The Reliability of a New Sleep Screening Questionnaire for Large Population-Based Studies: The Third Nord-Trøndelag Health Study

Morten Engstrøm*,[1], Siv Steinsmo Ødegård[1], Trond Sand[1,2], Lars Jacob Stovner[1,3], John-Anker Zwart[1,4] and Knut Hagen[1,3]

[1]*Department of Clinical Neurosciences; Norwegian University of Science and Technology, N-7489 Trondheim;* [2]*Department of Neurology and Clinical Neurophysiology , St. Olavs Hospital, N-7006 Trondheim;* [3]*Norwegian National Headache Centre, St. Olavs Hospital, N-7006 Trondheim and* [4]*Department of Neurology, Oslo University Hospital and University of Oslo, N-0450 Oslo, Norway*

Abstract: The aim of the present study was to evaluate the reliability of a new sleep questionnaire (sHUNT-Q) used in the Nord Trøndelag Health Study 3 (HUNT 3) performed between 2006 and 2008. Six of nine items were modified from a Norwegian version of the Karolinska Sleep questionnaire (KSQ). Overall, 50,839 (54%) out of 94,194 invited participated in HUNT 3. In a randomly selected group of participants, 297 (53%) out of 563 invited persons attended an interview by neurologists. The reliability of the sleep related questions was evaluated with Cohen´s kappa by blindly comparing the answers in the questionnaire with those from the semi-structured interview. Summary measure scores were calculated by summing the responses of the six modified KSQ questions, the three insomnia-related questions, and the two questions related to respiratory disturbance during sleep. Agreement was calculated for dichotomized sum-score indexes based on 75% and 50% percentile cut-offs. Kappa values for the individual questions ranged from 0.35 to 0.52 in 3 x 3 tables while kappa for summary indexes in 2x2 tables ranged from 0.47 to 0.62. The best agreement between the interview and the questionnaire was found for respiratory disturbance defined as a summary measure score ≥50 percentile (kappa value 0.62, 95% CI 0.51-0.74). Reliability did not depend on the time between the HUNT questionnaire and the interview. The KSQ-based summary measure scores from the new sHUNT-Q sleep screening questionnaire may become a useful tool in epidemiological studies for identifying individuals with a persistent sleep disturbance.

Keywords: Karolinska Sleep Questionnaire, epidemiology, reliability, sleep.

INTRODUCTION

Changes in sleep duration or continuity have negative impacts on various health outcomes including an increased mortality risk [1], and obstructive sleep apnea is linked to fatal and nonfatal cardiovascular events [2]. In recognition of this huge impact on health, epidemiological studies may increase our knowledge of prevalence, causes, and consequences of sleep disturbance in the general population. Although questionnaires are not optimal diagnostic tools for sleep-related disorders, the use of standardized sleep questionnaire in large-scale population-based studies is an efficient and inexpensive strategy to identify subjects at risk and quantify the occurrence of sleep disturbance.

The Karolinska Sleep Questionnaire (KSQ) is relatively commonly used in Scandinavian studies of selected samples [3-6]. A slightly modified Norwegian version including questions about snoring and sleep apnea has been used in population-based surveys [7-9]). Previously, a good correlation between the KSQ and the Karolinska sleep diary has been reported in an abstract [10]. However, despite the widespread use of KSQ, no previous studies have to our knowledge estimated the reliability of the questions.

Most KSQ questions in the Norwegian version have five response alternatives varying from "never" to "always". Within this present large multi-disciplinary epidemiological HUNT study it was not possible to include the complete questionnaire due to required space limitations. For this reason, we evaluated the reliability of a brief screening questionnaire (with a three-response option scale), as compared to the more detailed evaluation with KSQ and additional sleep questions (using the five-option scale). Clinical interviews of 297 randomly selected participants in the Nord-Trøndelag Health Study 3 (HUNT 3) were performed by neurologists on the basis of the detailed questions.

METHODS

The HUNT 3 study was carried out between October 2006 and June 2008, and all inhabitants > 13 years in the Nord-Trøndelag County were invited to participate. Among 94,194 invited adults aged 20 years or more, 50,839 (54%) responded to the first general health-questionnaire (Q1) enclosed with the invitation letter. They handed in the questionnaire when they attended a brief medical examination. During this consultation they received the second questionnaire (Q2) to be filled in and returned from home. The Q2 included nine sleep-related questions, including six slightly modified questions from the KSQ (Table **1**). A total of 39,349 subjects (42%) answered at least one of these questions (Fig. **1**).

*Address correspondence to this author at the Department of Clinical Neurosciences, PB 8905, MTFS, Norwegian University of Science and Technology, N-7489 Trondheim, Norway; Tel: +4772575090; Fax: +47 73 59 87 95; E-mail: morten.engstrom@ntnu.no
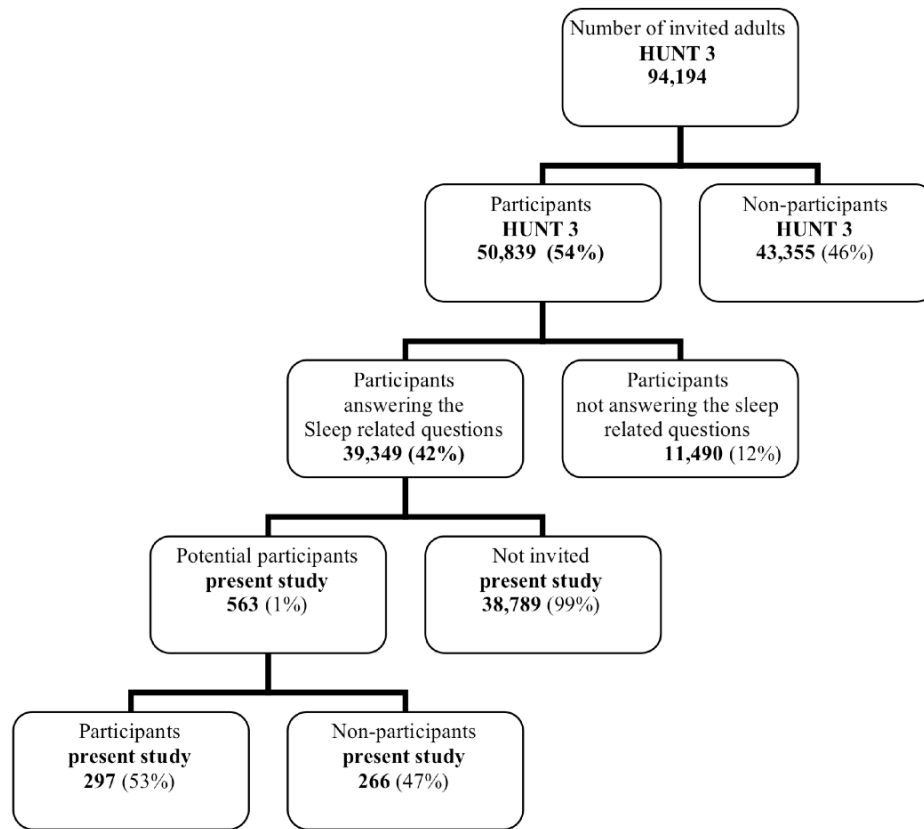
Fig. (1). Diagram of the invited population according to type of participation.

Study Population of the Present Study

The method of the present study has been described elsewhere [11]. The study was mainly designed to validate questionnaire-based headache diagnoses. A total of eight neurologists performed a semi-structured interview including questions about headache, alcohol consumption, musculoskeletal pain, and sleep.

Nord-Trøndelag County was divided into 25 study areas. In two of these areas, Verdal (14, 000 inhabitants) and Stjørdal (20,600 inhabitants), both chosen for practical reasons, the main survey was performed from September to November 2007, and from December 2007 to April 2008, respectively. The present reliability study was performed in December 2007 in Verdal and in February 2008 in Stjørdal. The participation rates in the HUNT 3 study were 52% of the whole adult population in Verdal and 50% in Stjørdal. The mean interval between answering the HUNT 3 questionnaire and the structured interview was 50 days (median 45 days, range, 9-100 days).

In HUNT 2 the participation rate was strongly age dependent, with the highest participation in the age group 60-69 years, and lowest in the age group 20-29 years [12]. To ensure acceptable balanced participation in the present study for both genders and all age groups, potential participants were picked out randomly from each subgroup in the following order: man ≤ 50 years, man > 50 years, woman ≤50 years, woman > 50 years. A total of 563 participants (279 males) were invited by mail to participate in the semi-structured interview. In the invitation letter they were informed about a telephone contact to be performed by re-

search assistants in order to make an appointment for the personal interview. The list of participants in Verdal and Stjørdal were received shortly before the time of the interview. Hence, due to lack of time, if the research assistants were unable to get in contact despite of a minimum of two attempts they were instructed to call the next person on the list.

Sleep Related Questions in Q2 (sHUNT-Q Questionnaire)

The nine different sleep-related questions used in Q2 (sHUNT-Q) are listed in Table **1**, including three questions about snoring/sleep apnea (q1-2), insomnia (q3-5), sleepiness (q6), restless legs, morning headache and night sweating during the last three months. The Norwegian KSQ includes 14 main questions about sleep problems for the last 3 months. The first six s-HUNT questions (q1-6) were selected and grammatically modified from the Norwegian KSQ. Due to the space limitation the response options were reduced from five (used in KSQ) to three in sHUNT-Q. For all the nine questions there were three response options; "seldom or never", "from time to time", and "several times per week" in sHUNT-Q.

Sleep-Related Questions in the Interview

While waiting for the interview, participants were given a retest-questionnaire (Q3) which included the same nine sleep-related questions as in sHUNT-Q. The questions were identical, but the more detailed five response alternatives were used (R1: never, R2: seldom/a few times per year, R3: sometimes/a few times a month, R4: often/several times per week, R5: always/every day). In addition, the Epworth

sleepiness scale (ESS) was recorded (score 0-3 for eight questions about the tendency to fall asleep during daytime). The summed ESS responses range from 0-24. A further categorization was required to compare the ESS results with the results of the three-level sHUNT day-time sleepiness question by kappa calculations. An ESS-index was defined as 1 for ESS less than 5, 2 for ESS 5 to 9, or 3 for ESS 10 or greater in order to assess the validity of the interview question about daytime sleepiness.

Most participants accepted the opportunity to answer the Q3 by themselves in the waiting room. The answers were then reviewed by the neurologist together with the patient. If symptoms indicated an obstructive sleep apnea syndrome, advice regarding available medical investigations and treatment options were discussed.

To clarify the participants' interpretation of the response options in Q2 (sHUNT-Q) "seldom or never" and "from time to time", the participants were also asked to quantify these options into a more detailed frequency scale with six response options (less than once a month , less than once a week, once a week, twice a week, 3-5 times a week, daily or almost daily).

Statistics

Demographic data were compared between participants and non-participants of the present study with independent samples Student's t-test for continuous variables and with the Chi-square test for categorical variables. Two-tailed estimations of significance were used, and the level of significance was set at p<0.05.

The reliability of the sleep related questions was evaluated by Cohen's kappa statistics with 95% confidence interval (CI), comparing the answers in Q2 to those made in Q3. The R1+R2 responses and the R4+R5 responses from the interview (Q3) were combined for this comparison. In the 3 x 3 tables there were one cell for each main (Q2) response option.

We constructed a KSQ-based summary measure score (KSQ score) from the six questions from KSQ, a respiratory disturbance score from the two questions about apnea and snoring (q1-2, Table **1**), an insomnia summary score from the three questions concerning insomnia (q3-5, Table **1**), and finally hypersomnia was evaluated from the response to one question (q6,Table **1**).

Each score was then dichotomized into a corresponding index twice; at the level closest to the sHUNT-Q 50% percentile and the 75% percentile respectively. The highest quartiles defined subjects with the most severe sleep related symptoms. Because data are categorical, it was impossible to achieve the exact 50% and 75% percentile cut offs. The calculated percentiles for the chosen cut offs varied between 29-50% and 77-86% respectively.

There is no absolute definitions for interpreting kappa values, but we adopted Douglas Altman's interpretation [13]. Values of kappa between 0.21 to 0.40 are suggested to indicate fair agreement, values ranging for 0.41 to 0.60 are considered moderate, and values over 0.60 are considered good or very good [13]. Statistical analysis was performed using SPSS 16 and Excel.

Ethics

The present study was an integrated part of the HUNT 3 project which was approved by the Regional Committee for Ethics in Medical research and the Norwegian Data Inspectorate.

RESULTS

Among the 563 potential participants (279 males), 171 were not reached by at least two telephone calls. A total of 297 out of the 392 persons contacted by telephone participated in the clinical interview (53% out of the total invited group, Fig. **1**).

Compared to the 266 non-participants, the 297 participants were older (mean age 52.3 vs. 48.6 years, p=0.004) and slightly more likely to be men (51% vs. 47%, p=0.37).

Table 1.   Agreement Between HUNT-Questionnaire (Q2) and Interview for Sleep-Related Questions (Kappa Values with 95% CI for 3 x 3 Tables)

	Kappa (95% CI)
How often during the last 3 months has it occurred that you:	
Snore loud and embarrassingly? (q1)	0.46 (0.38-0.54)
Have periods with breathing pauses during the night? (q2)	0.52 (0.43-0.60)
Have difficulties with falling asleep at night?(q3)	0.44 (0.35-0.52)
Wake up repeatedly during the night?(q4)	0.38 (0.31-0.46)
Awake early and can not fall asleep again?(q5)	0.35 (0.26-0.44)
Feel sleepy during the day? (q6)	0.38 (0.30-0.45)
Have bothersome sweat during the night?	0.43 (0.34-0.53)
Wake up with headache?	0.45 (0.35-0.55)
Have discomfort, tingling or prickling in the legs?	0.45 (0.36-0.53)

q1-6: Grammatically slightly modified questions from the Karolinska sleep questionnaire.

Table 2.   **Reliability of Sleep Disturbance Indexes in HUNT (Q2) Compared to the More Detailed Questionnaire in the Interview. Two-By Two Tables and Kappa Values (95% Confidence Intervals in Parentheses). Data for 75 Percentile and 50 Percentile Cutoffs**

	75% percentile limit			50% percentile limit		
	HUNT questionnaire			HUNT questionnaire		
	No	Yes	Kappa (95% CI)	No	Yes	Kappa (95% CI)
Interview questionnaire						
Respiratory disturbance index (q1-2)						
No	188	19		107	24	
Yes	28	46	0.55 (0.43-0.67)	29	121	0.62 (0.51-0.74)
Insomnia index (q3-5)						
No	202	13		86	41	
Yes	37	34	0.47 (0.36-0.58)	27	132	0.51 (0.40-0.63)
Hypersomnia index (q6)						
No	213	12		67	69	
Yes	33	30	0.48 (0.37-0.59)	16	136	0.40 (0.29-0.50)
Karolinska questionnaire index (q1-6)						
No	190	17		110	32	
Yes	33	40	0.50 (0.39-0.62)	28	110	0.57 (0.45-0.69)

q1-6: the six questions from the Karolinska sleep questionnaire.

No significant difference was found between participants and non-participants with regard to mean score of the six question from the KSQ (1.60 versus 1.56, p=0.16), prevalence of KSQ score ≥75 percentile (25.3% for participants and 21.9% for non-participants p=0.18), bodily pain lasting ≥ 6 months (40% vs. 36%, p=0.30), self-reported good or very good health (75% vs. 79%, p=0.25), and regarding proportion of working individuals (76% v. 82%, p=0.12).

Reliability

As demonstrated in Table **1**, the kappa values for single questions ranged from 0.35 to 0.52 in the 3 x 3 tables. The questions concerning respiratory disturbance were most reliable, but kappa values did not differ significantly between questions.

The best agreement were found for a positive respiratory disturbance index (kappa value 0.62, 95% CI 0.51-0.74) and a positive total sleep disturbance index (kappa value 0.60, 95% CI 0.48 - 0.71), defined as scores ≥ the 50% percentile (Table **2**). Agreement for indexes did not differ significantly between men and women, or between those below and over 50 years of age (data not reported).

There were no significant differences in reliability between those with the longest test-retest time and those with the shortest test-retest time (Table **3**).

295 subjects had scored ESS during the interview (median 5, range 0-18). The change-corrected agreement between daytime sleepiness (q6 in Table **1**) and the ESS-index in a 3x3 table was 0.27 (95% CI 0.19-0.35) (Table **4**).

Interpretation of the Response Options

During the interview the participants were asked what they actually meant when they had stated that sleep related symptoms occurred "seldom or never". It appeared that this meant less than once a month for the vast majority of the participants (87%) (Table **5**). Similarly, the majority (88%) reported that "from time to time" meant approximately once a week, whereas 98% answered that the option "several times per week" was twice a week or more.

DISCUSSION

The reliability of a new brief sleep screening questionnaire (sHUNT-Q) used in the large Norwegian epidemiological HUNT-3 study was evaluated by blindly comparing the answers in sHUNT-Q with those in a more detailed questionnaire. The best agreement between the interview and the questionnaire was found for respiratory disturbance as defined by a summary measure score.

A similar approach, computing sleep index scores from selected KSQ items, has been used in several previous studies [4, 6, 14-16]. Each summary measure score was dichothomized by the 50% and 75% percentile (50% and 75% index) because no validated cut-off score had been reported in those studies [4, 6, 14-16].

The present results indicated that the sHUNT-Q index scores may be useful screening tools for identifying individuals with a significant sleep disturbance. Reliability for the hypersomnia question was also fairly good, particularly for those with most frequent symptoms (75% percentile cut-off).

Table 3.   **Reliability of Sleep Indexes Related to the Time Interval Between the Two Questionnaires: Kappa Values (95% CI)**

75% percentile limit	Less than 45 days (n=138)	Equal or more than 45 days (n=159)
Respiratory disturbance index (q1-2)	0.57 (0.40-0.74)	0.54 (0.38-0.70)
Insomnia index (q3-5)	0.39 (0.23-0.55)	0.52 (0.36-0.67)
Hypersomnia index (q6)	0.50 (0.34-0.66)	0.46 (0.31-0.62)
Karolinska questionnaire index (q1-6)	0.41 (0.24-0.57)	0.55 (0.39-0.71)
50% percentile limit		
Respiratory disturbance index (q1-2)	0.62 (0.44-0.79)	0.63 (0.47-0.78)
Insomnia index (q3-5)	0.46 (0.29-0.63)	0.56 (0.40-0.72)
Hypersomnia index (q6)	0.38 (0.22-0.54)	0.40 (0.26-0.55)
Karolinska questionnaire index (q1-6)	0.66 (0.49-0.83)	0.49 (0.33-0.65)

Table 4.   **The Agreement Between the Daytime Sleepiness Question and Epworth Sleepiness Scale (ESS) Score**

	Daytime sleepiness (interview)		
	Never/seldom	Sometimes	Often
Epworth sleepiness scale (interview)			
ESS<5	81	28	21
ESS 5-9	45	58	24
ESS >9	11	7	20

Kappa=0.27 (95% CI 0.19-0.35).

Table 5.   **The Interpretation of the Questionnaire Response Alternatives "Seldom or Never", "from Time to Time", and "Several Times Per Week" Given Six Different Frequency Options in the Interview**

Frequency options	Seldom or never % (n)	From time to time % (n)	Several times a week % (n)
Daily or almost daily	0	0	10.9% (29)
3-5 times per week	0	0.8% (2)	57.9% (154)
Twice a week	0	11.1% (29)	29.3% (78)
Once a week	1.5% (4)	35.9% (94)	1.9% (5)
< once a week	11.4% (30)	51.5% (135)	0
< once a month	87.1% (230)	0.8% (2)	0
Total	100.0% (264)	100.0% (262)	100.0% (266)

The agreement for each separate question was more moderate (kappa values between 0.35-0.52). These kappa values were somewhat lower than those found for a Norwegian version of ESS [17]. It should also be emphasized that the time span between the Q2 to the interview was longer in the present study (mean of 50 days versus 18 days).

Our questionnaire included only one question about daytime sleepiness. This question could possibly be useful to screen for hypersomnia although "sleepiness" in Norwegian is a wide term that probably target some patients with fatigue as well as those with a propensity to fall asleep during daytime. The questions about falling asleep during the day, included in the Norwegian version of the Karolinska sleep questionnaire, could not be included in the sHUNT-Q due to

space restrictions. In spite of this limitation, a fair agreement (validity) between sleepiness and ESS was demonstrated in the present study. It should also be noted that excessive daytime sleepiness (EDS) and hypersomnia are difficult to evaluate. Although ESS is widely used to evaluate EDS and hypersomnia, the correlation between ESS and objective measures like MSLT is moderate [18, 19]. This may partly explain why the agreement was only fair in the present study.

A long time frame between the questionnaire and the interview (median 45 days in the present study) may reduce the agreement between responses, because sleep related symptoms may vary over time in some subjects. Thus, if the time span between test-retest is long, there could be a poten-

tial risk of measuring symptom variation rather than questionnaire reliability. This may be considered as a weakness of our study. However, it should be emphasized that sHUNT-Q (and KSQ) asked for complaints during the last three months. It was accordingly a considerable overlap regarding the time periods most subjects were asked to evaluate. In addition, a long period minimizes the chance of recall bias, i.e. that the participants remember what they actually answered to the sHUNT questions. Thus, a relatively long time frame between the questionnaire and the interview may be considered as an advantage for independent scoring of long-lasting problems. Lastly no significant differences in kappa values between individuals with shortest and longest retest time were found. Thus, most likely the reported kappa values mainly reflect the reliability of the questionnaire. In accordance, a high agreement between an interview and a questionnaire has previously been found for chronic headache [20].

The vast majority of the participants reported that sleep related symptoms which occurred "seldom or never" was equal to occurrence less than once a month, whereas "from time to time" occurred with a maximum of once a week. Thus, the participants quantified the somewhat vague response options in sHUNT-Q into a frequency scale in a relatively uniform way. However, we can not rule out that there has been some kind of education of the participants in this respect during the HUNT surveys because the five response options included in the Norwegian KSQ and the Q3 questionnaire also included some frequency-scale guidance, as reported in Methods.

To our best knowledge no previous studies have estimated the reliability of KSQ questions in a random population-based study. It should be emphasised that less than half of the invited adults (42%) responded to the sleep related questions in HUNT 3, and that our results are only valid for these responders. Only 53% of the random sample of invited participants in this study was interviewed. A selective participation due to sleep disturbance seems even though less likely since the mean KSQ-based index score did not differ between participants and non-participants. This finding is of particular interest, because such a comparison between participants and non-participants rarely can be performed.

In conclusion, the best reliability between the sHUNT-Q questionnaire and the retest questionnaire with an interview was found for individuals with high burden of insomnia, obstructed respiration during sleep, and overall sleep disturbance defined as KSQ summary measure score of ≥ 75 percentile. The present results suggest that the KSQ-based index score may become a useful tool for identifying individuals with persistent sleep disturbance in epidemiological studies.

REFERENCES

[1]　Roth T. Sleep and society. Sleep Med 2009; 10: 1-2.
[2]　Selim B, Won C, Yaggi HK. Cardiovascular consequences of sleep apnea. Clin Chest Med 2010; 31: 203-20.
[3]　Carlsson ME. Fatigue in relatives of palliative patients. Palliat Support Care 2009; 7: 207-11.
[4]　Sterud T, Hem E, Ekeberg O, Lau B. Health problems and help-seeking in a nationwide sample of operational Norwegian ambulance personnel. BMC Public Health 2008; 8: 3.
[5]　Lineweber C, Kecklund G, Janszky I, Akerstedt T, Orth-Gomer K. Snoring and progression of coronary artery disease: The Stockholm Female Coronary Angiography Study. Sleep 2004; 27: 1344-9.
[6]　Lineweber C, Kecklund G, Orth-Gomer K. Prediction of cardiocerebrovascular and other significant disease from disturbed sleep and work strain. Scand J Work Environ Health 2007; 33: 215-22.
[7]　Ursin R, Bjorvatn B, Holsten F. Sleep duration, subjective sleep need, and sleep habits of 40- to 45-year-olds in the Hordaland Health Study. Sleep 2005; 28: 1260-9.
[8]　Oyane NM, Ursin R, Pallesen S, Holsten F, Bjorvatn B. Self-reported seasonality is associated with complaints of sleep problems and deficient sleep duration: the Hordaland Health Study. J Sleep Res 2008; 17: 63-72.
[9]　Forenklet søvnspørreskjema. Available from: http://www.helse-bergen.no/avd/SOVno/sporreskjema/utredning.htm.
[10]　Axelsson J, Kecklund G, Åkerstedt T, Ekstedt M, Menenga J. A comparison of the Karolinska sleep questionnaire and the Karolinska sleep diary: a methodological study. Eur Sleep Res Soc JSR 11 2002; 11: 8.
[11]　Hagen K, Zwart JA, Aamodt AH, *et al*. A face-to-face interview of participants in HUNT 3: the impact of the screening question on headache prevalence. J Headache Pain 2008; 9: 289-94.
[12]　Holmen J, Midthjell K, Kruger Ø, *et al*. The Nord-Trøndelag Health Study 1995-97 (HUNT 2): Objectives, contents, methods and participation. Norsk Epidemiol 2003; 13(1): 19-32.
[13]　Altman DG. Practical statistics for medical research. Chapman & Hall 1991.
[14]　Akerstedt T, Ingre M, Broman JE, Kecklund G. Disturbed sleep in shift workers, day workers, and insomniacs. Chronobiol Int 2008; 25: 333-48.
[15]　Akerstedt T, Knutsson A, Westerholm P, Theorell T, Alfredsson L, Kecklund G. Sleep disturbances, work stress and work hours: a cross-sectional study. J Psychosom Res 2002; 53: 741-8.
[16]　Lineweber C, Kecklund G, Akerstedt T, Janszky I, Orth-Gomer K. Snoring and the metabolic syndrome in women. Sleep Med 2003; 4: 531-6.
[17]　Beiske KK, Kjelsberg FN, Ruud EA, Stavem K. Reliability and validity of a Norwegian version of the Epworth sleepiness scale. Sleep Breath 2009; 13: 65-72.
[18]　Fong SY, Ho CK, Wing YK. Comparing MSLT and ESS in the measurement of excessive daytime sleepiness in obstructive sleep apnoea syndrome. J Psychosom Res 2005; 58: 55-60.
[19]　Miletin MS, Hanly PJ. Measurement properties of the Epworth sleepiness scale. Sleep Med 2003; 4: 195-9.
[20]　Hagen K, Zwart JA, Aamodt AH, *et al*. The validity of questionnaire-based diagnoses: the third Nord-Trondelag Health Study 2006-2008. J Headache Pain 2010; 11(1): 67-73.