations with depression (Rugulies *et al.*, 2012; Gullander *et al.*, 2014; Bonde *et al.*, 2016) and depressive symptoms (Bonde *et al.*, 2016) were found in studies using the self-labelling method to measure the exposure to workplace bullying. With regard to previous research adopting the behavioural experience method, we are not aware of studies examining the association between empirically derived cut-off points for the S-NAQ and mental health outcomes. A few studies examined the association between continuous S-NAQ scores and depressive symptoms, obtaining mixed results (Hauge *et al.*, 2010; Reknes *et al.*, 2014). Using the NAQ-R, Einarsen and Nielsen (2015) observed a 5-year prospective association between a dichotomous exposure based on the ≥33 cut-off point proposed by Notelaers and Einarsen (2013) and symptoms of anxiety and depression. In a 2-year follow-up study adopting a slightly modified 23-item version of the NAQ-R, Hogh *et al.* (2016) failed to find significant associations between

Table 6. Multivariate linear regression for the association between S-NAQ thresholds and depressive symptoms (*N* = 4071).

	Total	Mean (SD)	Model 1 unadjusted *B*[d] (95% CI)	Model 2 adjusted *B*[e]
	N (column %)			
Depressive symptoms	4071	1.27 (1.61)		
S-NAQ				
Reference[a]	3040 (74.7)	1.01 (1.44)	1	1
First threshold[b]	812 (19.9)	1.79 (1.79)	0.81 (0.69–0.93)***	0.78 (0.66–0.90)***
Second threshold[c]	219 (5.4)	2.78 (1.90)	1.77 (1.56–1.98)***	1.65 (1.44–1.86)***

[a]Participants scoring <12 on the S-NAQ sum score scale.
[b]Participants scoring between 12 and 15 on the S-NAQ sum score scale.
[c]Participants scoring ≥16 on the S-NAQ sum score scale.
[d]Unstandardized B coefficients.
[e]Model 2 adjusted for gender, age, cohabitation, education, smoking status, personal history of depression, family history of depression, psychotropic drug prescriptions, sense of coherence.
***P < 0.001.

Table 7. Multivariate logistic regression for the association between S-NAQ thresholds and diagnoses of depression (*N* = 4844).

	Total	Cases	Model 1 crude, OR (95% CI)	Model 2[d], OR (95% CI)
	N (column %)	*N* (row %)		
Depression		63 (1.3)		
S-NAQ				
Reference[a]	3573 (74.7)	22 (0.6)	1	1
First threshold[b]	942 (19.7)	25 (2.6)	4.31 (2.42–7.68)***	3.55 (1.98–6.38) ***
Second threshold[c]	266 (5.6)	16 (5.7)	9.78 (5.07–18.82) ***	5.90 (2.93–11.88) ***

[a]Participants scoring <12 on the S-NAQ sum score scale.
[b]Participants scoring between 12 and 15 on the S-NAQ sum score scale.
[c]Participants scoring ≥16 on the S-NAQ sum score scale.
[d]Model 2 adjusted for gender, age, cohabitation, and sense of coherence.
***P < 0.001.

four different categories of potential bullying behaviours (measured using three-level categorical exposure variables) and subsequent diagnoses of depression. The comparison with previous studies thus suggests that categorizing the S-NAQ using empirically derived cut-off points may enhance the criterion validity of this instrument in relation to depressive symptoms and diagnosis of depression.

Strengths and limitations

The two major strengths of this study are the large sample size used to derive the cut-off points for the S-NAQ and the fact that we employed diagnosis of depression, in addition to self-reported depressive symptoms, to test the criterion validity of the new three-level categorical exposure.

Nevertheless, this study presents some limitations worth considering. A first limitation relates to our choice of the reference standard, i.e. the self-labelling method with a definition. The subjectivity inherent in the self-labelling method may introduce information bias in the measurement of bullying. The reporting may in fact be influenced by defensive mechanisms such as the difficulty to admit to one's victimization and the presence of different views about what constitutes bullying (Nielsen *et al.*, 2011). Yet, the definition of bullying provided before the question seems effective in calibrating the individual answers and reducing the bias deriving from different interpretations of the phenomenon (Nielsen *et al.*, 2010; Nielsen *et al.*, 2011).

A second limitation is the high prevalence of participants belonging to the groups delimited by the first and

the second threshold. The prevalence rates of these two thresholds, which include participants who are exposed to negative acts and perceive these acts as occasional or frequent bullying, were higher than those observed by Notelaers and Einarsen (2013) when they applied to their sample the cut-off points they obtained for the NAQ-R. In addition, our thresholds overestimate the prevalence of workplace bullying if compared to previous studies conducted on the Danish working population (Ortega *et al.*, 2009). The high prevalence rates obtained in this study result from the aforementioned low S-NAQ cut-off points, implying that a lower threshold of exposure was needed in our study to classify respondents as targets of behaviours perceived as occasional or frequent bullying. However, it must be noted that the specificity of our first cut-off point was 77.4%, while the corresponding specificity was 88.6% in the study by Notelaers and Einarsen (2013). Thus, our study presents a higher number of false positives. The fact that we used a 12-months' time window, instead of the 6-months' time window adopted by Notelaers and Einarsen (2013) and in previous studies using the NAQ-R (Einarsen *et al.*, 2009), could have also partially contributed to the higher prevalence of exposed observed in our study.

A third limitation is that our study design permitted us to establish only the concurrent, but not the predictive validity of the new S-NAQ categorical exposure in relation to both depressive symptoms and diagnosis of depression. Regarding diagnosis of depression, even if the interviews were conducted three to five months after the questionnaires, our study cannot be considered prospective given the lack of diagnostic information available to exclude the depressed at baseline. The direction of causality could therefore not be determined in our study, as some of the participants might have already shown depressive symptoms or might have been already clinically depressed at baseline as a result of a previous possible exposure to bullying behaviours and/or to other risk factors. We were therefore unable to exclude reverse causation, reflecting the hypothesis that people with depression or depressive symptoms show a higher risk of being exposed to bullying (e.g. Nielsen *et al.*, 2014; Hogh *et al.*, 2016). Further research based on prospective designs is thus needed to establish the predictive validity of the new S-NAQ categorical exposure proposed in the present study.

A fourth limitation is that misclassification may have affected the coefficients calculated in the linear and logistic regression models. In particular, exposure contrast might have been reduced as a consequence of misclassification, leading to risk estimates being biased toward the null. A fifth and final limitation is that our sample,

despite involving numerous occupational sectors, was not representative of the Danish working population. The generalizability of our findings to the Danish working population is also affected by the high proportion of women in our sample (75.1%). Our cut-off points need therefore to be verified in representative samples of the Danish workforce. Furthermore, the cut-off points may not be generalizable to other countries; the identification of country-specific cut-off points for the S-NAQ is thus recommended (Notelaers and Einarsen, 2013).

Practical implications

Despite the two S-NAQ cut-off points demonstrated high classification accuracy, some caution is warranted in their use in the domain of practice, given the above-mentioned limitations. The main precaution is related to the presence of possible misclassified respondents. In particular, when applying the first cut-off point, some respondents may be erroneously classified as exposed to behaviours that they perceive as occasional bullying, while they may actually not perceive themselves as bullied (false positives). On the one hand, the identification of false positives seems acceptable given the high individual and organizational costs associated with workplace bullying. In particular, the detection of low-threshold exposure to negative behaviours may be useful to organizations to capture emergent relational problems that can later escalate into bullying situations. On the other hand, false positives may incur significant costs given that remedial actions against bullying are resource consuming.

Despite these limitations, the identification of empirical cut-off points for a well-known inventory such as the S-NAQ may improve the use of this instrument in order to inform workplace interventions to prevent workplace bullying. Specifically, empirically derived cut-off points can be applied to the S-NAQ scores to identify thresholds of exposure to negative behaviours that reflect a possible perceived exposure to workplace bullying of increased severity. Given the importance of the subjective perception of being bullied in determining the health impact of such exposure, identifying targets who may perceive the negative acts as bullying behaviours can improve the estimation, and the subsequent prevention, of the risk associated with bullying in organizational settings. On the contrary, this estimation is problematic when using arbitrary cut-off points as a way to distinguish between targets and non-targets (as it occurs with the operational criterion), or when relying on continuous scale scores.

Obviously, if an organization is simply interested in estimating the prevalence of employees perceiving themselves

as bullied, the self-labelling method with a definition is the method of choice. However, as also recognized previously (see, e.g. Nielsen *et al.*, 2009, 2011), a complete assessment of workplace bullying should rely on a mix of instruments, which enables the different aspects of the phenomenon to be captured. In particular, behavioural inventories such as the S-NAQ are necessary to obtain information, not provided by the self-labelling method, on emergent relational problems as well as on the nature of the specific bullying behaviours occurring in organizations (Nielsen *et al.*, 2009). Once groups of employees with different levels of exposure to bullying behaviours are detected, the organization can then tap into the specific information provided by the behavioural inventories in order to identify the acts that mostly contributed to the reported exposure, and devise targeted intervention plans accordingly.

Conclusions

In conclusion, in the present study we determined, by means of ROC curve analysis, two optimal cut-off points (i.e. ≥12 and ≥16) for the 9-item S-NAQ that can be used to classify employees experiencing a possible self-perceived exposure to occasional and frequent bullying behaviours. Based on these thresholds, we formed a new three-level categorical exposure variable that demonstrated criterion validity in relation to depressive symptoms and diagnosed depression. However, due to limitations in the study design it was not possible to establish the predictive validity of the new S-NAQ categorical exposure, which should be therefore tested in future prospective studies.

Declaration

Conflict of Interest

The authors declare no conflict of interest relating to the material presented in this article. Its contents, including any opinions and/or conclusions expressed, are solely those of the authors.

References

Agervold M. (2007) Bullying at work: a discussion of definitions and prevalence, based on an empirical study. *Scand J Psychol*; **48**: 161–72.

Andersen I, Thielen K, Nygaard E *et al.* (2009) Social inequality in the prevalence of depressive disorders. *J Epidemiol Community Health*; **63**: 575–81.

Antonovsky A. (1993) The structure and properties of the sense of coherence scale. *Soc Sci Med*; **36**: 725–33.

Balducci C, Fraccaroli F, Schaufeli WB. (2011) Workplace bullying and its relation with work characteristics, personality, and post-traumatic stress symptoms: an integrated model. *Anxiety Stress Coping*; **24**: 499–513.

Boden JM, Fergusson DM, Horwood LJ. (2010) Cigarette smoking and depression: tests of causal linkages using a longitudinal birth cohort. *Br J Psychiatry*; **196**: 440–6.

Bonde JP, Gullander M, Hansen ÅM *et al.* (2016) Health correlates of workplace bullying: a 3-wave prospective follow-up study. *Scand J Work Environ Health*; **42**: 17–25.

Burcusa SL, Iacono WG. (2007) Risk for recurrence in depression. *Clin Psychol Rev*; **27**: 959–85.

Burton P. (2015) The role of job embeddedness in the relationship between bullying and aggression. *Eur J Work Organ Psy*; **24**:518–29.

Christensen KS, Fink P, Toft T *et al.* (2005) A brief case-finding questionnaire for common mental disorders: the CMDQ. *Fam Pract*; **22**: 448–57.

Einarsen S. (1999) The nature and causes of bullying at work. *Int J Manpower*; **20**:16–27.

Einarsen S, Nielsen MB. (2015) Workplace bullying as an antecedent of mental health problems: a five-year prospective and representative study. *Int Arch Occup Environ Health*; **88**: 131–42.

Einarsen S, Hoel H, Notelaers G. (2009) Measuring exposure to bullying and harassment at work: Validity, factor structure and psychometric properties of the Negative Acts Questionnaire-Revised. *Work Stress*; **23**: 24–44.

Einarsen S, Hoel H, Zapf D *et al.* (2011) The concept of bullying at work: The European tradition. In: Einarsen S, Hoel H, Zapf D, Cooper CL, editors. *Bullying and harassment in the workplace: Developments in theory, research, and practice*. London, England: CRC Press; p. 3–30. ISBN 9781439804896.

Einarsen S, Skogstad A. (1996) Bullying at work: Epidemiological findings in public and private organizations. *Eur J Work Organ*; **5**: 185–201.

Gullander M, Hogh A, Hansen ÅM *et al.* (2014) Exposure to workplace bullying and risk of depression. *J Occup Environ Med*; **56**: 1258–65.

Hasin DS, Goodwin RD, Stinson FS *et al.* (2005) Epidemiology of major depressive disorder: results from the National Epidemiologic Survey on Alcoholism and Related Conditions. *Arch Gen Psychiatry*; **62**: 1097–106.

Hauge LJ, Skogstad A, Einarsen S. (2010) The relative impact of workplace bullying as a social stressor at work. *Scand J Psychol*; **51**: 426–33.

Hewett R, Liefooghe A, Visockaite G *et al.* (2016) Bullying at work: cognitive appraisal of negative acts, coping, well-being, and performance. *J Occup Health Psychol*. doi: 10.1037/ocp0000064.

Hogh A, Hansen AM, Mikkelsen EG *et al.* (2012) Exposure to negative acts at work, psychological stress reactions and physiological stress response. *J Psychosom Res*; **73**: 47–52.

Hogh A, Conway PM, Grynderup MB *et al.* (2016) Negative acts at work as potential bullying behavior and depression:

examining the direction of the association in a 2-year follow-up study. *J Occup Environ Med*; **58**: e72–9.

Kessler RC. (1997) The effects of stressful life events on depression. *Annu Rev Psychol*; **48**: 191–214.

Kolstad HA, Hansen AM, Kærgaard A *et al.* (2011) Job strain and the risk of depression: is reporting biased? *Am J Epidemiol*; **173**: 94–102.

Lazarus RS, Folkman S. (1984) *Stress, appraisal, and coping*. New York: Springer. ISBN 9780826141910.

Leymann H. (1990) Mobbing and psychological terror at workplaces. *Violence Vict*; **5**: 119–26.

Mikkelsen EG, Einarsen S. (2001) Bullying in Danish worklife: prevalence and health correlates. *Eur J Work Organ Psy*; **10**:393–13.

Neall A, Tuckey MR. (2014) A methodological review of research on the antecedents and consequences of workplace harassment. *J Occup Organ Psychol*. ;**87**:225–57.

Nielsen MB, Skogstad A, Matthiesen SB *et al.* (2009) Prevalence of workplace bullying in Norway: comparisons across time and estimation methods. *Eur J Work Organ Psy*; **18**:81–101.

Nielsen MB, Magerøy N, Gjerstad J *et al.* (2014) Workplace bullying and subsequent health problems. *Tidsskr Nor Laegeforen*; **134**: 1233–8.

Nielsen MB, Matthiesen SB, Einarsen S. (2010) The impact of methodological moderators on prevalence rates of workplace bullying: a meta-analysis. *J Occup Organ Psychol*; **83**: 955–79.

Nielsen MB, Notelaers G, Einarsen S. (2011) Measuring exposure to workplace bullying. In: Einarsen S, Hoel H, Zapf D, Cooper CL, editors. *Bullying and harassment in the workplace: Developments in theory, research, and practice*. London, England: CRC Press. p. 149–175. ISBN 9781439804896.

Notelaers G, Einarsen S. (2013) The world turns at 33 and 45: defining simple cutoff scores for the Negative Acts Questionnaire–Revised in a representative sample. *Eur J Work Organ Psy*; **22**: 670–82.

Notelaers G, Einarsen S. (2008) The construction and validity of the Short Negative Acts Questionnaire. In: Soares A, Jeanneau N, Geneviève P, Geneviève A, editors. Proceedings of 6th International Conference on Workplace Bullying; 2008 June 4–6; Montreal, Canada. p. 104–5.

Notelaers G, Einarsen S, De Witte H *et al.* (2006) Measuring exposure to bullying at work: the validity and advantages of the latent class cluster approach. *Work Stress*; **20**: 288–301.

Ortega A, Høgh A, Pejtersen JH *et al.* (2009) Prevalence of workplace bullying and risk groups: a representative population study. *Int Arch Occup Environ Health*; **82**: 417–26.

Reknes I, Pallesen S, Magerøy N *et al.* (2014) Exposure to bullying behaviors as a predictor of mental health problems among Norwegian nurses: results from the prospective SUSSH-survey. *Int J Nurs Stud*; **51**: 479–87.

Rugulies R, Madsen IE, Hjarsbech PU *et al.* (2012) Bullying at work and onset of a major depressive episode among Danish female eldercare workers. *Scand J Work Environ Health*; **38**: 218–27.

Salin D. (2001) Prevalence and forms of bullying among business professionals: a comparison of two different strategies for measuring bullying. *Eur J Work Organ Psy*; **4**: 425–41.

Salin D. (2003) Ways of explaining workplace bullying: a review of enabling, motivating and precipitating structures and processes in the work environment. *Hum Relat*; **56**: 1213–32.

Salin D. (2015) Risk factors of workplace bullying for men and women: the role of the psychosocial and physical work environment. *Scand J Psychol*; **56**: 69–77.

Stephan C, Wesseling S, Schink T *et al.* (2003) Comparison of eight computer programs for receiver-operating characteristic analysis. *Clin Chem*; **49**: 433–9.

Swets JA, Dawes RM, Monahan J. (2000) Psychological science can improve diagnostic decisions. *Psychol Sci Public Interest*; **1**: 1–26.

Theorell T, Hammarström A, Aronsson G *et al.* (2015) A systematic review including meta-analysis of work environment and depressive symptoms. *BMC Public Health*; **15**: 738.

Verkuil B, Atasayi S, Molendijk ML. (2015) Workplace bullying and mental health: a meta-analysis on cross-sectional and longitudinal data. *PLoS One*; **10**: e0135225.

Vie TL, Glasø L, Einarsen S. (2011) Health outcomes and self-labeling as a victim of workplace bullying. *J Psychosom Res*; **70**: 37–43.

Youden WJ. (1950) An index for rating diagnostic tests. *Cancer*; **3**: 32–5.

Zhou XH, Obuchowski NA, Obushcowski DM. (2002) *Statistical methods in diagnostic medicine*. New York: Wiley & Sons; ISBN 9780470183144.