Scand J Work Environ Health 2002;28(4):232-237

<https://doi.org/10.5271/sjweh.670>

Issue date: Aug 2002

Reliability assessment of a coding scheme for the physical risk factors of work-related musculoskeletal disorders

by [Chen Y](#), [Meyer JD](#), [Song J](#), [McDonald JC](#), [Cherry N](#)

Affiliation: Centre for Occupational & Environmental Health, The Faculty of Medicine, The University of Manchester, Block D, 4th Floor, Humanities Building, Oxford Road, Manchester M13 9PL, UK. Yiqun.chen@man.ac.uk

Refers to the following texts of the Journal: [1999;25 suppl 4:25-30](#)
[1999;25\(2\):81-83](#) [1998;24\(5\):376-385](#)

Key terms: [epidemiology](#); [exposure assessment](#); [kappa](#); [occupational musculoskeletal disorder](#); [physical risk factor](#); [reliability assessment](#); [surveillance](#); [work-related musculoskeletal disorder](#)

This article in PubMed: www.ncbi.nlm.nih.gov/pubmed/12199424

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

Reliability assessment of a coding scheme for the physical risk factors of work-related musculoskeletal disorders

by Yiqun Chen, MSc,¹ John D Meyer, MD,¹ Jaechul Song, MD,² J Corbett McDonald, MD,³ Nicola Cherry, MD^{1, 4}

Chen Y, Meyer JD, Song J, McDonald JC, Cherry N. Reliability assessment of a coding scheme for the physical risk factors of work-related musculoskeletal disorders. *Scand J Work Environ Health* 2002;28(4):232–237.

Objectives This study assessed the reliability of a novel coding scheme for physical risk factors for musculoskeletal disorders reported to an occupational surveillance scheme.

Methods Since 1997 new cases of musculoskeletal disease have been reported as part of a surveillance scheme by over 300 consultant rheumatologists in the United Kingdom; the rheumatologists also gave a short description of the tasks and activities they considered to be causal. With the use of a summary of the activities described, a coding scheme was developed comprising 16 categories of task codes and another 16 categories of movement codes. Four reviewers coded the work activities independently for 576 cases. The fourth rater coded the cases twice. With the use of a single summary kappa statistic and the matrix of kappa coefficients, both interrater reliability and intrarater reliability were assessed.

Results The overall interrater agreement on the task codes was good (kappa = 0.73), with the best agreement for keyboard work (kappa = 0.96) and the worst for assembly work (kappa = 0.40, kappa = 0.37). The interrater agreement on movement codes was also good (kappa = 0.79), with the best agreement for kneeling (kappa = 0.94) and the worst for materials handling (kappa = 0.10). The intrarater agreement was somewhat better than the interrater agreement with both codes.

Conclusions The results suggest that the coding scheme was, on the whole, reliable for classifying the physical risk factors reported as causal.

Key terms epidemiology, exposure assessment, kappa, occupational, risk factors, surveillance.

Musculoskeletal disorders are the most frequently reported type of work-related illness in the United Kingdom (1–3). Many epidemiologic studies have explored the relationship between work-related physical risk factors and musculoskeletal disorders. For such research, reliable and valid methods of measurement are necessary. The complex task of exposure assessment in musculoskeletal epidemiology entails the use of methods broadly classifiable as subjective judgment, systematic observation, and direct measurement (4). Subjective judgments may have limited precision and accuracy, but the cost of data collection through interviews and ques-

tionnaires is low, application is usually feasible, and it may provide the only opportunity to describe physical factors in the population under study (5). Information can also be collected from clinicians who treat patients with a musculoskeletal disease. Although physicians may probe into exposure details, they still depend on the patient's own description, unless they visit the workplace. Nevertheless, such self-reported information on physical demands at work has been found to be reasonably valid (6).

Since October 1977 some 80% of consultant rheumatologists in the United Kingdom have participated in

- 1 Centre for Occupational & Environmental Health, University of Manchester, Manchester, United Kingdom.
- 2 Hanyang University College of Medicine, Seoul, Korea.
- 3 Imperial College of Science, Technology and Medicine, London, United Kingdom.
- 4 Department of Public Health Sciences, University of Alberta, Alberta, Canada.

Reprint requests to: Yiqun Chen, Centre for Occupational & Environmental Health, The Faculty of Medicine, The University of Manchester, Block C, 4th Floor, Humanities Building, Oxford Road, Manchester M13 9PL, UK. [E-mail: Yiqun.chen@man.ac.uk]

a voluntary reporting scheme for new cases of musculoskeletal disease caused or made substantially worse by work. This scheme, known as MOSS (the musculoskeletal occupational surveillance scheme) forms part of ODIN (the occupational disease intelligence network) which, since 1998, has been coordinated from Manchester and now covers all types of work-related disease (7). The methods used in MOSS and an analysis of the main findings, 1997–2000, have recently been published (8). Each new case reported includes a brief description of the tasks and activities considered by the physician to have been causal. This paper presents the results from a study of the reliability of a coding procedure for these factors; the study is based on reports received from some 300 rheumatologists during the two-year period, October 1997 to December 1999.

Material and methods

Between October 1997 and December 1999, a total of 667 new cases of musculoskeletal disorder were reported to MOSS by consultant rheumatologists, each randomly assigned to only one month a year. Reporters were asked to include a new case of a disease or illness caused by work, with additional elaboration, in the guidelines, that such a condition could be considered occupational if it would not have occurred in the absence of the occupational exposure or if the occupational exposure “made a substantial difference to severity”. From those reported, we selected all 576 cases with a specific diagnosis, attributed to repeated exposure. (See the following text.)

The card used for the reporting required the participating rheumatologists to specify disease categories within a particular body region: hand-wrist-arm, elbow, shoulder, neck-thoracic spine, lumbar spine-trunk, hip-knee, ankle-foot, or other. Within each body region, more specific categories, such as nerve entrapment in the arm, epicondylitis in the elbow, and disc problem in the lumbar region, were listed. Physicians could report more than one disease category for each patient; about 6% of the cases reported appeared in more than one category. In the present study, each disorder was considered independently. Occupation and industry were also recorded, and the physician was asked to indicate whether the disorder was thought to have resulted from single or repeated exposure; only the latter was used in this study. The rheumatologists did not receive training on how to complete the card but were sent guidelines on the admissibility of cases and on how to complete each section of the card.

Guidelines for completing the section entitled “activity, task or event” were as follows: “Please indicate

the activity, task or event that you feel precipitated the illness (eg, lifting boxes, filleting fish, etc)”. Initial attempts to use theoretically derived codes reflecting, for example, physiological or ergonomic demand, were unproductive, as the information recorded was too limited. This approach was abandoned and replaced by a more pragmatic scheme, attempting to summarize the data for all reported musculoskeletal conditions in all activities, tasks or events, in terms of (i) task, (ii) movement, (iii) repetition, and (iv) vibration. For the latter two dimensions attempts were only made to code presence or absence (as probable, possible, unlikely, or none). The task and movement codes are listed in table 1. Both were necessary, as job tasks may not specify the physical activity, whereas movement alone may not describe the type of work being performed.

Three coders working independently coded the cases in the following order: hand-wrist-arm first, then elbow, and finally ankle-foot. Each coder was given a copy of the coding scheme with examples as shown in table 1, together with a printout including the reporting physician’s description of the work activity, the diagnostic group to which the disorder had been ascribed, and the job and industry recorded. No further training or instruction was given, and discussion was not allowed between the coders during the procedure. However, after completion, coder 1 and coder 3, chosen simply because of time availability rather than specialist knowledge, reviewed anything they had coded differently and, after discussion, assigned a reconciled code. These reconciled codes were considered the “gold standard” for calculating false negative and positive rates. The fourth rater coded the cases twice. On the second occasion, 4 weeks after the first, the same cases were presented in random order, and the results were compared for intrarater reliability. One of these two sets, selected randomly, was then included with output from the first three coders to assess interrater reliability.

As the physicians’ description of work activity could not always be incorporated into a single code, the coders were allowed to use up to three task and three movement codes for each. Some 10% of the cases had two task codes, but only very few (0.07%) required three; 19% required two movement codes, whereas 3% required three. The raters were asked to record first the code they considered most important and, for simplicity, the first code only was used in this analysis. Information on activity, task, or event contributed to the coding of both task and movement; as such the codes ascribed are not strictly independent, but are considered separately.

Interrater agreement and intrarater agreement were assessed by means of the kappa statistic, together with 95% confidence intervals (9). Matrices of kappa coefficients were prepared using a program designed by

Table 1. Task and movement codes (with some examples, in parentheses).

Code	Tasks	Movements
1	Keyboard work (typing, work with visual display unit or computer, checkout, piano playing)	Fine hand (keyboarding, sewing)
2	Driving screws, cutting (using scissors)	Forceful upper-limb grip (bricklaying, hammering, beating, painting, meat boning)
3	Hammering, chopping, sawing (carpentrying)	Torque upper limb (driving screws)
4	Guiding or holding tools (painting, boring, drilling, ironing)	Lifting
5	Meat boning or filleting	Carrying
6	Packing or sorting (meat packing, assembly-line sorting)	Pushing
7	Assembly (small or delicate parts)	Pulling
8	Assembly, large or heavy parts (vehicle assembly or repair)	Forceful leg movement (machine treadling or foot switching)
9	Materials manipulation (pottering, carpet laying, masseuring, gardening)	Overhead work
10	Machine operation, heavy or forceful (using jackhammer or farming equipment)	Materials handling, not elsewhere specified
11	Machine operation, light or technical or scientific (sewing machine operation)	Bending (digging, with lifting as well)
12	Heavy lifting-carrying-pushing-pulling (digging, loading bins, patient lifting)	Sitting (if relevant, eg, not with keyboard use)
13	Light lifting-carrying-pushing-pulling (filing and office work, domestic cleaning)	Standing-walking
14	Coordinated whole body movement (climbing, participating in sports and athletics, performing arts)	Kneeling (carpet laying)
15	Driving: heavy plant, forklift	Twisting, postural
16	Driving: automobiles	Postural, not elsewhere specified (postures not covered under 11–15)
0	Other	Other
99	Uncodable	Uncodable

Roberts & McNamee (10). The kappa statistic measures the degree of agreement compared with that expected by chance. Values greater than 0.75 are generally considered excellent, values between 0.40 and 0.75 are fair to good, and those less than 0.40 represent poor agreement, little better than chance (9).

A single overall measure of reliability may have limited value in developing a new coding scheme, since poor reliability may arise from ambiguity in the definition of a particular category or through confusion between pairs of categories. To investigate the reliability of nominal scales, both Kraemer (11, 12) and Schouten (13, 14) suggested a summary matrix of coefficients, as contrasted with a single kappa statistic. In the matrix suggested by Schouten, the intraclass kappa coefficients allow particularly unreliable categories to be identified and therefore permit further investigation to show the extent to which they are confused with the other categories.

Results

Diagnoses

Of the 576 cases considered in this analysis, 47.6% (N=274) were disorders of the hand, wrist, or forearm, 12.3% (N=71) were of the shoulder, 11.6% (N=67) were of the neck or thoracic spine, 11.3% (N=65) were of the elbow, 11.1% were of the lumbar spine or trunk, 3.3%

(N=19) were of the foot or ankle, and 2.8% (N=16) were of the hip or knee.

Repetition and vibration

Low agreement between the reviewers was found for both repetition (kappa=0.27) and vibration (kappa=0.41), and a more-detailed analysis was not indicated.

Task codes

Not all the cases could be assigned a meaningful task code. Sometimes no task had been recorded (eg, "task not known"), and in others the description was inadequate (eg, "heavy manual activity" or "nursing"). Among 576 eligible cases, the proportion of tasks given a specific code (codes 1–16) by four coders was 91%, 89%, 95%, and 87%, respectively; 80% of the tasks (459 cases) was coded by all four. The remaining 20% were classified by at least one rater as either "uncodeable" (code=99) or "other unspecified" (code=0) and excluded from further analysis. The degree of interrater agreement between "codeable" and "uncodeable" was only fair (kappa = 0.44). In the intrarater reliability analysis, 84% of the activities were given a specific code in both ratings, and the degree of agreement between "codeable" and "uncodeable" was excellent (kappa = 0.85).

The overall agreement between the four coders on 16 task codes was good (kappa = 0.73). It was particularly high for keyboard work (kappa = 0.96), coordinated

whole-body movement ($\kappa = 0.91$), meat boning or filleting ($\kappa = 0.83$), and heavy lifting ($\kappa = 0.80$) (table 2). The agreement was poor for light and heavy assembly, and heavy machine operation (κ values all ≤ 0.40).

By comparing the initial four ratings with the derived "gold standard", the proportions of false positives (ie, coder recorded a specific task code not assigned in the consensus code) and of false negatives (code omitted) were calculated. Table 2 shows that all four coders had a generally high proportion of false positives and very few false negatives (ie, coders tended to include more codes that agreed in the consensus coding which acted as the "gold standard"). It can be noted that a very high proportion of the true positives were accounted for by three types of task (keyboard work, heavy lifting-carrying-pushing-pulling, and guiding or holding a tool), the analysis of other tasks (particularly assembling of small or large parts) being based on only very small numbers of "true" observations.

Levels of coding agreement were then considered in the following four main diagnostic groups: "upper limb", "neck-thoracic spine", "lumbar spine-trunk", and "lower limb". The overall agreement was good for tasks associated with disorders of the upper limb ($\kappa = 0.71$), neck ($\kappa = 0.82$), and lumbar spine ($\kappa = 0.74$), but less so for the lower limb ($\kappa = 0.63$).

The intrarater agreement was, in general, better than the interrater agreement, but the pattern of high or low κ values among 16 task codes was very similar, apart from category 8 (assembly of large or heavy parts).

A consistently low agreement was found in both intra- and interrater reliability for category 7 (assembly of small or delicate parts) and category 10 (machine operation, heavy or forceful).

Movement codes

Essentially the same methods were used to assess the reliability of the movement codes. Among 576 eligible cases, the proportion of tasks given a specific movement code (codes 1–16) by four coders was 89%, 88%, 95%, and 81%, respectively; 76% were coded by all four coders. The remaining 24% cases had been coded at least by one as "uncodeable" (code=99) or "other unspecified" (code=0) and were excluded from further analysis. The degree of interrater agreement between "codeable" and "uncodeable" was fair ($\kappa = 0.46$). The proportion given a specific code in both ratings in the intrareliability analyses was 80%; the agreement between "codeable" and "uncodeable" was also good ($\kappa = 0.85$).

As with the task codes, the agreement between coders for the 16 movement codes (table 3) was generally excellent ($\kappa = 0.79$), particularly for kneeling, fine hand (work), standing-walking, and lifting. For some, however, the agreement was very poor, for example, for materials handling, postures not elsewhere specified, and even carrying. The false positive rates were mostly high, and the false negatives were very low. Again, certain codes (fine hand movements, forceful movements of the upper limb and lifting) accounted for a very high

Table 2. Level of interrater agreement for the 16 categories of task codes. (95% CI = 95% confidence interval, FP = false positive, FN = false negative)

Task headline code	Summary κ	95%CI	Rater 1		Rater 2		Rater 3		Rater 4		"True" positives	
			FP	FN	FP	FN	FP	FN	FP	FN	N	%
			(%)	(%)	(%)	(%)	(%)	(%)	(%)	(%)		
1 Keyboard work	0.96	0.92–1.00	2.1	0.6	2.2	1.2	2.1	1.2	1.5	2.7	138	30
2 Driving screws, cutting	0.76	0.72–0.79	23.1	0.4	0.0	0.2	9.1	0.4	28.6	4.5	12	3
3 Hammering, chopping, sawing	0.75	0.71–0.78	12.5	0.2	23.6	0.5	12.5	0.2	16.7	1.1	15	3
4 Guiding or holding tool	0.56	0.53–0.60	20.6	3.7	12.1	6.8	25.0	6.0	27.5	5.1	58	13
5 Meat boning or filleting	0.83	0.79–0.87	0.0	0.0	23.0	0.0	0.0	0.4	0.0	0.0	6	1
6 Packing or sorting	0.73	0.70–0.77	0.0	0.5	0.0	1.6	28.0	0.2	17.4	0.0	19	4
7 Assembling, small parts	0.40	0.36–0.44	50.0	0.4	63.7	0.0	25.0	0.0	50.0	0.4	4	1
8 Assembling, large parts	0.37	0.33–0.40	33.3	0.2	70.0	0.0	50.0	0.2	76.9	0.0	3	1
9 Materials manipulating	0.49	0.45–0.53	28.4	0.4	20.0	0.9	53.3	1.1	58.3	0.4	12	3
10 Machine operating, heavy or forceful	0.39	0.35–0.42	42.3	0.2	71.8	1.2	68.7	2.5	38.5	1.8	16	4
11 Machine operating, light or technical	0.55	0.51–0.58	10.6	2.3	16.7	2.7	22.0	1.2	22.2	2.9	27	6
12 Heavy lifting-carrying-pushing-pulling	0.80	0.76–0.83	6.0	1.9	9.1	2.8	8.5	3.8	13.8	5.2	100	22
13 Light lifting-carrying-pushing-pulling	0.61	0.57–0.65	16.7	1.1	35.5	1.2	48.7	1.2	35.7	1.6	25	5
14 Coordinated whole-body moving	0.91	0.87–0.94	0.0	0.0	0.0	0.2	0.0	0.2	18.2	0.0	9	2
15 Driving: heavy plant, forklift	0.85	0.80–0.88	22.2	0.0	0.0	0.0	32.2	0.0	12.5	0.0	7	2
16 Driving: automobiles	0.81	0.77–0.85	14.3	0.2	12.5	0.0	25.0	0.2	22.2	0.0	7	2
Overall	0.73	0.77–0.85									459	100

Table 3. Level of interrater agreement for the 16 categories of movement codes. (95% CI = 95% confidence interval, FP = false positive, FN = false negative)

Movement Headline code	Summary kappa	95%CI	Rater 1		Rater 2		Rater 3		Rater 4		"True" positives	
			FP	FN	FP	FN	FP	FN	FP	FN	N	%
			(%)	(%)	(%)	(%)	(%)	(%)	(%)	(%)		
1 Fine hand	0.89	0.85–0.93	3.2	3.9	1.4	4.2	1.8	1.4	7.4	4.3	163	37.1
2 Forceful upper-limb grip	0.77	0.73–0.80	6.4	7.3	11.4	2.9	10.8	3.2	15.3	8.1	126	28.7
3 Torque, upper limb	0.63	0.59–0.67	25	0.2	0.0	0.0	33.3	0.0	62.5	0.2	4	0.9
4 Lifting	0.84	0.80–0.88	12.6	2.4	6.9	1.8	6.5	3.8	14.4	6.6	100	22.8
5 Carrying	0.33	0.29–0.36	100.0	0.2	100.0	0.2	83.3	0.0	100.0	0.2	1	0.2
6 Pushing	0.44	0.40–0.48	0.0	0.0	40.0	0.2	50.0	0.7	0.0	0.7	4	0.97
7 Pulling	0.66	0.62–0.70	28.6	0.0	25.0	0.5	0.0	0.2	60.0	0.7	5	1.1
8 Forceful leg movement	0.72	0.69–0.76	0.0	0.0	0.0	0.0	0.0	0.0	60.0	0.0	4	0.9
9 Overhead work	0.57	0.53–0.61	0.0	0.0	0.0	0.0	66.7	0.0	50.0	0.0	1	0.2
10 Materials handling, not elsewhere specified	0.10	0.06–0.14	92.3	0.0	50.0	0.0	–	0.2	100.0	0.2	1	0.2
11 Bending	0.58	0.54–0.62	75.0	0.5	33.3	0.2	60.0	0.2	50.0	0.2	3	0.7
12 Sitting	0.63	0.59–0.67	33.4	0.2	50.0	0.5	50.0	0.2	33.3	0.7	5	1.1
13 Standing-walking	0.85	0.81–0.88	0.0	0.2	0.0	9.3	8.3	0.2	0.0	0.9	12	2.7
14 Kneeling	0.94	0.90–0.98	0.0	0.0	0.0	0.0	0.0	0.0	20.0	0.0	4	0.9
15 Twisting, postural	0.53	0.49–0.57	0.0	0.0	0.0	0.0	0.0	0.2	–	0.5	2	0.5
16 Postural, not elsewhere specified	0.18	0.14–0.21	40.0	0.2	42.9	0.0	–	0.9	83.3	0.7	4	0.9
Overall	0.79	0.77–0.81									439	100.0

proportion of the true positives, with very small numbers for the analysis of other movement codes.

The interrater agreement for each of the four main body regions was generally high [upper limb (kappa = 0.79), neck (kappa = 0.74), lumbar spine (kappa = 0.65), and lower limb (kappa = 0.72)]. Inconsistencies did occur; for example, in one case of back pain attributed to the use of a visual display unit or desk work, one rater coded the movement as "fine hand" and three as "sitting". A problem also arose with disorders at more than one anatomical site. For example, a man with both shoulder pain and bursitis of the hip-knee had exposure described as "roofing and use of hand tools", but without mention of which aspect of roofing affected the lower limb (presumably kneeling, but not coded as such by all coders).

Overall, the intrarater agreement was better than the interrater agreement for the 16 movement codes. For categories 5 (carrying), 6 (pushing), 10 (materials handling, not elsewhere specified), and 16 (postural, not elsewhere specified), the interrater agreements were much lower than the intrarater agreements.

Further identification of unreliability

Matrix analyses (available from the authors upon request) showed a main diagonal consisting of binary coefficients that measured the reliability of each category relative to all others combined. The other (off-diagonal) elements were interclass kappa coefficients for pairs of categories, reflecting the degree of confusion between

them. The more negative the interclass coefficients, the higher the degree of confusion between the pairs of categories. For infrequently used categories, the small numbers in the subgroup resulted in wide confidence intervals, which reduced the effectiveness of this analysis. For this reason, our analysis focused on categories in which reasonable numbers of cases were included.

For the task codes, keyboard work, heavy lifting, and light lifting were clearly distinguished from the rest of the categories, whereas guiding or holding tools (code 4) were confused with assembly work (codes 7 & 8) and with heavy machine operation (code 10) [$k(4,7) = -0.90$, $k(4,8) = -1.13$ and $k(4,10) = -1.93$, respectively]. For the movement codes, fine hand movements were clearly distinguished from the rest of the categories, forceful upper-limb grip (code 2) was confused with pushing (code 6) [$k(2,6) = -0.90$] and materials handling (code 10) [$k(2,10) = -0.40$], and lifting (code 4) was confused with carrying (code 5) [$k(4,5) = -0.84$]. This evidence of apparent confusion between categories suggested that existing categories should be combined to increase reliability. For example, task codes 10 and 11 were later combined to the single category of "machine operation".

Discussion

The purpose of a clinically based surveillance scheme such as MOSS is primarily to obtain data on the incidence of specific diseases (by gender, age, occupation,

and geographic location) that, in the opinion of the reporting physician, have been caused or made substantially worse by work. Such information from specialist physicians is likely to be reasonably correct, and the resulting data help to assess the extent of work-related disease, to initiate further epidemiologic study, and to establish preventive strategies (7, 8). A secondary objective is to throw light on causation, but the accuracy of an attribution to specific work activities may be open to question. Moreover, rheumatologists may only see patients at a late stage of their illness, and thus initiating factors may not be described accurately or completely. Nevertheless, since the information on the exposure activities reflects a wide range of occupations and industries and covers a wide spectrum of exposures, analysis of the physical risk factors and possible trends over time may point to emerging or unsuspected hazards that warrant closer examination. For this purpose, the opinions recorded as to the factors suspected require specifically designed validation. A first step is to ensure that the reported information can be coded reliably for statistical analysis.

Although the kappa statistics calculated should be regarded as upper estimates (cases were only included if all 4 coders offered a specific code), the agreement between coders was good to excellent for most, but not all, of the factors examined. The single summary kappa statistic has been supplemented in this study by the calculation of false positive rates, the comparison of inter-reliability and intrareliability, and the inspection of the matrix of kappa coefficients.

In this study three main reasons for poor reliability were found by examining cases with poor agreement. Inconsistencies in coding resulted most clearly when, in disorders affecting more than one anatomic site, it was uncertain to which specific disorder the reporting factor applied. Next, the matrix analyses identified several categories that were repeatedly "confused" by the coders. Finally, occasional inconsistencies resulted from the vagueness of some of the information recorded by the reporting physician. Indeed it was this imprecision that made it unrealistic to define categories more precisely and led to the decision to collapse codes where confusion persisted.

Several procedures would probably improve the reliability of the coding scheme considerably. First, easily confused categories could be combined if a more meaningful category could be derived; failing this, better specification, with examples, might improve the reliability in certain poorly defined categories. Second, clear descriptions of tasks and movements could be sent to reporting physicians for guidance; revision of the reporting card to ask specifically about repetition and vibration may produce data on these dimensions that

would add considerably to the understanding of causation in some conditions. Third, double coding and reconciliation between coders should be used both in training and for subsequent quality control. This use may also increase the number of usable cases; although only 80% of the tasks, movements, or activities were given a usable code by all four coders in the present methodological study, at least one of the four coders was able to assign a code for 95% of the cases. Finally, since the reliability of less frequently coded categories remains uncertain, but not necessarily unimportant, their reliability should be reassessed after further use.

References

1. Cherry NM, Meyer JD, Holt DL, Chen Y, McDonald JC. Surveillance of work-related diseases by occupational physicians in the UK: OPRA 1996–1999. *Occup Med* 2000; 50(7):496–503.
2. McDonald JC, Meyer JD, Holt DL, Chen Y, Cherry NM. Disease surveillance by occupational physicians. *Occup Health Rev* 2001;90:10–4.
3. Health & Safety Commission. Health and safety statistics 1998/1999. Norwich: Her Majesty's Stationery Office, 1999.
4. Burdorf A, van der Beek A. Exposure assessment strategies for work-related risk factors for musculoskeletal disorders. *Scand J Work Environ Health* 1999;25 suppl 4:25–30.
5. Burdorf A, Van der Beek A. In musculoskeletal epidemiology are we asking the unanswerable in questionnaires on physical load? [editorial]. *Scand J Work Environ Health* 1999;25(2):81–3.
6. Pope DP, Silman AJ, Cherry NM, Pritchard C, Macfarlane GJ. Validity of a self-completed questionnaire measuring the physical demands of work. *Scand J Work Environ Health* 1998;24(5):376–85.
7. Cherry NM. Recent advances occupational disease. *BMJ* 1999;338:1397–9.
8. Cherry NM, Meyer JD, Chen Y, Holt DL, McDonald JC. The reported incidence of work-related musculoskeletal disease in the UK: MOSS 1997–2000. *Occup Med* 2001;51(7):450–5.
9. Fleiss JL. Statistical methods for rates and proportions. 2nd ed. New York (NY): John Wiley & Sons, 1981.
10. Roberts C, McNamee R. A matrix of kappa-type coefficients to assess the reliability of nominal scales. *Stat Med* 1998; 17:471–88.
11. Kraemer HC. Ramification of a population model for K as a coefficient of reliability. *Psychometrika* 1979;44:461–472.
12. Kraemer HC. Measurement of reliability for categorical data in medical research. *Stat Methods Med Res* 1993;1:183–99.
13. Schouten HJA. Measuring pair-wise agreement among many observers, II: some improvements and additions. *Biomet J* 1982;24:431–5.
14. Schouten HJA. Nominal scale agreement among observers. *Psychometrika* 1986; 51:453–66.

Received for publication: 19 December 2001