Original article

Scand J Work Environ Health 2008;34(2):113–119

Test–retest reliability and validity of self-reported duration of computer use at work

by Stefan IJmker, MSc,[1, 2, 3] Josien NM Leijssen, MSc,[3] Birgitte M Blatter, PhD,[1, 3] Allard J van der Beek, PhD [1, 2] Willem van Mechelen, PhD,[1, 2] Paulien M Bongers, PhD [1, 2, 3]

IJmker S, Leijssen JNM, Blatter BM, van der Beek AJ, van Mechelen W, Bongers PM. Test–retest reliability and validity of self-reported duration of computer use at work. *Scand J Work Environ Health.* 2008;34(2):113–119.

Objectives The aims of this study were to evaluate the test–retest reliability and the validity of self-reported duration of computer use at work.

Methods Test–retest reliability was studied among 81 employees of a research department of a university medical center. The employees filled out a web-based questionnaire twice with an in-between period of 14 days. Validity was studied among a group of 572 office workers who participated in an epidemiologic field study. A software program recorded the duration of computer use at work during the 3 months preceding the questionnaire.

Results The percentages of agreement for test–retest reliability were 75% [95% confidence interval (95% CI) 64–84] for total computer use and 67% (95% CI 55–77) for mouse use. The percentages of agreement between self-report and registration were 18% (95% CI 15–21) for total computer use and 16% (95% CI 13–19) for mouse use. Misclassification was mainly nondifferential in nature, since all of the evaluated subgroups showed at least 75% misclassification.

Conclusions The use of self-reports lead to the misclassification of exposure to computer use for more than 80% of all persons. This misclassification is predominantly nondifferential in nature and can only partly be explained by limited test–retest reliability.

Key terms occupational exposure; questionnaire; reproducibility of results; validation study.

Most longitudinal studies that address the association between the duration of computer use at work and musculoskeletal symptoms and disorders have used questionnaires to measure the duration of computer use at work. However, insufficient information is available regarding the reliability and validity of the questionnaires used in these studies (1). This lack of information may hamper the determination of a potential threshold dose (ie, number of hours of computer use at work) that may be relevant in the prevention of musculoskeletal symptoms and disorders at work (2).

Previous studies have shown that computer workers generally overestimate their duration of computer use at work (2–5). In epidemiologic studies, the overestimation by individual participants leads to the misclassification of exposure (ie, exposure values that are too high are assigned to participants). Depending on the nature of the misclassification, the strength of the association can be underestimated (most likely in the case of nondifferential misclassification) or either underestimated or overestimated (in the case of differential misclassification).

The literature on this topic suggests that differential misclassification is likely. Faucett & Rempel (3) reported more overestimation among younger workers and workers with less psychological workload. Heinrich et al (4) reported less overestimation among workers who worked fewer hours with the computer (according to their questionnaire). Homan & Armstrong (2) reported less overestimation among workers with longer durations of computer use (according to work sampling) than among those with shorter durations of computer use. In addition, managers seem to overestimate the duration to

1 Body@Work TNO VUmc, Research Centre Physical Activity, Work and Health, Amsterdam, Netherlands.
2 Department of Public and Occupational Health, EMGO Institute, VU University Medical Centre, Amsterdam, Netherlands.
3 TNO Quality of Life, Hoofddorp, Netherlands.

Correspondence to: Dr A van der Beek, Body@Work TNO VUmc, EMGO Institute, Van der Boechorststraat 7, 1081 BT Amsterdam, Netherlands. [E-mail: a.vanderbeek@vumc.nl]

a less extent than nonmanagers do (6). Misclassification was not related to symptom status in most studies that investigated the issue (4, 7). However, small sample sizes may have obscured actual differences. A recent large-scale study showed that prevalent arm symptoms can influence the self-reported duration of computer use (8).

This study aimed at exploring the following two aspects related to the misclassification of the duration of computer use at work, as obtained by self-reports: test–retest reliability and validity. In addition, this study evaluated the relative importance of differential and nondifferential misclassification. A limited test–retest reliability of questionnaires may contribute to misclassification. In addition, limited validity between self-reporting and objective registration may contribute to misclassification. All studies to date have used continuous self-report estimates of the duration of computer use at work (ie, the number of hours or percentage of time). In general, self-reports lack measurement precision (9), and this lack offers a challenge to the collection of continuous self-report data. Therefore, we constructed a self-report measurement with a categorical response scale. Using a categorical instead of a continuous measurement scale may possibly reduce misclassification.

The first study question addressed the degree of test–retest reliability of the self-report measurement of the duration of computer use at work. The second question focused on the agreement between the self-reported duration of computer use at work and the registered duration of computer use at work.

Table 1. Characteristics of the study populations for the evaluation of test–retest reliability (N=81) and validity (N=572).

Characteristic	Test-retest		Validity	
	N	%	N	%
Gender				
Male	26	32	356	52
Female	55	68	326	48
Age				
<30 years	27	33	71	12
30–39 years	25	31	183	32
40–49 years	16	20	189	33
≥50 years	13	16	124	22
Missing	–	–	5	1
Self-reported total computer use at work				
0–<2 hours/day	3	4	2	1
2–<4 hours/day	12	15	70	12
4–<6 hours/day	30	37	254	44
6–<8 hours/day	35	43	240	42
≥8 hours/day	1	1	6	1
Self-reported mouse use at work				
0–<2 hours/day	13	16	61	11
2–<4 hours/day	26	32	147	25
4–<6 hours/day	33	41	229	40
6–<8 hours/day	8	10	132	23
≥8 hours/day	1	1	3	1
Arm, wrist or hand symptoms [a]				
No	74	91	478	84
Yes	7	9	94	16
Neck or shoulder symptoms [a]				
No	71	88	414	72
Yes	10	12	158	28

[a] Regular or long-lasting symptoms in the previous 3 months.

Study population and methods

Study population and design

We used two different study populations to answer the study questions. To answer the first question (on the test–retest reliability of the questionnaire), we included a group of employees of a research department of a university medical center in the Netherlands. The participants filled out a web-based questionnaire twice with an in-between period of 2 weeks. To answer the second question (on the validity of the questionnaire), we included a group of office workers who had filled out the 1-year follow-up questionnaire of the PROMO study [for more details, see the report by IJmker et al (10)], and for whom software registration of the duration of computer use at work was available for the 3 months preceding the baseline questionnaire of the PROMO study. Table 1 presents the characteristics of the two study populations. In the test–retest reliability study, 84 participants filled out the first questionnaire, and 81 participants filled out both questionnaires. In the validity study, 572 participants filled out the questionnaire.

Measurements

The questionnaire included the following question on the duration of computer use at work: "How many *hours per day* do you use your computer during your *work at the office* (including reading from the screen)?" The question had the following seven response categories: never, 0–1, 1–2, 2–4, 4–6, 6–8, >8 hours/day. The same question was used for the duration of mouse use: "How many *hours per day* do you use your mouse during your *work at the office*?" The participants were invited by e-mail to fill out the web-based questionnaire. The e-mail contained a hyperlink that directed the participants to the web page with the questionnaire.

Data on computer use at work were collected with the software program WorkPace, version 3.0 (Niche Software Ltd/ErgoDirect, Christchurch, New Zealand). The program was installed on the personal computers, and data were periodically sent to a central folder on the computer network. Data were stored for each person as cumulative totals for each calendar day. The program

estimated the duration of computer use on the basis of the duration of the time interval between two consecutive active events (ie, keying, mouse clicking, or mouse movements). If a participant hit a key, moved or clicked the mouse within 30 seconds of previously hitting a key or moving or clicking the mouse, then the "interevents period" (in seconds) was stored as a usage period of total computer use. If the threshold time of 30 seconds was exceeded, then the elapsed time period between two usage periods was stored as a break from total computer use. This threshold value for total computer use reflects the use of the keyboard or mouse, reading from the screen, and performing combinations of these activities. The threshold time for mouse use, which reflects clicking or moving the mouse (and not reading from the screen), was 5 seconds. Previous research has shown good agreement between the WorkPace estimate and systematic observation for the duration of total computer use (ie, using the keyboard or mouse, reading from the screen, or performing combinations of these activities). The average duration of total computer use, based on the WorkPace estimates, was within 10% of the average duration of total computer use, which was based on systematic observation (6, 11).

The mean daily duration of the registered duration of total computer use and mouse use was calculated by dividing the cumulative duration of registration during 3 months by the number of days for which the software program recorded activity (ie, at least 1 second of use). The data from the participants who worked at least 2 days per week at another location of their organization where no recordings could be made and the data from the participants who shared a computer account with a colleague were excluded from the analyses. In addition, data were excluded if the number of recorded days was less than 70% of the number of actual workdays.

Statistical analysis

We calculated percentages of agreement, percentages of misclassification (ie, 1 – the percentage agreement), and subdivisions of percentage misclassification (ie, difference of 1 and ≥2 categories) to evaluate the test-rest reliability of the self-report measure.

We performed four analyses to answer the second study question concerning validity. At first, we recoded both self-reported and registered data into the same categories (ie, 0 to <2, 2 to <4, 4 to <6, 6 to <8, and ≥8 hours per day). Then we calculated, for computer use and mouse use separately, the percentages of agreement, percentages of misclassification (ie, 1% agreement), and subdivisions of the percentage misclassification (ie, difference of 1 and ≥2 categories, underestimation and overestimation). In addition, we dichotomized the self-reported and registered data with cut-off values of 2, 4, and 6 hours per day and calculated the percentage of agreement, the sensitivity, and the specificity. Then we performed separate analyses for subgroups of the participants to investigate the potential effect modification (ie, the differential misclassification of exposure). On the basis of previous findings in the literature (2–4, 6–8), we defined the subgroups according to symptom status (ie, the presence of regular or long-lasting arm–wrist–hand symptoms or neck–shoulder symptoms in the previous 3 months), self-reported computer use at work (cut off: 4 hours/day), gender, age, cognitive demands, decision authority, task variation, effort, reward, and variation in the registered computer use (cut-off: median). In addition, we calculated the difference between the self-reported duration of computer use and the registered duration of computer use for each participant by using the middle score of the self-report categories as the estimate of self-reported duration of computer use. We used this data to plot the difference between the self-reported duration of computer use and the registered duration of computer use against the registered duration of computer use.

We used SPSS (version 12.0, SPSS Inc, Chicago, IL, USA) for the data transformation. Summary statistics and 95% confidence intervals were calculated using web-based applications (http://faculty.vassar.edu/lowry/kappa.html and http://faculty.vassar.edu/lowry/clin1.html).

Results

Test–retest reliability

Tables 2 and 3 show the results of the test–retest analysis of the self-reported duration of computer use at work. The percentage agreements were 75% and 67% for total computer use and mouse use, respectively. The percentage of misclassification was primarily restricted to neighboring categories [ie, a difference of 1 category between the two questionnaires (22% for total computer use and 33% for mouse use)].

Validity

Tables 4 and 5 show the agreement between self-reported and registered computer use at work. The percentage agreements were 18% and 16% for total computer use and mouse use, respectively. Almost all of the misclassification was due to overestimation in the self-reports. For the total computer use at work, 51% of all of the participants overestimated their computer use by 1 category, and 30% did so by at least 2 categories (ie, at least 2 hours/day). For mouse use at work, 35% of all of the participants overestimated their mouse use by 1 category, and 48% did so by at least 2 categories (ie, at least 2 hours/day).

Table 2. Results of the test–retest analysis of self-reported total computer use at work—cross table displaying frequencies and summary statistics (N=81). [a] (h/d = hours per day, 95% CI = 95% confidence interval)

Questionnaire 1	Questionnaire 2				
	0-<2 h/d	2-<4 h/d	4-<6 h/d	6-<8 h/d	≥8 h/d
0–<2 h/d	–	1	2	–	–
2–<4 h/d	1	8	3	–	–
4–<6 h/d	–	4	25	1	–
6–<8 h/d	–	–	8	27	–
≥8 h/d	–	–	–	–	1

[a] Percentage of agreement = 75 (95% CI 64–84); percentage of misclassification = 25 (95% CI 16–36); percentage of misclassification for 1 category of difference = 22 (95% CI 14–33); and percentage of misclassification for ≥2 categories of difference = 3 (95% CI 0–10).

Table 3. Results of the test–retest analysis of self-reported mouse use at work—cross table displaying frequencies and summary statistics (N=81). [a] (h/d = hours per day, 95% CI = 95% confidence interval)

	Questionnaire 2				
	0-<2 h/d	2-<4 h/d	4-<6 h/d	6-<8 h/d	≥8 h/d
0–<2 hours/day	10	3	–	–	–
2–<4 hours/day	2	15	9	–	–
4–<6 hours/day	–	3	23	7	–
6–<8 hours/day	–	–	3	5	–
≥8 hours/day	–	–	–	–	1

[a] Percentage of agreement = 67 (95% CI 55–77); percentage of misclassification = 33 (95% CI 23–45); percentage of misclassification for 1 category of difference = 33 (95% CI 23–45); and percentage of misclassifi-

Table 4. Agreement between self-reported and registered total computer use at work—cross table displaying frequencies and summary statistics (N=572). [a] (h/d = hours per day, 95% CI = 95% confidence interval)

Self-report	Registration				
	0-<2 h/d	2-<4 h/d	4-<6 h/d	6-<8 h/d	≥8 h/d
0–<2 h/d	2	–	–	–	–
2–<4 h/d	16	51	3	–	–
4–<6 h/d	22	185	47	–	–
6–<8 h/d	9	138	92	1	–
≥8 h/d	–	4	2	–	–

[a] Percentage of agreement = 18 (95% CI 15–21); percentage of misclassification = 82 (95% CI 74–85); percentage of misclassification for 1 category of underestimation by self-report = 1 (95% CI 0–2); percentage of misclassification for 1 category of overestimation by self-report = 51 (95% CI 47–55); and percentage of misclassification for ≥2 categories of overestimation by self-report = 30 (95% CI 27–35).

Table 5. Agreement between self-reported and registered mouse use—cross table displaying frequencies and summary statistics (N=572). [a] (h/d = hours per day, 95% CI = 95% confidence interval)

Self-report	Registration				
	0-<2 h/d	2-<4 h/d	4-<6 h/d	6-<8 h/d	≥8 h/d
0–<2 h/d	55	6	–	–	–
2–<4 h/d	111	36	–	–	–
4–<6 h/d	141	88	–	–	–
6–<8 h/d	60	71	1	–	–
≥8 h/d	2	1	–	–	–

[a] Percentage of agreement = 16 (95% CI 13–19); percentage of misclassification = 84 (95% CI 81–87); percentage of misclassification for 1 category of underestimation by self-report = 1 (95% CI 0–2); percentage misclassification for 1 category of overestimation by self-report = 35 (95% CI 31–39); and percentage of misclassification for ≥2 categories of overestimation by self-report = 48 (95% CI 44–52).

Table 6. Agreement between self-reported and registered computer use at work for the dichotomized data—summary statistics (N=572). (95% CI = 95% confidence interval)

Cut-off	Percentage of agreement		Sensitivity		Specificity	
	%	95% CI	%	95% CI	%	95% CI
Total computer use at work						
2 hours/day	92	89–94	92	89–94	100	19–100
4 hours/day	37	33–41	28	25–33	96	87–99
6 hours/day	57	53–61	0	0–3	100	99–100
Mouse use at work						
2 hours/day	44	40–48	39	34–43	90	79–96
4 hours/day	37	33–41	0	0–2	100	97–100
6 hours/day	76	73–80	0	0–3	100	98–100

The dichotomized data showed higher percentages of agreement. For total computer use, the highest percentage of agreement was present at the cut-off of 2 hours/day (92%), and for mouse use the optimal cut-off was 6 hours/day (76%). [See table 6.] In general, the sensitivity values were low due to the overestimation in the self-reports.

The managers and the participants who reported low task variation, arm–wrist–hand symptoms, or less than 4 hours of computer use at work a day showed a higher agreement than the nonmanagers and participants who reported high task variation, no arm, wrist or hand symptoms, or at least 4 hours of computer use at work per day (data not shown). However, the agreement between the self-reported and registered data was low for all the subgroups. The percentages of agreement were below 25% [ie, the percentages of misclassification were at least 75% for all of the subgroups (data not shown)]. Figure 1 shows that the difference between the self-reported and registered duration of computer use at work decreased as the duration of registered computer use at work increased.

Discussion

In this study, we investigated the test–retest reliability and the validity of the self-reported duration of computer use at work. Imperfect test–retest reliability introduced misclassification of exposure for 25% (total computer use) to 33% (mouse use) of all the participants. Misclassification in the test–retest analysis was restricted to one category (ie, a difference of up to 2 hours). The validity study showed that the use of self-reports led to misclassification for 82% (total computer use) to 84% (mouse use) of all the participants. This misclassification was a result of overestimation in the self-reports in almost all cases. Altogether 30% (for total computer use) to 48% (for mouse use) of all the participants overestimated their computer use by more than 2 hours a day. The misclassification was mainly nondifferential in nature since, in all of the evaluated subgroups, at least 75% misclassification was present.

Our results of the test–retest reliability study are in line with the results of Karlqvist (12), who found correlations of 0.92 and 0.75 for the estimated percentage of total worktime spent in work with a visual display unit and mouse use during such work. Self-reports seem to be prone to a limited amount of random measurement error, which can express itself as misclassification if categorical data are used, as in the current study.

Our results also seem to be in line with the results of other studies that compared self-reported data on the duration of computer use at work and objective measurements. All of the published studies found that workers generally overestimated their duration of computer use at work in comparison with objective measurements (2–5). The amount of overestimation between previous studies and the current study is difficult to compare due to the use of continuous data in previous studies versus categorical data in our study, and variation between studies in the time window over which comparisons have been made. Contrary to what was expected, the use of categorical data did not seem to improve the agreement between the self-reports and the objective measurements to a large extent, since 30% (total computer use) to 48% (mouse use) of all the participants overestimated their computer use for more than 2 hours a day in this study.

As already mentioned in the introduction, Heinrich et al (4) reported less overestimation among workers who worked fewer hours with the computer (according to their questionnaire). Homan & Armstrong (2) reported less overestimation among workers with longer durations of computer use (according to work sampling) than among workers with shorter durations of computer use. These observations seem to be in contradiction. However, our data showed that both phenomena could be present at the same time (figure 1). It should be noted that an artificial "ceiling effect" may be present. Workers cannot report longer durations of computer use than the number of hours they work. Despite the aforementioned trend, the differences in the agreement between workers who had a long duration of computer use versus those who had a short duration of computer use were small. For both groups the percentage of agreement was lower than 25%. In general, nondifferential misclassification seems to play a larger role than differential misclassification with respect to all of the variables evaluated in this study.

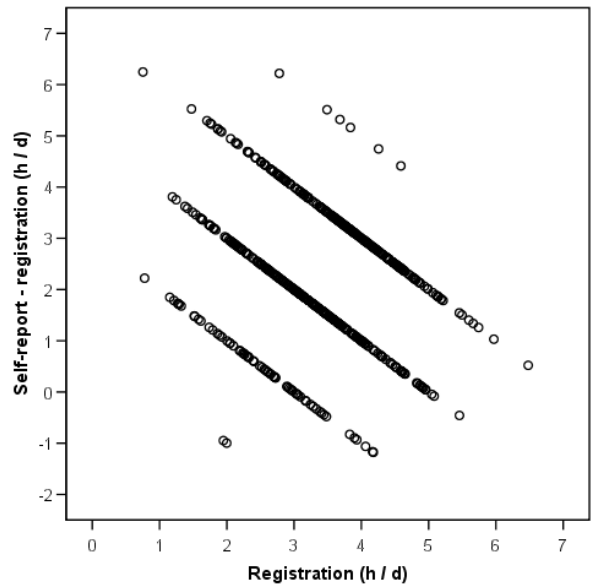

Figure 1. Scatter plot of differences between self-reported and registered total computer use at work (Y-axis) and registered total computer use at work (X-axis) [N=572]. (h/d = hours per day)

The strengths of this study are the large sample size compared with that of most published studies, the extended time period over which objective data were collected, and the evaluation of both test–retest reliability and validity in one paper. Moreover, we used a "gold standard" with known measurement characteristics (6, 11).

Several factors related to the design of this study may have biased our findings. At first, in the validity study, we compared the self-reports with the average daily duration of computer use, based on software registration from the preceding 3 months before the questionnaire. It is possible that the participants rated their computer use over a shorter recall period. However, the estimates of the daily duration of computer use were similar for the preceding week and the preceding month; it was therefore unlikely that the time period over which the daily duration of computer use was calculated biased our findings to a significant extent. In addition, it is still unclear how the participants interpreted "computer use at work". It is possible that they interpreted it as including sitting at the desk. However, the Spearman correlation

between the self-reported duration of sitting at the desk and the self-reported duration of computer use has been reported to be 0.30 (13). It follows that this explanation covers only a small part of the overestimation in the self-reports.

Another factor that may have been responsible for the overestimation in the self-reports is inaccuracy of the software program. However, previous studies in populations with various job functions have shown that the average duration of registration was within 10% of the observed duration of computer use (6, 11). It follows that the inaccuracy of the software program may explain part of the overestimation of one category by self-report [ie, 0–2 hours/day, observed for 35–51% of all the participants (tables 2 and 3)]. However, we think that it is unlikely that the overestimation of at least two categories can be explained by limitations in the software program [ie, overestimation of at least 2 hours/day, observed for 30% to 48% of all the participant (tables 4 and 5)].

Dichotomization of self-reported data improved the agreement with the registered data. However, the relatively high percentage agreement for the 2-hour and 6-hour cut-off for total computer use (ie, 92% and 57%, respectively) and the 6-hour cut-off for mouse use (ie, 76%) were biased due to skewed distributions. Only two participants reported less than 2 hours of total computer use, and one participant had a registration for more than 6 hours of total computer use. No participants had a registration of more than 4 hours per day of mouse use. It follows that only the 4-hour cut-off for total computer use and the 2-hour cut-off for mouse use should be used in epidemiologic studies to determine associations between self-reported computer use at work and musculoskeletal symptoms and disorders. Misclassification was still present for these cut-offs for most of the participants (ie, 63% for total computer use and 56% for mouse use).

In several reviews, it has been reported that an increasing duration of computer use at work is associated with an increased risk of musculoskeletal symptoms and disorders (1, 7). Most studies draw their conclusions on the basis of the self-reported duration of computer use at work. Given the results of our study, it is likely that the use of self-reports leads to nondifferential misclassification and, consequently, underestimation of the strength of the association. In addition, the increased risk may be present at a lower duration of computer use at work. Future studies should increase our knowledge on the association between the duration of computer use at work and musculoskeletal symptoms and disorders by including more precise measurements of the duration of computer use at work. One possibility is to use event-driven diaries in which workers record the actual tasks and time period spent on each task throughout the day (14). This approach obviously requires more time and effort on the part of the worker and may not be feasible in large-scale epidemiologic field studies with repetitive measurements of the duration of computer use. The most promising approach, however, would be to use software or other objective measurements to register the duration of computer use at work. It should be noted that objective measurements may need considerable resources to ensure reliable registration throughout a longer time period.

In conclusion, the use of self-reports leads to the misclassification of exposure to computer use for more than 80% of all persons. This misclassification is predominantly nondifferential in nature and can only partly be explained by limited test–retest reliability.

Acknowledgments

We would like to thank Janneke Mikkers for her assistance with the data collection.

References

1. IJmker S, Huysmans M, Blatter BM, van der Beek AJ, van Mechelen W, Bongers PM. Should office workers spend fewer hours at their computer?: a systematic review of the literature. Occup Environ Med. 2007;64(4):211–22.
2. Homan MM, Armstrong TJ. Evaluation of three methodologies for assessing work activity during computer use. AIHA J. 2003;64:48–55.
3. Faucett J, Rempel D. Musculoskeletal symptoms related to video display terminal use: an analysis of objective and subjective exposure estimates. AAOHN J. 1996;44(1):33–9.
4. Heinrich J, Blatter BM, Bongers PM. A comparison of methods for the assessment of postural load and duration of computer use. Occup Environ Med. 2004;61:1027–31.
5. Lassen CF, Mikkelsen S, Kryger AI, Andersen JH. Risk factors for persistent elbow, forearm and hand pain among computer workers. Scand J Work Environ Health. 2005;31(2):122–31.
6. Douwes M, de Kraker H, Blatter BM. Validity of two methods to assess computer use: self-report by questionnaire and computer use software. Int J Ind Ergon. 2007;37(5):425–31.
7. Punnett L, Bergqvist U. Visual display unit work and upper extremity musculoskeletal disorders: a review of epidemiological findings. Solna (Sweden): Arbetslivsinstitutet; 1997.
8. Mikkelsen S, Vilstrup I, Lassen CF, Kryger AI, Thomsen JF, Andersen JH. Validity of questionnaire self-reports on computer, mouse and keyboard usage during a four-week period. Occup Environ Med. 2007;64(8):541–7.
9. van der Beek AJ, Frings-Dresen MH. Assessment of mechanical exposure in ergonomic epidemiology. Occup Environ Med. 1998;55(5):291–9.
10. IJmker S, Blatter BM, van der Beek AJ, van Mechelen W, Bongers PM. Prospective research on musculoskeletal disorders in office workers (PROMO): study protocol. BMC Musculoskelet Disord. 2006;7:55.
11. Blangsted AK, Hansen K, Jensen C. Validation of a commer-

cial software package for quantification of computer use. Int J Ind Ergon. 2004;34:237–41.
12. Karlqvist L. Assessment of physical work load at visual display unit workstations: ergonomic applications and gender aspects. Solna (Sweden): Arbetslivsinstitutet; 1997.
13. Evans O, Patterson K. Predictors of neck and shoulder pain in non-secretarial computer users. Int J Ind Ergon. 2000;26:357–65.
14. Unge J, Hansson GA, Ohlsson K, Nordander C, Axmon A, Winkel J, et al. Validity of self-assessed reports of occurrence and duration of occupational tasks. Ergonomics. 2005;48:12–24.

Received for publication: 20 April 2007