ORIGINAL ARTICLE

Psychometric properties of the Need for Recovery after work scale: test-retest reliability and sensitivity to detect change

E M de Croon, J K Sluiter, M H W Frings-Dresen

Occup Environ Med 2006;**63**:202–206. doi: 10.1136/oem.2004.018275

An appendix to this paper is available on the *OEM* website (www.occenvmed.com/supplemental)

See end of article for authors' affiliations

Correspondence to: Dr J K Sluiter, Coronel Institute of Occupational Health, Academic Medical Center, PO BOX 227700, 1100 DE Amsterdam, the Netherlands; j.sluiter@amc.uva.nl

Accepted 11 November 2005

Background: Monitoring worker health and evaluating occupational healthcare interventions requires sensitive instruments that are reliable over time. The Need for Recovery scale (NFR), which quantifies workers' difficulties in recovering from work related exertions, may be a relevant instrument in this respect.
Objectives: To examine (1) the NFR's test-retest reliability and (2) the NFR's sensitivity to detect the effect of a fatigue inducing change, namely an increase in working hours.
Methods: Two year longitudinal data of 526 truck drivers and 144 nurses were used. Two week, one year, and two year test-retest reliability was examined in both stable and unstable work environments by calculating intraclass correlations (ICCs). Work environmental (in)stability was quantified by four events that might have occurred during the follow up period: (1) a reorganisation or merge (0 = *yes*, 1 = *no*), (2) a change of supervisor or management (0 = *yes*, 1 = *no*), (3) a change in working hours or work schedules (0 = *yes*, 1 = *no*), and (4) a change in work activities, position, or duties (0 = *yes*, 1 = *no*). The four scores constituted a work (in)stability index ranging from 0 to 4. The NFR's sensitivity to detect the effect of the increase in working hours was assessed indirectly by comparing it with an alternative scale, namely the Checklist Individual Strength.
Results: Test-retest reliability over a two year interval was good to excellent when applied in stable work environments (ICCs 0.68 to 0.80) but, as expected, poor to fair when applied in unstable work environments (ICCs 0.30 to 0.55). The NFR was sensitive in detecting an increase in work related fatigue due to the increase in working hours (effect size 0.40).
Conclusions: The NFR's test-retest reliability and sensitivity to detect change are favourable. This implicates that the NFR may form a valuable part of health surveys and may be a useful tool for evaluating occupational healthcare interventions.

The interest in recovery from work related exertions has intensified over the last decades.[1 2] The explanation for this increased interest is twofold. Firstly, incomplete recovery after work has been found an outcome of unfavourable working conditions such as long working hours,[3] high job demands, and low job control.[4] Secondly, workers with incomplete recovery have been found at an increased risk of developing occupational diseases such as burnout, cardiovascular disease, and musculoskeletal disorders.[5–11]

Monitoring difficulties in recovering from work and implementing interventions to reduce these difficulties may be a fruitful strategy to prevent these occupational diseases from occurring. Several instruments, including diaries[12] and the collection of catecholamine or cortisol in urine or saliva,[13] can assist practitioners for these purposes. Another instrument is the Need for Recovery scale (NFR)[14] which quantifies the difficulties workers experience in recovering from work.

Although the NFR has been frequently used,[13 15–19] psychometric evidence about this scale is scarce. To our knowledge, two peer reviewed studies have shown that the NFR is internally consistent and has favourable construct validity.[20 21] The NFR's test-retest reliability (reproducibility) and sensitivity to detect change (hereafter sensitivity) have, however, not been examined until now. Because monitoring recovery and evaluating interventions asks for sensitive instruments that are reliable over time, this lack of evidence seems surprising.

Three reasons may account for the observed lack of evidence. Firstly, producing evidence about test-retest reliability and sensitivity requires studies with a longitudinal design. These studies are comparatively expensive. Secondly, determining test-retest reliability of the NFR may be difficult because low test-retest correlations may denote low reliability of the NFR (measurement error) as well as true work induced change in need for recovery. Thirdly, examining the NFR's sensitivity requires a gold standard with which to define change in need for recovery. Such a standard, however, is not available.

This longitudinal study was designed to overcome these difficulties. For assessing test-retest reliability, measurement error and true change were set apart in two ways: (1) by calculating test-retest correlations over short and long intervals, and (2) by asking participants to indicate whether work related changes occurred during the follow up. It was expected that true work induced change in need for recovery was unlikely to occur over short intervals or among the participants who did not report work related changes. Under these conditions, then, test-retest reliability of the NFR was predicted to be good.

The NFR's sensitivity was examined by measuring its ability to detect the effect of an increase in working hours. Considering the fatiguing effect of long working hours,[22–25] this naturally occurring intervention was expected to increase need for recovery. The absence of a gold standard to define change in need for recovery was contended with by assessing the NFR's sensitivity in relative terms through comparison with the sensitivity of the established Checklist Individual Strength (CIS).[26] The CIS, which is an excellent measure of general fatigue, is a reliable and validated questionnaire.[20 26–29] Both NFR and CIS were expected to detect the effect of the quasi-experiment. However, because the NFR measures *work*

Abbreviations: NFR, Need for Recovery after work scale; CIS, Checklist Individual Strength; ICC, intraclass correlation

induced fatigue, the NFR was expected to detect the effect of the increase in working hours more efficiently.

METHODS

Subjects

Two year longitudinal data of 526 truck drivers and 144 hospital nurses were used for this study. The sample of truck drivers was taken from the directory of the Dutch Central Bureau of Occupational Health Care in Road transport. At baseline, 1225 of the 2000 truck drivers returned the questionnaire. Two years later, 102 of these 1225 drivers were no longer traceable. Therefore, 1123 drivers received a second questionnaire of which 820 were returned (72%). Non-response analyses revealed one significant difference in 1998 between non-respondents in 2000 and all respondents in 2000. Compared with the respondents, non-respondents were younger in 1998 (mean age 38 years *v* 40 years, $t = -3.41$, $p<0.01$). Because of retirement, permanent disability, dismissal, turnover (n = 243), or incomplete questionnaires (n = 51), 294 of the 820 respondents were excluded from the analyses. Eventually, this resulted in 526 available questionnaires that were used for this study too.

The sample of nurses was taken from a group of 540 Dutch nurses who participated in a two year, three wave prospective cohort study in one academic hospital (adjusted response = 51%).[4] For this study, a random sample of 200 nurses was taken from the prospective cohort of 540 nurses for the purpose of short term test-retest testing.[30] The sample size was chosen according to the 1:10 rule of thumb for reliability analysis to sample 10 subjects per item on the scale of interest and an expected 60% response. Two weeks after the third and last measurement of the prospective cohort study, these 200 nurses received a fourth questionnaire of whom 144 returned a completed questionnaire (72%). The analyses of this study were restricted to this sample of 144 nurses with valid baseline, one year, two year, and two year and two week follow up data. For reasons of transparency, it should be noted that the above described samples of truck drivers and nurses have also been used for other studies.

Questionnaire

The Need for Recovery after work scale (NFR)[14] comprises 11 dichotomous items (see http://www.occenvmed.com/supplemental). Typical items of this scale are: ''At the end of a working day I am really feeling worn-out'' and ''I find it hard to relax at the end of a working day''. The NFR score is calculated by adding the individual's scores on the 11 (recoded) items. This scale score is transformed into a scale ranging from 0 to 100. Higher scores indicate a higher degree of need for recovery after work.

The Checklist Individual Strength (CIS)[26] is a multidimensional fatigue questionnaire with 20 statements for which the worker has to indicate on a seven point scale to what extent they apply to him or her (1 = Yes, that is true; to 7 = No, that is not true). The statements refer to four fatigue aspects: (1) subjective fatigue (for example, *I feel tired*), (2) reduced motivation (for example, *I feel no desire to do anything*), (3) reduced activity (for example, *I don't do much during the day*), and (4) reduced concentration (for example, *my thoughts easily wander*). The CIS is validated in the clinical[31] and working situation.[27] In this study, a composite CIS score (ranging from 20 to 140) was calculated by adding the individual's scores on the four factors. Higher scores indicate a higher degree of fatigue, more concentration problems, reduced motivation, or low levels of activity.

Stability of the work environment

To determine whether the participants were employed in stable or unstable work environments they reported whether a reorganisation or merge (0 = *yes*, 1 = *no*), a change of supervisor or management (0 = *yes*, 1 = *no*), a change in working hours or work schedules (0 = *yes*, 1 = *no*), and a change in work activities, position, or duties (0 = *yes*, 1 = *no*) occurred during the follow up period. The four scores constituted a work stability index ranging from 0 to 4.

Increase in working hours

The naturally occurring intervention (that is, increase in working hours) was assessed at the two year follow up measurement among the truck drivers by asking whether, on average, the number of working hours had increased over the past two years. As a check on the reliability of the answers, for each driver the difference in working hours per week reported at baseline and the number of working hours per week reported at follow up were calculated. For 19 drivers this difference was not in agreement with the self reported change in working hours. Therefore, these drivers were excluded from the analyses.

Data analyses

Test-retest reliability was examined in truck drivers and nurses separately by calculating intraclass correlations (ICCs) for agreement (two-way mixed model).[30] To establish the level of clinical significance, guidelines for the evaluation of instruments measuring psychological constructs were used.[32] These guidelines state that when ICCs are below 0.40, the level of clinical significance is poor; when ICCs are between 0.40 and 0.59, the level of clinical significance is fair; when ICCs are between 0.60 and 0.74, the level of clinical significance is good; and when ICCs are between 0.75 and 1.00, the level of clinical significance is excellent.

Sensitivity to detect change was assessed by calculating changes in NFR and CIS scores among the drivers whose working hours did and did not increase during the follow up period. The sensitivity was examined by calculating the effect size ([*mean scale after intervention – mean scale before intervention*]/*SD of scale before intervention*). Clinically irrelevant effects were considered 0.00–0.19; small effects were considered 0.20–0.50, moderate effects were considered 0.51–0.80, and large effects were considered 0.81–1.00.[33]

Skewness and kurtosis

Examination of the NFR scores revealed a positively skewed distribution (that is, skewness ranges between 0.22 and 0.61; kurtosis ranges between −1.31 and −0.61). This was caused by the comparatively large group of respondents (approximating 30%) with the minimal score on the NFR (no recovery complaints at all), which has been found consistently in NFR research. Square root transformation of the scale did not improve the normality of the NFR scores distribution meaningfully (skewness ranges between −0.18 and 0.46; and kurtosis ranges between –1.44 and −0.88). Because of this finding and because transformation would just have increased the ICC outcomes artificially, no transformation of original scores was applied.

RESULTS

Descriptives

The truck drivers' and nurses' baseline age averaged 40 years (SD 9.7 and 9.3 years, respectively). Almost all drivers were male (99%) whereas the majority of the nurses was female (82%).

Test-retest reliability

More than one third of the truck drivers (37%) and the nurses (35%) reported no work related change during the follow up period (that is, stable work environment). Thirty eight per cent of the drivers were employed in a moderately

Table 1 Intra class correlations (ICCs) with 95% confidence intervals of the Need for Recovery after work scale over a period of two weeks, one year, and two years in the truck drivers and nurses

	Two week		One year		Two year	
	ICC (95% CI)		ICC (95% CI)		ICC (95% CI)	
Truck drivers (n = 526)	–	–	–	–		
Stable environment (no work related changes)	–	–	–	–	0.68 (0.59–0.75)	***
Moderately unstable environment (1 work related change)	–	–	–	–	0.69 (0.59–0.76)	***
Unstable environment (⩾2 work related changes)	–	–	–	–	0.55 (0.42–0.66)	**
Hospital nurses (n = 144)	–	–	–	–		
Stable environment (no work related changes)	0.67 (0.48–0.80)	***	0.72 (0.55–0.83)	***	0.80 (0.67–0.88)	****
Moderately unstable environment (1 work related change)	0.67 (0.52–0.78)	***	0.58 (0.40–0.72)	**	0.48 (0.28–0.64)	**
Unstable environment (⩾2 work-related changes)	0.65 (0.32–0.84)	***	0.69 (0.36–0.87)	***	0.30 (0.08–0.62)	*

*Level of clinical significance is poor; **level of clinical significance is fair; ***level of clinical significance is good; ****level of clinical significance is excellent.

stable work environment (that is, they reported one work related change during follow up) and 25% in an unstable work environment (that is, they reported two or more work related changes during follow up). Among the nurses these percentages were 50% and 15%, respectively.

The two year ICCs of the NFR are good among the truck drivers in stable work environments (0.68; 95% CI 0.59 to 0.75) and excellent among the nurses in stable work environments (0.80; 95% CI 0.67 to 0.88) (see table 1). However, they are poor (0.30; 95% CI 0.08 to 0.62) to fair (0.55; 95% CI 0.42 to 0.66) among the truck drivers and nurses in unstable work environments. Furthermore, the results show that the longer the time lag is, the smaller the ICC is. For instance, among the nurses in moderately unstable work environments, the ICC of the NFR drops from 0.67 over a two week lag (95% CI 0.52 to 0.78) to 0.58 (95% CI 0.40 to 0.72) over a one year lag to 0.48 (95% CI 0.28 to 0.64) over a two year lag.

Sensitivity to detect change

Almost a quarter of the truck drivers (24%) reported an increase in working hours during the follow up period. The average increase in number of working hours per week was 5.0 (SD 6.5). Before examining the NFR's sensitivity to detect the increase in working hours among these drivers, NFR and CIS scores at baseline of the drivers whose working hours increased were compared with NFR and CIS scores of the drivers whose working hours did not increase. These one-way analyses of variance revealed that the NFR scores ($F = 2.30$; $df = 1, 500$) and CIS-scores ($F = 2.53$; $df = 1, 492$) of the drivers in the natural intervention group were already significantly higher on baseline as compared to the scores of the drivers in the comparison group.

Table 2 presents the results of the analyses addressing the NFR's sensitivity. The NFR scores of the drivers whose working hours went up increased, but the scores remained stable among the drivers without an increase in working hours. The CIS scores of the drivers in the natural intervention group increased as well. However, although to a lesser extent, the CIS scores also increased in the drivers without an increase in working hours. The pre-post effect in the drivers with an increase in working hours was small with regard to the NFR (0.40) and with regard to the CIS (0.41). Among the drivers with stable working hours the pre-post effect detected by the NFR (0.05) and the CIS (0.18) was clinically irrelevant.

DISCUSSION

The NFR has been found to be internally consistent[20 21] which, however, does not imply that it can be reliably used to monitor health or to make inferences across long periods.[34 35] After all, instruments with many similar items may be internally consistent but, simultaneously, may display low test-retest correlations over time due to measurement error or true change in the variable of interest.[36] To distinguish measurement error of the NFR from true work induced change in need for recovery, the NFR's test-retest reliability was examined over (1) short and long intervals, and (2) among workers with and without self reported work related changes. In accordance with our expectation, ICC analyses revealed that, irrespective of the length of the time interval, reliability was good to excellent among workers who did not report work related changes. The analyses further revealed that, over a short term two week interval, reliability was also good among the workers who *did* report work related changes during the two year follow up. Over the full two year span of time, however, test-retest reliability among

Table 2 Mean (SD), paired *t* values, and effect size of the NFR and CIS at baseline and two year follow up among the drivers with (natural intervention group) and without (comparison group) an increase in working hours during the follow up

	Baseline (before increase in working hours)		Two year follow up (after increase in working hours)		Difference score (follow up–baseline)	Paired *t* value	Effect size	n
Natural intervention group (increase in working hours)								
NFR, mean (SD)	38.1	(31.3)	50.7	(37.0)	12.6 (30.6)	4.20*	0.40	104
CIS, mean (SD)	52.3	(23.7)	62.1	(26.3)	9.8 (24.9)	3.99*	0.41	103
Comparison group (no increase in working hours)								
NFR, mean (SD)	31.2	(31.1)	32.7	(32.9)	1.5 (26.5)	1.12	0.05	393
CIS, mean (SD)	46.5	(20.3)	50.1	(21.5)	3.6 (19.3)	3.68*	0.18	381

*$p<0.05$ (two-tailed).
Effect sizes:[32] large effect (0.81–1.00); moderate effect (0.51–0.80); small effect (0.20–0.50); clinically irrelevant effect (0.00–0.19).

these workers was found to be poor. This suggests that true *work induced* changes in need for recovery caused the low test-retest correlations among these workers and supports the reliability of the NFR.

Considering the relevance of the NFR as an outcome measure of occupational healthcare interventions, its sensitivity was also examined. To this end, the NFR's ability to distinguish an increase in need for recovery due to a fatigue inducing change (that is, increase in working hours)[23 24 37] was analysed. As noted already, the assessment of the NFR's sensitivity to detect this change is complex because no gold standard to define change in need for recovery is available. Stated differently, evidence that the NFR is able to detect change requires an intervention that produces change. At the same time, evidence that an intervention—such as the naturally occurring increase in working hours—actually produces change in need for recovery requires the availability of an instrument such as the NFR that is sensitive in detecting this change. To deal with this tricky question, the NFR's sensitivity was assessed in relative terms by comparing it with the CIS. Both scales were expected to detect the effect of the increase in working hours. However, because the NFR measures work induced fatigue, its sensitivity was expected to be more favourable. In accordance with our expectation, both scales detected the increase in working hours. Moreover, the NFR was found to discriminate most efficiently between the drivers whose working hours increased and the drivers whose working hours did not increase.

Four methodological aspects of this study should be commented upon. The first aspect involves the generalisability of the test-retest findings (ICCs). Because the NFR's test-retest reliability was confirmed in two disparate occupations (nurses and truck drivers) the generalisability of these findings seems good. It should be noted, however, that the mean ICCs are estimates of the true test-retest reliability in terms of agreement. Consequently, interpretation necessitates carefulness. To assist the reader in this respect we presented the 95% confidence intervals. These intervals indicate that even at the lower bounds, the NFR's test-retest reliability is (moderately) good in stable work environments.

The second aspect that requires consideration is the Cicchetti criteria[33] which were used to interpret the clinical relevance of the ICCs, to decide how well workers can be distinguished from each other, despite measurement error. These criteria are generally accepted to evaluate instruments that measure psychological constructs. Nevertheless, it should be noted that reliability cannot be conceived of as a property that a particular instrument does or does not possess. The adequacy of an ICC depends on several factors including the time lag, the particular population under investigation, and whether or not the instrument is used at the individual or group level. Correspondingly, as noted by Streiner and Norman,[30] there is no absolute basis for a recommendation on how much reliability is ''good enough''. The criteria we used to interpret the ICCs, therefore, should not be considered as absolute.

The third aspect that should be commented upon involves the quasi-experimental design that was used to examine the NFR's sensitivity. In particular, the increase in working hours was not planned but occurred naturally during the follow up. We did not determine the increase in working hours, which averaged only 8% (that is, five more hours per week). Likewise, we did not assign the drivers to the natural intervention group and comparison group at random. Possibly, differences between the natural intervention group and comparison group influenced the changes in NFR scores. One relevant difference involves the higher baseline NFR scores of the drivers in the natural intervention group as compared to the drivers in the comparison group. This and the small increase in working hours may have resulted in an underestimation of the effect of the quasi-experiment.

The fourth aspect that requires consideration concerns the response among the truck drivers, which, despite the publicity given to the research and three repeated mailings, was 61% at the first measurement. Unfortunately, we could not retrieve the characteristics of the non-respondents at the first measurement. However, in our opinion, further non-response did not have a strong influence on the research findings. Firstly, at the first measurement, no significant differences in need for recovery scores between non-respondents and respondents at the second measurement were observed. Secondly, in a similar questionnaire study as ours,[38] a telephone investigation among the 150 non-responding Dutch truck drivers was conducted. Of the 150 non-responding drivers, 102 drivers participated (response 68%) and no meaningful differences between the non-respondents and respondents were found.

The results carry implications for practitioners in occupational health care. The favourable test-retest reliability of the NFR indicates that the NFR can be safely used to monitor difficulties in recovering from work related exertions over time or to make inferences across longer periods of time. Used in this way, for instance as part of a periodic health survey, the detection of an increase in NFR score may forecast the manifestation of occupational diseases such as burnout, cardiovascular disease, and musculoskeletal disorders. Thereupon, actions may be undertaken to prevent these problems from appearing. The NFR's favourable sensitivity indicates that this scale may provide useful information to those engaged in the evaluation of occupational healthcare interventions directed at the reduction of difficulties in recovering from work related exertions. Cost effectiveness research should evidence whether the benefits of applying the NFR in occupational healthcare practice for these purposes (for example, forecasting and prevention of occupational diseases) outweighs the expenditure (for example, time and financial costs associated with completing and scoring the NFR).

Main messages

- When applied in stable work environments, test-retest reliability of the Need for Recovery after work scale (NFR) is good to excellent up to a period of two years.
- When applied in unstable work environments, test-retest reliability of the NFR is, as may be expected, poor to fair.
- The NFR is sensitive in detecting an increase in working hours.

Policy implications

- The NFR can be used in occupational health care to monitor difficulties workers experience in recovering from work related exertions.
- It can be used to evaluate occupational healthcare interventions.
- The practical utility of the NFR in occupational health care (for example, whether benefits outweigh the costs) for the above stated purposes requires further evaluation.

ACKNOWLEDGEMENTS

This study was supported by a grant from the Netherlands Organisation for Scientific Research (NWO), Research Programme ''Fatigue at work''.

Authors' affiliations

E M de Croon, J K Sluiter, M H W Frings-Dresen, Academic Medical Center, Universiteit van Amsterdam, Department: Coronel Institute of Occupational Health, Research Institute AmCOGG, Amsterdam, the Netherlands

REFERENCES

1 **Frankenhaeuser M**, Lundberg U, Fredrikson M, *et al.* Stress on and of the job as related to sex and occupational status in white-collar workers. *J Organ Behav* 1989;**10**:321–46.
2 **Meijman TF**, Mulder G, Van Dormolen M, *et al.* Workload of driving examiners: a psychophysiological field study. In: Kragt H, eds. *Enhancing industrial performance*. London: Taylor & Francis, 1992:245–58.
3 **Jansen NWH**, Kant IJ, van Amelsvoort LGPM, *et al.* Need for recovery from work: evaluating short-term effects of working hours, patterns and schedules. *Ergonomics* 2003;**46**:664–80.
4 **Sluiter JK**, De Croon EM, Meijman TF, *et al.* Need for recovery from work related fatigue and its role in the development and prediction of subjective health complaints. *Occup Environ Med* 2003;**60**:62i–70i.
5 **Westman M**, Eden D. Effects of a respite from work on burnout: vacation relief and fade-out. *J Appl Psychol* 1997;**82**:516–27.
6 **Sluiter JK**, Van der Beek AJ, Frings-Dresen MHW. The influence of work characteristics on the need for recovery and experienced health: a study on coach drivers. *Ergonomics* 1999;**42**:573–83.
7 **Sihm I**, Delholm G, Hansen ES, *et al.* The psychosocial work environment of younger men surviving acute myocardial infarction. *Europ Heart J* 1991;**12**:203–9.
8 **Bergqvist U**, Wolgast E, Nilsson B, *et al.* Musculoskeletal disorders among visual display terminal workers: individual, and work organizational factors. *Ergonomics* 1995;**38**:763–76.
9 **Kuiper JI**, Van der Beek AJ, Meijman TF. Psychosomatic complaints and unwinding of sympathoadrenal activation after work. *Stress Med* 1998;**14**:7–12.
10 **Magnavita N**, Bevilacqua L, Mirk P, *et al.* Work-related musculoskeletal complaints in sonologists. *J Occup Environ Med* 1999;**41**:981–8.
11 **Van Amelsvoort LGPM**, Kant IJ, Bültmann U, *et al.* Need for recovery after work and the subsequent risk of cardiovascular disease in a working population. *Occup Environ Med* 2003;**60**:83–7.
12 **Sonnentag S**. Work, recovery activities, and individual well-being: a diary study. *J Occup Health Psychol* 2001;**6**:196–210.
13 **Sluiter JK**, Frings-Dresen MHW, Van der Beek AJ, *et al.* The relation between work-induced neuroendocrine reactivity and recovery, subjective need for recovery, and health status. *J Psychosom Res* 2001;**50**:29–37.
14 **Van Veldhoven MJPM**, Meijman TF. Het meten van psychosociale arbeidsbelasing met een vragenlijst: de Vragenlijst Beleving en Beoordeling van de Arbeid (VBBA)[Questionnaire on psychosocial job demands and job stress]. Amsterdam: NIA, 1994.
15 **Alexopoulos EC**, Burdorf A, Kalokerinou A. Risk factors for musculoskeletal disorders among nursing personnel in Greek hospitals. *Int Arch Occup Environ Health* 2003;**76**:289–94.
16 **Van Amelsvoort LGPM**, Kant IJ, Bültmann U, *et al.* Need for recovery after work and the subsequent risk of cardiovascular disease in a working population. *Occup Environ Med* 2003;**60**:83i–7i.
17 **Van Amelsvoort LGPM**, Janssen N, Swaen G, *et al.* Direction of shift rotation in three-shift workers in relation to psychological health and work-family conflict. *Scand J Work Environ Health* 2004;**30**:149–59.
18 **Bekker MHJ**, Nijssen A, Hens G. Stress prevention training: sex differences in types of stressors, coping, and training effects. *Stress Health* 2001;**17**:207–18.
19 **Swaen GMH**, van Amelsvoort LGPM, Bültmann U, *et al.* Fatigue as a risk factor for being injured in an occupational accident: results from the Maastricht Cohort Study. *Occup Environ Med* 2003;**60**:88–92.
20 **De Vries J**, Michielsen HJ, Van Heck GL. Assessment of fatigue among working people: a comparison of six questionnaires. *Occup Environ Med* 2003;**60**:10i–5i.
21 **Van Veldhoven MJPM**, Broersen JPJ. Measurement quality and validity of the ''need for recovery scale''. *Occup Environ Med* 2003;**60**:3i–9i.
22 **Sparks K**, Cooper C, Fried Y, *et al.* The effects of hours of work on health: a meta-analytic review. *J Occup Organ Psychol* 1997;**70**:391–408.
23 **Harrington JM**. Health effects of shift work and extended hours of work. *Occup Environ Med* 2001;**58**:68–72.
24 **Van der Hulst M**. Long workhours and health. *Scand J Work Environ Health* 2003;**29**:171–88.
25 **Horne J**, Reyner L. Vehicle accidents related to sleep: a review. *Occup Environ Med* 1999;**56**:289–94.
26 **Vercoulen JHMM**, Swanink CMA, Fennis JFM, *et al.* Dimensional assessment of chronic fatigue syndrome. *J Psychosom Res* 1994;**38**:383–92.
27 **Beurskens AJHM**, Bültmann U, Kant IJ, *et al.* Fatigue among working people: validity of a questionnaire measure. *Occup Environ Med* 2000;**57**:353–7.
28 **Vercoulen JHMM**, Swanink CM, Zitman FG, *et al.* Randomised, double-blind, placebo-controlled study of fluoxetine in chronic fatigue syndrome. *Lancet* 1996;**347**:858–61.
29 **Bültmann U**, De Vries M, Beurskens AJ, *et al.* Measurement of prolonged fatigue in the working population: determination of a cutoff point for the Checklist Individual Strength. *J Occup Health Psychol* 2000;**5**:411–16.
30 **Streiner DL**, Norman GR. Health Measurement Scales. Oxford: OUP, 2003.
31 **Vercoulen JHMM**, Swanink CM, Fennis JFM, *et al.* Prognosis in chronic fatigue syndrome: a prospective study on the natural course. *J Neurol Neurosur Psychiatry* 1996;**60**:489–94.
32 **McGraw KO**, Wong SP. Forming inferences about some intraclass correlation coefficients. *Psychol Methods* 1996;**1**:30–46.
33 **Cicchetti DV**. Guidelines, criteria and rules of thumb for evaluating normed and standardized assessment instruments in psychology. *Psychol Assessment* 1994;**6**:284–90.
34 **Kane RL**. Outcome measures. In: Kane RL, eds. *Understanding health care outcomes research*. Aspen: Gaithersberg, MD, 1997:17–8.
35 **Murphy KR**, Davidshofer CO. Reliability: the consistency of test scores. In: Murphy KR, Davidshofer CO, eds. *Psychological testing principles and applications*. New Jersey: Prentice-Hall, Inc, 1994.
36 **Cortina JM**. What is coefficient alpha—an examination of theory and applications. *J Appl Psychol* 1993;**78**:98–104.
37 **Spurgeon A**, Harrington JM, Cooper CL. Health and safety problems associated with long working hours: a review of the current position. *Occup Environ Med* 1997;**54**:367–75.
38 **Van der Beek AJ**, Frings-Dresen MHW, Van Dijk FJH, *et al.* Loading and unloading by lorry drivers and musculoskeletal complaints. *Int J Ind Ergon* 1993;**12**:13–23.