



Original article

Scand J Work Environ Health [1999;25\(3\):246-254](#)

doi:10.5271/sjweh.431

Evaluation of questionnaire-based information on previous physical work loads

by [Torgén M](#), [Winkel J](#), [Alfredsson L](#), [Kilbom Å](#), [Stockholm MUSIC 1 Study Group](#)

The following articles refer to this text: [2001;27\(1\):30-40](#); [2004;30\(4\):261-278](#); [2009;35\(4\):245-260](#)

Key terms: [differential misclassification](#); [exposure assessment](#); [occupational exposure](#); [questionnaire](#); [reproducibility](#); [retrospective study](#); [validity](#)

This article in PubMed: www.ncbi.nlm.nih.gov/pubmed/10450776



This work is licensed under a [Creative Commons Attribution 4.0 International License](http://creativecommons.org/licenses/by/4.0/).

Evaluation of questionnaire-based information on previous physical work loads

by Margareta Torgén, MD,^{1,2} Jörgen Winkel, PhD,¹ Lars Alfredsson, PhD,³ Åsa Kilbom, MD,¹ Stockholm MUSIC 1 Study Group⁴

Torgén M, Winkel J, Alfredsson L, Kilbom Å, Stockholm MUSIC 1 Study Group. Evaluation of questionnaire-based information on previous physical work loads. *Scand J Work Environ Health* 1999;25(3):246—254.

Objectives The principal aim of the present study was to evaluate questionnaire-based information on past physical work loads (6-year recall).

Methods Effects of memory difficulties on reproducibility were evaluated for 82 subjects by comparing previously reported results on current work loads (test-retest procedure) with the same items recalled 6 years later. Validity was assessed by comparing self-reports in 1995, regarding work loads in 1989, with worksite measurements performed in 1989.

Results Six-year reproducibility, calculated as weighted kappa coefficients (k_w), varied between 0.36 and 0.86, with the highest values for proportion of the workday spent sitting and for perceived general exertion and the lowest values for trunk and neck flexion. The six-year reproducibility results were similar to previously reported test-retest results for these items; this finding indicates that memory difficulties was a minor problem. The validity of the questionnaire responses, expressed as rank correlations (r_s) between the questionnaire responses and workplace measurements, varied between -0.16 and 0.78. The highest values were obtained for the items sitting and repetitive work, and the lowest and "unacceptable" values were for head rotation and neck flexion. Misclassification of exposure did not appear to be differential with regard to musculoskeletal symptom status, as judged by the calculated risk estimates.

Conclusions The validity of some of these self-administered questionnaire items appears sufficient for a crude assessment of physical work loads in the past in epidemiologic studies of the general population with predominantly low levels of exposure.

Key terms differential misclassification, exposure assessment, occupational, questionnaire, reproducibility, retrospective, validity.

In epidemiologic studies of work-related musculoskeletal disorders, direct measurements or systematic observations at the workplace are recommended in preference to questionnaire-based information (1—3). However, exposures that occurred in the past are often no longer available for observation. Thus, in most cases, questionnaires are the method of choice both because of their relatively low cost and because the subjects can be asked to recall past exposures. The accuracy of much self-reported

information has been doubted, mainly because of presumed recall problems. The magnitude of memory difficulties is likely to be closely related to the kind of information requested; for example there may be major problems with details such as percentages of workhours with parts of the body in specified positions. It is also reasonable to presume that memory difficulties increase with time (ie, more mistakes are made when answering questions regarding work loads 25 years back in time when

1 National Institute for Working Life, Solna, Sweden.

2 Department of Medical Sciences, Clinical Physiology, University Hospital, Uppsala, Sweden.

3 Institute of Environmental Medicine, Karolinska Institute, Stockholm, Sweden.

4 Karolinska Hospital and Institute, Departments of Epidemiology /IMM, Occupational Health, Orthopaedics, Rheumatology, and Rehabilitation and Physical Medicine, Stockholm, Sweden; National Institute of Occupational Health, divisions of Applied Work Physiology, Occupational Medicine, Psychophysiology, and Work and Environmental Physiology, Solna, Sweden; National Institute of Psychosocial Factors and Health, Stockholm, Sweden.

Reprint requests to: Dr Margareta Torgén, National Institute for Working Life, S-171 84 Solna, Sweden. [E-mail: margareta.torgen@niwl.se]

compared with questions on work loads a few years ago) and that the risk of differential misclassification of exposure due to symptom status may also increase with the length of the recall period.

The validity of questionnaire information on physical work loads is crucial for the accuracy of risk estimates for associations between work loads and musculoskeletal outcome. The validity of some self-reported gross activities in current work (eg, the fraction of the workday spent sitting) has been reported as sufficient for use in epidemiologic studies, while the validity of information on body postures has been found to be lower (4, 5). Validity studies performed so far have been restricted to present work load, while, to our knowledge, the validity of self-reported historical physical work loads has only been evaluated once before in the literature (6).

The aim of our study was to evaluate questionnaire information about past physical work loads (6 years ago), in terms of test-retest reliability and validity, by questionnaire responses and workplace measurements obtained 6 years back in time.

Subjects and methods

Study design

Questionnaire information on historical physical work loads was validated by comparing self-reports (collected in 1995), on physical work loads 6 years previously (in 1989), with worksite measurements obtained in 1989 and used as criterion values. Reports on perceived general exertion (RPE) was validated by heart-rate measurements.

Disagreement between measurements performed in 1989 and questionnaire information collected later (in 1995) probably has 2 main sources, recall problems and methodological differences between questionnaire items and workplace measurements. Recall difficulties were evaluated by comparing previously reported results on test-retest reproducibility (2 week test-retest) with the 6-year reproducibility results presented in this study. Disagreement between questionnaire definition and workplace measurements may arise as a result of differences in the response scales used, especially with regard to the duration of recorded as opposed to self-reported work loads and to differences in task definition between workplace measurements and questionnaire items. The applicability of different response scales and different phrasings of items was also investigated.

Differential bias related to the musculoskeletal symptom status was analyzed by comparing the reproducibility, sensitivity, and specificity between subjects with and without low-back symptoms. The musculoskeletal symptom status both in 1989 and in 1995 was considered.

Subjects

The study group consisted of 39 men and 58 women, a subgroup of the Stockholm MUSIC 1 study. This group was originally selected in 1989, and individual workplace measurements were then performed. "MUSIC" is an acronym for Musculoskeletal Intervention Center and is a network of 10 departments in Stockholm with the aim of preventing musculoskeletal disorders. The subjects constituted 4 subgroups [furniture removers (N=12), a sample from the general male population (N=27), medical secretaries (N=13), and a sample from the general female population (N=45)] in order to cover a broad spectrum of physical work loads (7). These subjects were all working in the Stockholm area during the Stockholm MUSIC 1 study, and most of them were still doing so in 1995, when they were contacted and asked to participate in a follow-up. Their present addresses and phone numbers were obtained from national population and phone number registers. The main reasons for nonparticipation were (i) could not be located at a recent phone number or address, (ii) did not respond to phone calls or letters, (iii) refused to participate (2 subjects).

The study was reviewed and approved by the regional Ethics Committee of Human Research at the Karolinska Institute, Stockholm, Sweden.

Data collection instruments

Measurements at the workplaces in 1989. Physical work loads were recorded individually at the workplaces during a normal workday for each of the 97 subjects. Heart rate was recorded during a whole workday by a Sport Tester PE 3000 (Polar Electro, Finland), and the percentage of time spent sitting was determined with a Posimeter device (8). Body postures, repetitive finger and hand movements, and manual handling were recorded by experienced ergonomists and physiotherapists with the portable ergonomic observation (PEO) method (9). The heart rates and the PEO results of the observation day were weighted for a "typical workweek" according to interview information on task duration during the week (7, 9).

Questions in 1989. Illustrated questions about work postures and manual handling were used (appendix). Their reproducibility for present work loads have previously been evaluated with a test-retest procedure in a population study of 343 subjects (10) and validated in relation to worksite measurements comprising the same subjects as in the present study (5). The physical activities described by these items were quantified as a proportion of the workday (6-degree ordinal scale) or as the frequency per hour (5-degree ordinal scale).

Questions in 1995. In 1995 new items (appendix) were added to the previously described items in order to

explore the influence on the validity of different response scales. These items have been evaluated for test-retest reliability (11) and were quantified as the proportion of the workday, using a 100-millimeter visual analogue scale (VAS) or using the number of days per month spent at a certain activity as recorded on a 5-degree ordinal scale.

Physical exertion at work was asked for both in 1989 and 1995, regarding the general perceived exertion in 1989, with a modified RPE scale (0–14). Information on past musculoskeletal symptoms (last 12 months) requested in 1989 and in 1995 was obtained by the Nordic questionnaire (12) and additional questions were asked in 1995 regarding present employment or retirement status.

Procedure

Subjects living in the Stockholm area in 1995 were contacted by telephone and invited to meet with a member of the research team. Those living far away or too busy for a personal meeting were sent instructions and questionnaires and contacted by telephone (3 subjects). When the subjects answered the self-administered questionnaire items in 1995, they were instructed to consider their work situation as it was during the time of the Stockholm MUSIC study in 1989.

A total of 7 questionnaire items that could be evaluated in relation to workplace measurements were identified and checked for comparability with the workplace measurements performed in 1989. Thus repetitive finger and hand movements were recorded as “repetitive work” during the workplace measurements in 1989, but they were given as 1 or 2 separate questions in the questionnaires. Only the item on repetitive finger movements was chosen for the validation of “repetitive work” in the past. Manual handling of loads was measured in 1989 as the total number of lifts, independent of weight, and the answers concerning these items in the questionnaires could therefore not be validated.

Two more items were not evaluated. One was because of the incongruent definition between the questionnaire item and the workplace measurement (work with hands above shoulder level). The other item (trunk rotation more than 45 degrees) was deleted from the analysis since too few exposed subjects (7 subjects) exceeded 0% of their worktime in this activity during an estimated normal week.

Data analysis

The distribution of questionnaire responses was summarized using mean values, ranges, and 10th, 50th and 90th percentiles (P_{10} , P_{50} , P_{90}).

The 6-year reproducibility of the questionnaire items was analyzed with weighted Cohen's kappa coefficients (k_w) with 95% confidence intervals (95% CI). Kappa co-

efficients describe agreement beyond chance and produce results identical with those of intraclass correlation coefficients, if calculated with quadratic weights on categorical responses. Kappa values exceeding 0.75 were regarded as “excellent agreement” beyond chance, values below 0.40 represented “poor agreement”, and values between 0.40 and 0.75 were rated “fair to good agreement” (13). Tests of the statistical significance of differences in the kappa coefficients between independent groups (men, women; with symptoms, without symptoms) were made according to Fleiss (13).

Validity was analyzed for the retrospective questionnaire responses obtained in 1995 with worksite measurements performed in 1989 as the reference values. The response scales differed between the measurements at the workplaces in 1989 and the questionnaire items in 1995 so that an exact matching of scales was not possible, and the Spearman rank correlation coefficient (r_s) was chosen for the calculation of agreement. A correlation coefficient of at least 0.6 was regarded as an indicator of high agreement. Tests of statistical significance of differences in rank correlation coefficients between independent groups (men, women; with symptoms; without symptoms) were made according to Blalock (14).

For calculations of sensitivity and specificity the cut-off point between the exposed and unexposed subjects was defined as the median value because of the incomparability between the scales used, and also in order to have enough subjects in both exposure groups. However, some analyses of sensitivity and specificity, using cut-off points based on current knowledge about adverse effects on the musculoskeletal system, were also performed. The distributions of questionnaire responses and measurement values were highly skewed in some cases, with most responses at the lower end of the scales. The sensitivity and specificity were only calculated if there were 10 or more subjects in each exposure group.

Results

Eighty-two of the 97 subjects (85%) participated in our current study (table 1). The reported percentages of neck and shoulder or low-back symptoms during the last 12 months were approximately equal in 1989 and 1995. About 1/3 of the subjects reported neck and shoulder or low-back symptoms in both of these years, 1/3 only in 1989 and 1/3 only in 1995.

Six-year reproducibility

Slightly higher work loads were reported in 1995 than in the original responses in 1989 for some of the items, such as perceived general exertion, hands above shoulder level, and trunk bent forward (table 2).

The 6-year reproducibility (ρ_w) of different questionnaire items varied between 0.36 and 0.86, with the highest values for the proportion of the workday spent sitting (0.86) and perceived general exertion (0.70) and the lowest values for forward bending of the trunk (0.36) and head bent forward (0.37). There was no general trend for gender differences in the kappa values. The reproducibility (ρ_w) for the men significantly exceeded that for the women for the items on perceived exertion ($P=0.004$) and the lifting of loads between 1 and 5 kilograms ($P=0.002$). Higher reproducibility among the women than the men was found for the item on repetitive finger movements ($P<0.001$). The presence of low-back symptoms during the last 12 months before answering the questionnaire in 1995 did not seem to affect the test-retest reliability in any obvious direction (table 2), and no statistically significant differences were found between those with low-back symptoms and those without. The influence on reproducibility for neck and shoulder symptoms was also examined, with similar results. The influence of symptom status in 1995 on reproducibility did not differ when compared with the influence of symptom status in 1989.

Validity of retrospective questionnaire information

The correlation between workplace measurements and questionnaire reports was high for sitting and repetitive work, moderate for perceived exertion, trunk flexion, and kneeling and squatting, but poor for neck flexion and neck rotation (table 3).

The correlation (r_s) among the men exceeded that among the women for the item on perceived exertion ($P=0.02$). Higher correlation among the women than among the men was found for the item on repetitive work, especially with the 6-degree ordinal scale ($P=0.07$), but to a less extent with the 5-degree ordinal scale ($P=0.27$). The items on sitting, kneeling and squatting, and trunk bent forward showed only minor gender differences. The 2 questions on head postures yielded low a correlation for the women and could not be evaluated for the men because of the few measurements performed (table 3).

Among the subjects with low-back symptoms the coefficient of correlation (r_s) was higher than among those without symptoms for the item on perceived physical exertion ($P=0.03$). Higher (not statistically significant) agreement among those without low-back symptoms was noted for the item on forward bending. The item on kneeling and squatting showed significantly higher agreement among the subjects with low-back symptoms ($P=0.01$). No trend could be seen for the 1995 responses regarding sensitivity and specificity between those with or without low-back symptoms.

The distribution of the responses was highly skewed for some items, and sensitivity and specificity were therefore not calculated for all the subgroups (table 3). When cut-off points based on current knowledge about adverse

effects on the musculoskeletal system were used, slightly lower sensitivity and specificity values were obtained in comparison with the corresponding values for cut-off points based on median values.

Discussion

In this study of people interviewed on 2 occasions, the agreement between self-reports of current conditions and self-reports of the same conditions after a delay of 6 years was fair to excellent ($k_w > 0.40$) for 9 of 13 items, namely, perceived exertion, sitting, kneeling and squatting, head rotation, hands above shoulder level, repetitive finger movements, and the 3 items on the manual handling of loads. Low reproducibility ($k_w < 0.40$) was found for 4 items on head and trunk posture and repetitive hand movements (table 2).

Differences between the previously reported 2-week test-retest reproducibilities of the questionnaire items used in the Stockholm MUSIC study (10) and the 6-year reproducibility results in this study were regarded as a measure of memory difficulties during the time period between 1989 and 1995. For 7 of the 13 items in table 3

Table 1. Participants (N=82) in the reexamination in 1995 and dropouts (N=15). Employment status and changes in occupations between 1989 and 1995 are given. The subjects have been divided into 4 groups, established in 1989 (furniture removers N=12, male population N=27, medical secretaries N=13, female population N=45).

	Male groups ^a (N)		Female groups ^b (N)		All ^c (N)
	Furniture removers	Male population	Medical secretaries	Female population	
Subjects participating in 1995	8	23	11	40	82
Dropouts	4	4	2	5	15
Participants in 1995 (N=82)					
Gainfully employed	7	21	10	31	69
Subjects who changed occupation between 1989 and 1995	2	4	1	6	13
Unemployed	0	0	0	2	2
Old age pension	0	1	1	5	7
On sick leave	1	1	1	2	5
Neck or shoulder symptoms ^d	6 (3)	8 (8)	9 (8)	20 (23)	43 (42)
Low-back symptoms ^d	4 (6)	9 (7)	6 (7)	13 (15)	32 (35)

^a Mean (range) age: furniture removers 46 (40–57) years; male population 47 (29–67) years.

^b Mean (range) age: medical secretaries 54 (27–69) years; female population 51 (26–70) years.

^c Mean (range) age: 50 (26–70) years.

^d Any feeling of pain or discomfort during the last 12 months in 1995. The numbers of subjects with symptoms during the last 12 months in 1989 are in parentheses.

Table 2. Reproducibility comparing questionnaire responses in 1989 with those obtained 6 years later, analyzed by weighted kappa coefficients (k_w) with the 95% confidence interval (95% CI). The distribution of work loads has been given as the 10th, 50th, and 90th percentiles (P_{10} , P_{50} , P_{90}) for 82 subjects. Results are also given for subgroups (ie, gender and low-back symptoms during the last 12 months before the retest in 1995).

Question*	1989			1995 about 1989			k_w	95% CI
	P_{10}	P_{50}	P_{90}	P_{10}	P_{50}	P_{90}		
Perceived exertion (scale 0—14)								
All	2	5	9	3	6	9	0.70	0.60—0.81
Men	2	6	10	3	7	11	0.81	.
Women	2	5	7	3	6	8	0.51	.
Low-back symptoms	1	7	10	3	7	11	0.73	.
No low-back symptoms	2	5	7	3	5	8	0.62	.
Sitting (scale 1—6)								
All	2	4	6	2	4	6	0.86	0.81—0.90
Men	2	4	6	2	3	5	0.85	.
Women	2	4	6	2	4	6	0.86	.
Low-back symptoms	2	4	6	2	4	6	0.87	.
No low-back symptoms	1	4	6	2	4	5	0.85	.
Kneeling and squatting (scale 1—6)								
All	1	1	2	1	1	3	0.58	0.44—0.73
Men	1	1	3	1	1	2	0.63	.
Women	1	1	2	1	1	3	0.57	.
Low-back symptoms	1	1	3	1	2	3	0.67	.
No low-back symptoms	1	1	2	1	1	2	0.45	.
Head bent forward (scale 1—6)								
All	1	3	6	1	3	6	0.37	0.15—0.58
Men	1	2	5	1	2	4	0.16	.
Women	1	3	6	1	3	6	0.36	.
Low-back symptoms	1	3	6	1	3	5	0.34	.
No low-back symptoms	1	3	6	1	2	6	0.38	.
Head rotation (scale 1—6)								
All	1	1	3	1	1	2	0.55	0.35—0.76
Men	1	1	3	1	2	2	0.37	.
Women	1	1	2	1	1	2	0.66	.
Low-back symptoms	1	1	3	1	1	2	0.46	.
No low-back symptoms	1	1	2	1	1	2	0.62	.
Hands above shoulders (scale 1—6)								
All	1	1	3	1	2	3	0.45	0.23—0.67
Men	1	1	3	1	2	3	0.34	.
Women	1	1	3	1	2	3	0.51	.
Low-back symptoms	1	1	2	1	2	3	0.55	.
No low-back symptoms	1	1	3	1	2	3	0.42	.
Trunk bent forward (scale 1—6)								
All	1	1	3	1	2	3	0.36	0.16—0.56
Men	1	1	3	1	2	2	0.38	.
Women	1	1	2	1	2	3	0.35	.
Low-back symptoms	1	1	3	1	2	4	0.20	.
No low-back symptoms	1	1	2	1	2	2	0.50	.
Trunk rotation (scale 1—6)								
All	1	1	2	1	1	3	0.39	0.05—0.71
Men	1	1	2	1	1	2	0.20	.
Women	1	1	2	1	1	3	0.47	.
Low-back symptoms	1	1	3	1	2	3	0.48	.
No low-back symptoms	1	1	2	1	1	2	0.16	.
Repetitive hand movements (scale 1—6)								
All	1	1	6	1	2	6	0.38	0.17—0.59
Men	1	2	4	1	2	4	0.11	.
Women	1	1	6	1	3	6	0.46	.
Low-back symptoms	1	2	6	1	2	6	0.19	.
No low-back symptoms	1	1	5	1	3	6	0.49	.
Repetitive finger movements (scale 1—6)								
All	1	2	6	1	2	6	0.63	0.46—0.81
Men	1	1	4	1	2	4	0.07	.
Women	1	3	6	1	3	6	0.78	.
Low-back symptoms	1	3	6	1	2	5	0.51	.
No low-back symptoms	1	2	6	1	3	6	0.72	.

(continued)

Table 2. Continued.

Question ^a	1989			1995 about 1989			k _w	95% CI
	P ₁₀	P ₅₀	P ₉₀	P ₁₀	P ₅₀	P ₉₀		
Manual handling 1—5 kg (scale 1—5)								
All	1	2	3	1	2	3	0.40	0.20—0.61
Men	1	2	3	2	2	3	0.68	.
Women	1	2	3	1	2	3	0.19	.
Low-back symptoms	1	2	3	1	3	3	0.54	.
No low-back symptoms	1	2	3	1	2	3	0.26	.
Manual handling 6—15 kg (scale 1—5)								
All	1	1	3	1	2	3	0.64	0.51—0.77
Men	1	2	3	1	2	4	0.64	.
Women	1	1	2	1	1	3	0.55	.
Low-back symptoms	1	1	3	1	2	3	0.70	.
No low-back symptoms	1	1	3	1	1	3	0.59	.
Manual handling 16—45 kg (scale 1—5)								
All	1	1	3	1	1	3	0.58	0.38—0.78
Men	1	2	3	1	2	3	0.45	.
Women	1	1	1	1	1	1	0.15	.
Low-back symptoms	1	1	3	1	1	3	0.67	.
No low-back symptoms	1	1	2	1	1	2	0.39	.

^a A short version of the questions and scales used is presented in the appendix. There were 38% of the men in the whole study group, 40% of the men among those with low-back symptoms and 36% among those without low-back symptoms.

(sitting, kneeling and squatting, head rotation, hands above shoulder level, repetitive finger movements, manual handling of loads between 6 and 15 kilograms, and manual handling of loads between 16 and 45 kilograms), a decrease in the correlation coefficient of less than 0.10 between the 2-week and 6-year reproducibility was found, indicating only minor memory difficulties (average 2-week reproducibility 0.64, average 6-year reproducibility 0.62). The corresponding average reproducibility values for the 6 remaining items were 0.63 for 2 weeks and 0.43 for 6 years. The results are in accordance with the previously reported results, with comparisons of different time intervals between test and re-test (11), and also with the results of a study of physical activities during leisure time (15). The majority of the subjects in our study (about 75%) were still working in the same occupations in 1995 as in 1989. This circumstance may lead to an underestimation of memory difficulties in questions about physical work loads 6 years back, if the subjects used current work as a proxy for their work situation in 1989. Thirteen subjects in the present study changed jobs between 1989 and 1995, 8 of them to jobs with a similar level of physical load, and 5 to higher or lower physical load, based on an evaluation of the job titles. Physical work loads often decrease with age, especially among men (16), and reports of work loads in the past would thereby probably be underreported in cases of influence from the current work situation. However, this is less likely to be a big problem in this study, which showed mainly unchanged or even slightly higher responses in 1995 for work loads 6 years earlier, compared with the responses in 1989 (table 2).

The validity of the self-reported exposures was good for the items on sitting and repetitive work, and intermediate for the items on perceived exertion and forward bending of the trunk. The results of the items on head postures showed poor results, a finding in line with previously reported results for these items (5). The results of the item on kneeling and squatting showed less agreement than in the study by Wiktorin et al (5).

A major concern in epidemiologic studies of physical work loads and musculoskeletal health is whether the musculoskeletal symptom status may cause differential misclassification of exposure. If so, the point estimates of relative risk could be either under- or overestimated, depending on the direction of the bias. In view of possible gender differences, agreement between questionnaire responses and workplace measurements in relation to musculoskeletal symptom status should preferably be evaluated in groups with a similar gender distribution. This was the case regarding low-back symptoms, for which the men constituted approximately 35% of both the group with and that without symptoms (except for the items on head posture). A comparison of the subjects with and without symptoms did not indicate any tendency for the subjects with symptoms to report higher or lower exposure levels than the symptom-free subjects (table 2). However, the precision of the questionnaire responses (random error) could still cause differential misclassification if the subjects with musculoskeletal symptoms remember their previous work loads more exactly than those without symptoms. The rank correlation coefficients for the subjects with low-back symptoms during the last 12 months differed from the corresponding

Table 3. Agreement calculated as the Spearman rank correlation (r_s) between questionnaire responses in 1995 concerning physical load at work in 1989 and reference measurements at the workplaces in 1989. Sensitivity and specificity were calculated using the workplace measurements as criterion values, and the variables were dichotomized into 2 classes (exposed and unexposed) by the median value. The study group was divided into different groups according to gender and low-back symptom status during the last 12 months before answering the questionnaire in 1995.

Question ^a	N	Percentage of men	r_s	Sensitivity	Specificity
Perceived general exertion versus heart rate					
All	78	36	0.52	0.65	0.74
Men	28	0	0.71	0.71	0.86
Women	50	0	0.34	0.52	0.68
Low-back symptoms	30	37	0.65	0.89	0.75
No low-back symptoms	48	35	0.29	0.45	0.73
Sitting part of the workday (scale 0—100 %)					
All	75	37	0.73	0.79	0.81
Men	28	0	0.71	0.79	0.79
Women	47	0	0.74	0.68	0.76
Low-back symptoms	30	40	0.70	0.67	0.73
No low-back symptoms	45	36	0.76	0.87	0.86
Sitting part of the workday (scale 1—6)					
All	75	37	0.78	0.58	0.89
Men	28	0	0.78	0.79	0.86
Women	47	0	0.80	0.55	0.88
Low-back symptoms	30	40	0.75	0.47	0.80
No low-back symptoms	45	36	0.83	0.65	0.95
Kneeling and squatting (scale 1—6)					
All	64	33	0.46	0.64	0.72
Men	21	0	0.61	.	.
Women	43	0	0.40	0.63	0.67
Low-back symptoms	25	32	0.74	0.82	0.79
No low-back symptoms	39	33	0.23	0.50	0.68
Head bent forward (scale 1—6)					
All	40	23	0.07	0.55	0.45
Men	9	0	-	.	.
Women	31	0	0.12	0.53	0.50
Low-back symptoms	14	14	-0.17	.	.
No low-back symptoms	26	27	0.23	0.56	0.50
Head rotation (scale 1—6)					
All	41	22	-0.16	0.25	0.62
Men	9	0	-	.	.
Women	32	0	-0.13	.	.
Low-back symptoms	14	14	0.19	.	.
No low-back symptoms	27	26	-0.32	.	.
Trunk bent forward (scale 1—5)					
All	69	35	0.54	0.65	0.76
Men	24	0	0.56	0.67	0.92
Women	45	0	0.52	0.58	0.81
Low-back symptoms	27	33	0.48	0.70	0.71
No low-back symptoms	42	36	0.60	0.61	0.81
Trunk bent forward (scale 1—6)					
All	69	35	0.51	0.81	0.63
Men	24	0	0.55	.	.
Women	45	0	0.49	0.74	0.65
Low-back symptoms	27	33	0.39	.	.
No low-back symptoms	42	36	0.63	0.76	0.76
Repetitive work (scale 1—5)					
All	63	32	0.62	0.81	0.69
Men	20	0	0.49	0.67	0.64
Women	43	0	0.61	0.85	0.65
Low-back symptoms	26	31	0.69	0.81	0.82
No low-back symptoms	37	32	0.58	0.80	0.59

(continued)

Table 3. Continued.

Question ^a	N	Percentage of men	r_s	Sensitivity	Specificity
Repetitive work (scale 1—6)					
All	63	32	0.65	0.84	0.75
Men	20	0	0.36	.	.
Women	43	0	0.67	0.75	0.74
Low-back symptoms	26	31	0.74	0.82	0.87
No low-back symptoms	37	32	0.59	0.85	0.65

^a A short version of the questions and scales used is presented in the appendix.

correlation for those without symptoms for only a couple of items. However, these differences in agreement between symptom groups were not associated with differences between calculated prevalence ratios based on questionnaire responses (PR_{quest}) and those based on reference measurements (PR_{ref}). Thus there appears to have been no serious differential misclassification of exposure in our study. This conclusion is in accordance with a previously reported evaluation of present physical work loads obtained with the MUSIC questionnaire (5), but not with the results of a recently published study of forest industry workers in Finland (4).

The additional items used in 1995 were constructed on the basis of experiences from the Stockholm MUSIC study, but with a focus on physical work loads in the past and in the general population, and some changes in the wordings of items and the response scales were therefore considered necessary. These changes limited the items that could be validated in relation to workplace measurements performed in 1989. Both the old and the new (additional) items could be regarded as suitable for collecting information on historical physical work loads. However, the new items could be recommended in studies of the general population, as the ordinal response scale was found to generate better (ie, wider) response distributions in cases of predominantly low-exposure levels, which are often found in population studies. This finding is in accordance with the intentions when the new items were created.

Comparisons between previously reported validity of the self-reported physical work loads at present (5) and in the past (the present study) could be made for 6 items with comparable cut-off points in the 2 studies. Approximately unchanged and acceptable correlation values were found for daily time spent sitting and for forward bending of the trunk. Unchanged and "low agreement" were noted for the item on head rotation, and decreased agreement was found for kneeling and squatting and for forward bending of the head.

Despite the promising results for some of the items evaluated in this study, no general recommendation for the use of these items in epidemiologic studies can be

made. However, the validity of some of these self-administered questionnaire items (perceived exertion, sitting, kneeling and squatting, trunk flexion and repetitive work) are believed to be sufficient for the assessment of physical work loads in the past on a rough level, in studies of general population groups.

Acknowledgments

This study was supported by a grant from the Swedish Council for Working Life Research. We are grateful to Jan Fröberg, BSc, for the collection of the questionnaire information in 1995 and to those working in the Stockholm MUSIC study, especially to Lena Karlquist, PhD, and Christina Wiktorin, PhD, who contributed to the initial data collection in 1989. We also want to thank Professor Laura Punnett for her valuable comments on the manuscript.

References

1. Kilbom Å. Assessment of physical exposure in relation to work-related musculoskeletal disorders — what information can be obtained from systematic observations? *Scand J Work Environ Health* 1994;20 special issue:30—45.
2. Hagberg M. Exposure variables in ergonomic epidemiology. *Am J Ind Med* 1992;21:91—100.
3. Winkel J, Mathiassen SE. Assessment of physical work load in epidemiologic studies: concepts, issues and operational considerations. *Ergonomics* 1994;37:979—88.
4. Viikari-Juntura E, Rauas S, Martikainen R, Kuosma E, Riihimäki H, Takala E-P, et al. Validity of self-reported physical work load in epidemiologic studies on musculoskeletal disorders. *Scand J Work Environ Health* 1996;22:251—9.
5. Wiktorin C, Karlqvist L, Winkel J, Stockholm MUSIC I study group. Validity of self-reported exposures to work postures and manual materials handling. *Scand J Work Environ Health* 1993;19:208—14.
6. Evanoff B. Reliability and validity testing of a questionnaire measuring past physical exposures to the back and lower extremities [thesis]. Seattle (WA): University of Washington, 1993.
7. Karlqvist L, Winkel J, Wiktorin C. Direct measurements and systematic observations of physical workload among medical secretaries, furniture removers and male and female reference populations. *Appl Ergon* 1994;25:319—26.
8. Selin K, Winkel J, Stockholm MUSIC I study group. Evaluation of two instruments for recording sitting and standing postures and number of foot steps. *Appl Ergon* 1994;25:41—6.
9. Fransson-Hall C, Gloria R, Kilbom Å, Winkel J, Karlqvist L, Wiktorin C, et al. A portable ergonomic observation method (PEO) for computerized on-line recording of postures and manual handling. *Appl Ergon* 1995;26:93—100.
10. Wiktorin C, Wigaeus Hjelm E, Winkel J, Köster M. Reproducibility of a questionnaire for assessment of physical load during work and leisure time. *J Occup Environ Med* 1996;38:190—201.
11. Torgén M, Alfredsson L, Köster M, Wiktorin C, Smith K, Kilbom Å. Reproducibility of a questionnaire for assessment of present and past physical activities. *Int Arch Occup Environ Health* 1997;70:107—18.
12. Kuorinka I, Jonsson B, Kilbom Å, Vinterberg H, Biering-Sørensen F, Andersson G, et al. Standardised Nordic questionnaires for the analysis of musculoskeletal symptoms. *Appl Ergon* 1987;18:233—7.
13. Fleiss J. The measurement of interrater agreement. In: Fleiss J. *Statistical methods for rates and proportions*: chapter 13. 2nd ed. New York (NY): John Wiley & Sons, 1981:212—36.
14. Blalock HM. Correlation and regression (continued). In: Blalock HM. *Social statistics, revised second edition*. Singapore: McGraw-Hill Book Company, 1985:415—49.
15. Blair S, Dowda M, Pate R, Kronenfeld J, Howe H, Parker G, et al. Reliability of long-term recall of participation in physical activity by middle-aged men and women. *Am J Epidemiol* 1991;133:266—75.
16. Torgén M, Kilbom Å. The REBUS study: physical work load between 18 and 58 years of age — does it change? In: Seppälä P, Luopajarvi T, Nygård CH, Mattila M, editors. *Proceedings of the 13th triennial congress of the International Ergonomics Association, Tampere, Finland; vol 5*. Helsinki: Finnish Institute of Occupational Health, 1997:544—6.

Received for publication: 2 July 1998

Appendix

Questions used in 1989 and 1995 on work posture and manual materials handling

Items used both in 1989 and 1995

- Perceived general exertion — modified scale for rating of perceived exertion, range (0—14).
- Sitting, kneeling or squatting, head bent forward >20 degrees, head rotation >45 degrees, hands above shoulder level, trunk bent forward >60 degrees, trunk rotated >45 degrees, the same hand movements several times per minute, the same finger movements several times per minute — ordinal scale no 1, range (1—6).
- Lifting or carrying burdens weighing between 1 and 5 kg, 6 and 15 kg, 16 and 45 kg — ordinal scale no 2, range (1—5).

Items used only in 1995

- Proportion of the day spent sitting — continuous scale, range (0—100 %).
- Hands below knee level exceeding 30 minutes a day, repetitive hand or finger movements several times a minute exceeding 2 hours a day — ordinal scale no 3, range (1—5).

Continuous scale

Not at all _____ All the time

Ordinal scale no 1

Not at all	A little, about 1/10 of the time	About 1/4 of the time	Half of the time	About 3/4 of the time	Almost all the time
<input type="checkbox"/> (1)	<input type="checkbox"/> (2)	<input type="checkbox"/> (3)	<input type="checkbox"/> (4)	<input type="checkbox"/> (5)	<input type="checkbox"/> (6)

Ordinal scale no 2

Not at all	Less than once per hour	1—10 times per hour	11—30 times per hour	More than 30 times per hour
<input type="checkbox"/> (1)	<input type="checkbox"/> (2)	<input type="checkbox"/> (3)	<input type="checkbox"/> (4)	<input type="checkbox"/> (5)

Ordinal scale no 3

Almost never/ not at all	1—3 days per month	One day per week	2—4 days per week	Every work day
<input type="checkbox"/> (1)	<input type="checkbox"/> (2)	<input type="checkbox"/> (3)	<input type="checkbox"/> (4)	<input type="checkbox"/> (5)