Review

Scand J Work Environ Health 2005;31(6):409-437

doi:10.5271/sjweh.947

Reproducibility and validity of workers' self-reports of physical work demands

by Stock SR, Fernandes R, Delisle A, Vézina N

Affiliation: INSPQ, 500 René-Lévesque Ouest, bur 16.05.03, Montréal, Quebec, H2Z 1W7, Canada. susan.stock@inspq.qc.ca

The following articles refer to this text: 2008;34(2):120-132; 2008;34(4):250-259; 2009;35(2):134-144; 2009;35(4):245-260; 2009;35(6):429-436; 2012;38(6):568-576; 2012;38(6):582-589; 2019;45(3):239-247; [online first; 27 August 2019]

Key terms: ergonomics; MSD; musculoskeletal disorder; physical work demand; questionnaire; reproducibility; review; self-assessment; self-report; systematic review; validity; worker; workload

This article in PubMed: www.ncbi.nlm.nih.gov/pubmed/16425584

Scand J Work Environ Health 2005;31(6):409–437

Reproducibility and validity of workers' self-reports of physical work demands

by Susan R Stock, MD,[1, 2] Rita Fernandes, MD,[3] Alain Delisle, PhD,[4] Nicole Vézina, PhD[5]

Stock SR, Fernandes R, Delisle A, Vézina N. Reproducibility and validity of workers' self-reports of physical work demands. *Scand J Work Environ Health* 2005;31(6):409–437.

The objective of this paper is to provide a systematic review of the reproducibility and validity of self-report questions concerning physical work demands. After a bibliographic search of Medline and Ergonomic Abstracts for 1980–2003, 15 articles meeting the eligibility criteria were reviewed for methodological quality; 82 formulations of questions on physical work demands were evaluated for reproducibility and 83 for validity. Questions evaluated for both reproducibility and validity that performed well in both sets of studies included those on duration or presence of sitting and standing posture, the presence of walking, kneeling or squatting postures, duration or frequency of hands above shoulders, manual handling of more than or less than 10 kg, general level of physical effort, presence and duration of whole-body vibration, and duration of the use of visual display terminals. Suggestions for improving the design of reproducibility and validity studies and directions for future research in physical workload measurement are proposed.

Key terms ergonomics; musculoskeletal disorder; questionnaire; systematic review; workload.

Musculoskeletal disorders are an important public health problem in both industrialized and developing countries, and concern is growing about their social and economic consequences, particularly in the work environment. There is considerable evidence of a causal association between musculoskeletal disorders and physical work demands (1). Nonetheless the measurement of physical work demands has posed a major challenge for epidemiologic research in this area. Some occupational health researchers have questioned the validity and usefulness of using questionnaires to measure physical work demands and have assumed that observational methods or direct measurements are likely to be more accurate than workers' self-reports of physical workload.

However, observations and direct measurements of the various dimensions of physical workload also have major drawbacks and limitations. One of their main disadvantages is that they are very resource demanding, especially in large studies. Thus the number and scope of measurements are often reduced to a minimum, and such reduction may threaten their validity. Frequently, direct measurements and observations are taken over a short period of time and sample a small proportion of the workday for each worker. The representativeness of these measurements for the whole workday, week, or month may be very limited, particularly for tasks involving a high variability of exposure over the course of a workday or between different days. The same job title can include jobs with a wide variety of tasks, whose frequency and duration in a workday can change considerably among workers. Exposure may differ with variations in the work methods and strategies that different workers adopt (2). Variability of exposure may also be influenced by the characteristics of the work processes and production context that can lead to different peak and cumulative exposure (3).

It is therefore particularly challenging to capture the complexity and full range of relevant physical work

1 Institut national de santé publique du Québec (INSPQ) (Quebec National Institute of Public Health), Montreal, Québec, Canada.
2 McGill University, Montreal, Quebec, Canada.
3 Universidade Federal da Bahia (UFBA) (Federal University of Bahia), Collective Health Institute, Brazil.
4 Institut de recherche Robert-Sauvé en santé et en sécurite du travail (IRSST) (Robert-Sauvé Occupational Health and Safety Research Institute), Montréal, Québec, Canada.
5 Université du Québec à Montréal (UQAM) (University of Quebec at Montreal), Montreal, Québec, Canada.

Reprint requests to: Dr Susan Stock, INSPQ, 500 René-Lévesque Ouest, bur 16.05.03, Montréal, Quebec, H2Z 1W7, Canada. [E-mail: susan.stock@inspq.qc.ca]

demands in studies that present great variability of tasks and work conditions and a wide range of work methods among workers. In many work contexts it is impossible to measure the intensity, frequency, and duration of all the relevant physical work exposures. When the range and variability of the work situation is inadequately captured by limited individual observations and direct measures, self-reported measures may be able to provide a more complete picture of the overall physical work demands. The use of complementary self-reported and observational and direct measurement methods may therefore enhance the capacity to capture true physical demand exposure. Such combinations of methods are used by some researchers and have been recommended in ergonomic studies (4).

Self-reported measures may also be particularly relevant in the context of large population-based surveys that include many job titles, multiple workplaces, and a very wide variety of occupational tasks and are designed to monitor general trends over the time of exposure to important determinants of musculoskeletal disorders such as physical load. Questionnaires offer the possibility of studying a great number of persons at a modest cost, as well as allowing the investigation of a large number of variables. Questionnaires may be the most practical instrument for measuring exposure at work in this type of study and for obtaining retrospective data on exposure (5–7).

A wide variety of such questionnaires has been developed. To date, there has been no systematic review of the reproducibility and validity of self-report questions measuring physical work demands.

The objective of this paper is to identify which questions on physical work demands have been accurately and reliably answered by workers and which questions have been shown to have poor reproducibility or validity. It also reflects on the choice of response scales, the challenges associated with evaluating the reproducibility and validity of work demand questions and provides suggestions for future studies in the development of such questions and their validation.

Material and methods

A systematic review was carried out to identify articles that have presented results on the validity and reproducibility of questions measuring physical work demands.

The bibliographic search strategy included searches of Medline and Ergonomics Abstracts databases for the years 1980 to 2003 using the following key words: "questionnaire or self-report; reproducibility or validity or sensitivity or specificity or responsiveness to change or sensitivity to change; physical or mechanical or biomechanical or ergonomics or posture or force or repetition or lifting or manual material handling or vibrations or static or sitting or standing; exposure or load or workload or overload; musculoskeletal disorders or musculoskeletal diseases, musculoskeletal abnormalities or shoulder or neck or hand or wrist or back or elbow or arms or cervical or lumbar or muscular strength or muscular endurance". As well, bibliographies of relevant articles were reviewed, and all relevant references not previously identified by the database searches were sought. Once identified, the abstracts of each reference or the full article of references with no available abstract were scrutinized, and those meeting the following selection criteria were selected for more-detailed review: (i) articles evaluating the reproducibility or validity of self-administered or interviewer-administered questions assessing specific physical demands at current work and (ii) articles published in English, French, Spanish, or Portuguese. After the selection, two epidemiologists (RF, SS) evaluated the articles using the following methodological criteria for retaining them: (i) original studies and not reviews, (ii) studies that evaluated the test re-test reliability or validity of single questions evaluating the presence or level of physical work demands, (iii) no studies evaluating indices of physical work demands, (iv) studies with the evaluated work demand questions explicitly stated in the paper or available in another reference and the response scales described, (v) studies with the reproducibility or validity evaluated using statistical methods (eg, kappa, Pearson correlation coefficient, intraclass correlation coefficient (ICC), Spearman correlation coefficient (r_S), sensitivity and specificity), (vi) studies with at least 20 participants assessed in each reproducibility or validity study group, (vii) reproducibility studies in which short-term (<2 months) reproducibility was evaluated [in studies in which some subgroups were evaluated for short-term reproducibility and others for long-term reproducibility (8) or both short- and long-term reproducibility were evaluated (9) only the results pertaining to short-term reproducibility were included in this review], (viii) exclusion of work demand questions measuring more than one physical demand by the same question or ambiguous or problematic wording, and (ix) exclusion of questions measuring the contribution of a physical work demand to musculoskeletal symptoms.

Studies that met the criteria were evaluated for the following methodological issues: (i) proportion of persons responding to both administrations, (ii) sample size (iii) variability of the distribution of exposure, (iv) type of response scales, (v) confounders or effect modifiers taken into account (possible influence on results of gender, age, education level and presence or absence of musculoskeletal disorders), (vi) in validity studies, the validity of reference methods, and (vii) in validity studies, time interval between the assessments by questionnaire and by the reference method.

For interpreting the reproducibility and validity of categorical results, intervals stated by Landis & Koch (10) and Altman (11) were used to classify kappa results (0.81–1.0 = excellent agreement, 0.61–0.80 = substantial agreement, 0.40–0.60 = moderate agreement, 0.21–0.40 fair agreement, and ≥0.20 = slight to poor agreement). These same intervals were used to classify the concurrent criterion validity correlation coefficient results. To classify reproducibility correlation coefficient results, the criteria proposed by Currier (12) were used (0.90–0.99 = excellent correlation, 0.80–0.89 = good to very good, 0.70–0.79 = fair, <0.70 = poor).

Results

On the basis of the initial eligibility criteria, 31 studies were selected. After an evaluation taking into account the criteria for retention, 16 articles were excluded. Among the 15 articles retained, 5 were reproducibility studies, 8 were validity studies, and 2 evaluated both validity and reproducibility. From these studies, 82 formulations of questions on physical work demands were evaluated for reproducibility and 83 for validity. For some of the formulations of questions, there were separate results for several different response scales (eg, using a 5-point frequency scale or a dichotomized version of the scale). As well there were results from different studies using the same question formulation. Among the 83 question formulations evaluated for validity, 10 were also evaluated for reproducibility.

Reproducibility studies

Table 1 summarizes the seven studies on the reproducibility of work demand questions, and table 2 reveals the main findings of each study. These studies included reproducibility results for physical demand questions on general body posture (eg, sitting, standing, walking, etc), posture of the neck, shoulders, arms, wrists or hands, repetitive movements, hand use, vibration, and level of overall physical effort.

General body postures. Overall, questions concerning sitting and standing showed good-to-excellent reproducibility, with one exception. Questions on the duration or presence of sitting posture were tested in five studies (5, 8, 9, 13, 14). Weighted kappa or the ICC for the questions using a five- or six-point response scale measuring the proportion of the workday sitting (8, 13, 14) or a less-specific three-point scale (5) was generally excellent (>0.80) with the exception of two small subgroups in the Halpern et al study (8), which, though lower, had very acceptable kappa results (kappa 0.51 and 0.64). The only question about sitting that did not perform as well was the one using a 10-cm visual analogue scale measuring the duration with verbal anchors at each end (ICC 0.64) (9). Contrary to the other studies, that of Balogh et al (5), in addition to a question on sitting, included a specific question on standing posture that had a weighted kappa (K_w) of 0.66.

Two studies had good reproducibility for the questions on walking at work, showing a K_w of 0.67 (5), with a three-point, less specific scale, and an ICC of 0.78 and 0.80 (14) with duration and distance scales, respectively.

Questions on kneeling or squatting at work were tested in three studies. The studies using six-point duration scales had low ICC values (0.50–0.69) (14) and mixed results for the different subgroups of the Halpern et al, (8) study (kappa 0.32–0.58). Balogh et al (5), using their three-point scale, had a very good kappa (0.71). These results suggest that this question can be reliably answered when it asks workers whether they kneel or squat at work, but it does not allow reliable answers when it asks for the proportion of the workday doing so.

Similarly, in studies that tested jumping or climbing, the agreement statistic was very good (K_w 0.71) with

Table 1. Summary of studies on the reliability of questions measuring physical work demands. (VDU = visual display units, ICC = intraclass correlation coefficient, ANOVA = analysis of variance, REBUS = Rehabiliterings-Behovs-Undersölkningen i Stockholms län)

Study	Population	Physical work demands measured and type of questionnaire	Statistics	Methodological issues
Balogh et al, 2001 (5)	207 Swedish participants, gainfully employed >30 hours/week, participants in the Malmö Shoulder Neck Study (industry, type of work not described)	24 questions (from MUSIC 1 study, with revised 3-point quantity response scale) concerning mechanical exposure of the shoulder-neck region: body postures (11 items), repetitive movements (2 items), precise movements (1 item), vibration (2 items), manual materials handling (5 items), jumping and climbing (1 item), unexpected high loads (1 item), level of physical activity at work (1 item)	Weighted kappa	Interval between administrations: 2 weeks; variability of distribution of exposure: not stated; confounders/effect modifiers: not described for reliability study; proportion of participants responding to both administrations: 89.2%

(continued)

Table 1. Continued.

Study	Population	Physical work demands measured and type of questionnaire	Statistics	Methodological issues
Halpern et al, 2001 (8)	Convenience sample of 106 full-time American employees from an electric utility company with a wide variety of jobs or exposures (drafters, customer service representatives, warehouse workers, mechanics); group 1 (patients): 24, group 3 (supervisors 2): 33, group 4 (workers): 29	21 self-administered questions concerning handling and lifting activities, load features (bulk or grip), whole body and body regions, postures, environmental demands (vibration, slippery surfaces, elevated surfaces)	Kappa	Interval between administrations: group 1: 7 days (range 2–12), group 3: 9 days (range 2–21), group 4: 10 days (range 5–42); variability of distribution of exposure: 40% of the items did not have responses in one or two of the high-end scores of the rating scales; confounders/effect modifiers: not described; proportion of participants responding to both administrations: 82.8% for group 1, not stated for the other groups; other: small sample size in each group
Karlqvist et al, 1996 (16)	100 (64 women, 36 men) Swedish workers from 2 research institutes working on VDU for an average of 40% of the workday	2 questions: duration of VDU work, mouse use in VDU work; 3 questions: location of keyboard and mouse, height of elbow from keyboard	Pearson correlation coefficient; kappa	Interval between administrations: 1–2 weeks; variability of distribution of exposure: not stated; confounders/effect modifiers: gender; proportion of participants responding to both administrations: 90%
Leijon et al, 2002 (13)	203 employees (102 women, 101 men) sampled from 80 worksites in 5 Swedish counties, selected through a 4-step process to ensure variation in forms of work organization, type, size, and geographic location of employers, variation in worktasks, postures and qualification levels; to ensure similar distribution of gender, and similar distribution by age, level of education and type of work between the men and women; part of the Modern Work and Living Conditions for Women and Men (MOA) study	7 self-administered questions from the Stockholm Public Health Questionnaire concerning general physical activity, repetitive movements, general body postures, specific awkward postures, manual materials handling	Weighted kappa; percentage of full agreement; Spearman correlation coefficient; χ^2 to compare dichotomized version of responses	Interval between administrations: approximately 3 weeks for all of the participants, variability of distribution of exposure: skewed distribution (ie, frequency of highly exposed persons low for 4 questions [work with hands above shoulder level, bent trunk, manual handling of loads ≥10 kg, general physical activity)]; reference method rated 44% to 76% of the participants in the lowest exposure category (ie, unexposed) for these same 4 questions; confounders/effect modifiers: not described in reliability study; proportion of participants responding to both administrations: 93–97%
Spielholz et al, 1999 (15)	71 (41 women, 30 men) American tree nursery workers (out of 96), working in the field or inside the nursery	10 questions concerning upper extremities: grasping, wrist bending and exertion, neck bending, arms raised	ICC; weighted kappa; repeated measures ANOVA to compare within-participant differences for continuous scales; Wilcoxon paired-signed ranks test to compare within-participant differences for categorical scales	Interval between administrations: 1 week; variability of distribution of exposure: not stated; confounders/effect modifiers: gender and workplace; proportion of participants responding to both administrations: 74%; other: 25% respondents completed questionnaire in Spanish—no mention of validation of translation
Torgén et al, 1997 (9)	44 (24 women, 20 men) Swedish workers; subsample of REBUS-93 study in Stockholm	12 questions (from MUSIC I, with some response scales changed to express number of days per week exposed), supplemented with drawings illustrating physical activities at work: perceived exertion, sitting, VDT use, whole-body vibration, handheld vibrating tools, precision work, repetitive work, body postures (3 items), lifting and carrying (2 items)	ICC	Interval between administrations: 2 weeks; variability of distribution of exposure: distributions skewed towards low exposure for whole-body vibration, hand vibration, and precision work; confounders/effect modifiers: not described for 2-week reproducibility study; proportion of participants responding to both administrations: 100%
Wiktorin et al, 1996 (14)	343 (126 women, 217 men) Stockholm workers with at least 1 year of work experience and high, medium or low physical workload from 30 job titles	33 questions: general body posture, specific postures of trunk and head, material handling, hand use, vibration, level of physical effort	ICC; kappa for dichotomous versions of response scale	Interval between administrations: 2 weeks; variability of distribution of exposure: responses were concentrated at one end of the scale for many variables (frequency of highly exposed participants was low in these cases); confounders/effect modifiers: gender, presence of musculoskeletal complaints, age, educational level; proportion of participants responding to both administrations: 95%; no difference found between participants and nonparticipants with regard to job title, age, gender, education, musculoskeletal complaints; other: individual kappa results for the full version of the response scales not presented, only kappa ranges (individual kappa results provided for dichotomized version of the scales)

Table 2. Results of studies on the reliability of the questions measuring physical work demands. (ICC = intraclass correlation coefficient, K_w = weighted kappa; r = correlation coefficient, r_S = Spearman correlation coefficient, VAS = visual analogue scale, low education = grade school only, medium–high education = vocational and senior high school, high education = college, university)

Type of demand	Exposure variables	Response scales	Reliability statistics/results				
			K_w	ICC	r_S	Agreement (%)	r
General body posture							
Sitting and standing							
Leijon et al, 2002 (13)	Sitting work posture in current job	Duration: 5-point, percentage of workday	0.92	·	0.92	74	·
Wiktorin et al, 1996 (14)	Work involving sitting	Duration: 6-point, fraction of worktime	·	0.90	·	·	·
Torgén et al, 1997 (9)	Proportion of the day spent sitting	VAS: duration (anchors: not at all – all the time)	·	0.64	·	·	·
Halpern et al, 2001 (8)	Proportion of the time spent sitting	Duration: 6-point, percentage of worktime	·	·	·	·	·
Patients group 1			0.51	·	·	·	·
Supervisor group 3			0.83	·	·	·	·
Worker group 4			0.64	·	·	·	·
Balogh et al, 2001 (5)	Work involving sitting	Quantity: 3-point [a]	0.84	·	·	·	·
Balogh et al, 2001 (5)	Work involving standing	Quantity: 3-point [a]	0.66	·	·	·	·
Walking							
Balogh et al, 2001 (5)	Work involveing walking	Quantity: 3-point [a]	0.67	·	·	·	·
Wiktorin et al, 1996 (14)	Work involving walking	Duration: 6-point, fraction of worktime	·	0.78	·	·	·
Wiktorin et al, 1996 (14)	Work involving walking	Distance (km): 5-point	·	0.80	·	·	·
Kneeling or squatting							
Wiktorin et al, 1996 (14)	Work involving kneeling or squatting	Duration: 6-point fraction of worktime					
All participants			·	0.59	·	·	·
Without low-back complaints			·	0.68	·	·	·
With low-back complaints			·	0.52	·	·	·
Men			·	0.69	·	·	·
Women			·	0.50	·	·	·
Balogh et al, 2001 (5)	Work involving kneeling or squatting	Quantity: 3-point [a]	0.71	·	·	·	·
Halpern et al, 2001 (8)	Proportion of the time spent kneeling or squatting	Duration: 6-point, percentage of worktime					
Patients group 1			0.32	·	·	·	·
Supervisor group 3			0.58	·	·	·	·
Worker group 4			0.45	·	·	·	·
Jumping and climbing							
Balogh et al, 2001 (5)	Work involving jumping and climbing between different height levels	Quantity: 3-point [a]	0.71	·	·	·	·
Wiktorin et al, 1996 (14)	Work involving jumping and climbing between different height levels	Frequency: 5-point times/hour	·	0.71	·	·	·
Supine							
Balogh et al, 2001 (5)	Lying down	Quantity: 3-point [a]	0.72	·	·	·	·
Postures involving specific body parts							
Neck posture							
Spielholz et al, 1999 (15)	Neck bending	Duration: 100-mm VAS, percentage of workday	·	0.42	·	·	·
Spielholz et al, 1999 (15)	Neck bending	Frequency: 7-point numerical	0.29	·	·	·	·
Wiktorin et al, 1996 (14)	Holding head bent forward	Duration: 6-point fraction of worktime	·	0.63	·	·	·
Wiktorin et al, 1996 (14)	Holding head bent backward	Duration: 6-point fraction of worktime	·	0.53	·	·	·
Wiktorin et al, 1996 (14)	Holding head rotated a lot	Duration: 6-point fraction of worktime					
All participants			·	0.55	·	·	·
Low education			·	0.19	·	·	·
Medium–high education			·	0.67	·	·	·
High education			·	0.68	·	·	·
Balogh et al, 2001 (5)	Head bent backward	Quantity: 3-point [a]	0.60	·	·	·	·
Balogh et al, 2001 (5)	Head bent forward a little	Quantity: 3-point [a]	0.49	·	·	·	·
Balogh et al, 2001 (5)	Head bent forward a lot	Quantity: 3-point [a]	0.55	·	·	·	·

(continued)

Table 2. Continued.

Type of demand	Exposure variables	Response scales	Reliability statistics/results				
			K_W	ICC	r_S	Agreement (%)	r
Trunk posture							
Leijon et al, 2002 (13)	Trunk bent in current job	Duration: 4-point, percentage of workday	0.74	·	0.69	84	·
Torgén et al, 1997 (9)	Hands below knee >30minutes/day	Frequency: 5-point, days/week or month	·	0.61	·	·	·
Balogh et al, 2001 (5)	Back bent forward a lot	Quantity: 3-point [a]	0.59	·	·	·	·
Wiktorin et al, 1996 (14)	Back bent forward a lot	Duration: 6-point, fraction of worktime					
All participants			·	0.59	·	·	·
Without low-back complaints			·	0.68	·	·	·
With low-back complaints			·	0.49	·	·	·
Halpern et al, 2001 (8)	Bending the trunk forward hands	Duration: 6-point, percentage of worktime below knee height					
Patients group 1			0.70	·	·	·	·
Supervisor group 3			0.49	·	·	·	·
Worker group 4			0.62	·	·	·	·
Halpern et al, 2001 (8)	Bending the trunk forward slightly	Duration: 6-point, percentage of worktime					
Patients group 1			0.44	·	·	·	·
Supervisor group 3			0.52	·	·	·	·
Worker group 4			0.41	·	·	·	·
Wiktorin et al, 1996 (14)	Back bent forward a little	Duration: 6-point, fraction of worktime	·	0.52	·	·	·
Balogh et al, 2001 (5)	Back rotated a lot	Quantity: 3-point [a]	0.64				
Leijon et al, 2002 (13)	Bent or twisted body several times/hour	Frequency: 5-point, days/week or month	0.83	·	0.78	76	·
Torgén et al, 1997 (9)	Bent or twisted body several times/hour	Frequency: 5-point, days/week or month	·	0.74	·	·	·
Halpern et al, 2001 (8)	Twisting the trunk (>45 degrees) and bending sideways	Duration: 6-point, percentage of worktime					
Patients group 1			0.63	·	·	·	·
Supervisor group 3			0.57	·	·	·	·
Worker group 4			0.52	·	·	·	·
Arm posture							
Leijon et al, 2002 (13)	Hands above shoulder in current job	Duration: 4-point, percentage of workday	0.79	·	0.80	80	·
Torgén et al, 1997 (9)	Hands above shoulder >30 minutes/day	Frequency: 5-point, days/week or month	·	0.75	·	·	·
Wiktorin et al, 1996 (14)	Work involving work above shoulder level	Duration: 6-point, fraction of worktime					
All participants			·	0.54	·	·	·
Low education			·	0.20	·	·	·
Medium–high education			·	0.67	·	·	·
High education			·	0.53	·	·	·
Balogh et al, 2001 (5)	Arms elevated or stretched forward	Quantity: 3-point [a]	0.60	·	·	·	·
Spielholz et al, 1999 (15)	Arms raised	Frequency: 7-point, numerical	0.32	·	·	·	·
Spielholz et al, 1999 (15)	Arms raised	Duration: 100-mm VAS, percentage of workday	·	0.24	·	·	·
Hand or wrist posture							
Spielholz et al, 1999 (15)	Wrist bending	Duration: 100-mm VAS, percentage of workday	·	0.62	·	·	·
Spielholz et al, 1999 (15)	Wrist bending	Frequency: 7-point, numerical	0.40	·	·	·	·
Repetitive movements							
Leijon et al, 2002 (13)	Repetitive movements several times/hour	Duration: 5-point, percentage of workday	0.80	·	0.78	65	·
Balogh et al, 2001 (5)	Repetitive finger movements	Quantity: 3-point [a]	0.65	·	·	·	·
Balogh et al, 2001 (5)	Repetitive arm movements	Quantity: 3-point [a]	0.64	·	·	·	·
Torgén et al, 1997 (9)	Repetitive hand or finger movements several times/minute, >2 hour/day	Frequency: 5-point, days/week or month	·	0.64	·	·	·
Wiktorin et al, 1996 (14)	Same finger movements several times/minute	Duration: 6-point, fraction of worktime	·	0.71	·	·	·
Wiktorin et al, 1996 (14)	Same hand movements several times/minute	Duration: 6-point, fraction of worktime	·	0.65	·	·	·
Hand use							
Balogh et al, 2001 (5)	Making precise movements (eg, dentist, microscopist, musician)	Quantity: 3-point [a]	0.68	·	·	·	·
Wiktorin et al, 1996 (14)	Making precise movements (eg, dentist, microscopist, musician)	Duration: 6-point, fraction of worktime	·	0.70	·	·	·
Torgén et al, 1997 (9)	Precision work >2 hour/day	Frequency: 5-point, days/week or month	·	0.36	·	·	·
Spielholz et al, 1999 (15)	Grasping duration	Duration: 100-mm VAS, percentage of workday	·	0.59	·	·	·
Spielholz et al, 1999 (15)	Grasping frequency	Frequency: 7-point, numerical	0.36	·	·	·	·
Spielholz et al, 1999 (15)	Grasping force	Intensity: 100-mm VAS, percentage of maximum	·	0.69	·	·	·
Spielholz et al, 1999 (15)	Hand/wrist exertion	Intensity: 100-mm VAS, percentage of maximum	·	0.64	·	·	·

(continued)

Table 2. Continued.

Type of demand	Exposure variables	Response scales	Reliability statistics/results				
			K_W	ICC	r_S	Agree-ment (%)	r
Material handling							
Leijon et al, 2002 (13)	Handling ≥10 kg in current job	Frequency: 4-point, times/day	0.81	·	0.83	76	·
Balogh et al, 2001 (5)	Lifting and carrying a few 100 grams	Quantity: 3-point [a]	0.64	·	·	·	·
Halpern et al, 2001 (8)	Handling large and bulky objects at arms length	Duration: 6-point, percentage of worktime					
Patients group 1			0.72	·	·	·	·
Supervisor group 3			0.61	·	·	·	·
Worker group 4			0.45	·	·	·	·
Halpern et al, 2001 (8)	Handling objects difficult to grip—unstable, no handles	Duration: 6-point, percentage of worktime					
Patients group 1			0.80	·	·	·	·
Supervisor group 3			0.51	·	·	·	·
Worker group 4			0.58	·	·	·	·
Halpern et al, 2001 (8)	Carrying load with one hand	Duration: 6-point, percentage of worktime					
Patients group 1			0.59	·	·	·	·
Supervisor group 3			0.58	·	·	·	·
Worker group 4			0.69	·	·	·	·
Halpern et al, 2001 (8)	Carrying loads over 10 pounds [b], more than 40 feet [c]	Duration: 6-point, percentage of worktime					
Patients group 1			0.42	·	·	·	·
Supervisor group 3			0.67	·	·	·	·
Worker group 4			0.40	·	·	·	·
Halpern et al, 2001 (8)	Pushing and pulling loads (carts, drawers)	Duration: 6-point, percentage of worktime					
Patients group 1			0.45	·	·	·	·
Supervisor group 3			0.54	·	·	·	·
Worker group 4			0.37	·	·	·	·
Balogh et al, 2001 (5)	Liftingor carrying 1–5 kg	Quantity: 3-point [a]	0.71	·	·	·	·
Wiktorin et al, 1996 (14)	Lifting or carrying 1–5 kg	Frequency: 5-point, times/hour	·	0.62	·	·	·
Wiktorin et al, 1996 (14)	Carrying, pushing or pulling 1–5 kg	Duration: 6-point, fraction of worktime					
All participants			·	0.54	·	·	·
Men			·	0.62	·	·	·
Women			·	0.49	·	·	·
Halpern et al, 2001 (8)	Lifting <10 pounds [b]	Frequency: 5-point times/hour					
Patients group 1			0.65	·	·	·	·
Supervisor group 3			0.58	·	·	·	·
Worker group 4			0.77	·	·	·	·
Torgén et al, 1997 (9)	Lifting or carrying 5–15 kg	Frequency: 5-point, days/week or month	·	0.89	·	·	·
Balogh et al, 2001 (5)	Lifting or carrying 6–15 kg	Quantity: 3-point [a]	0.73	·	·	·	·
Wiktorin et al, 1996 (14)	Lifting or carrying 6–15 kg	Frequency: 5-point, times/hour	·	0.66	·	·	·
Wiktorin et al, 1996 (14)	Carrying, pushing or pulling 6–15 kg	Duration: 6-point, fraction of worktime					
All participants			·	0.57	·	·	·
Low education			·	0.30	·	·	·
Medium–high education			·	0.61	·	·	·
High education			·	0.67	·	·	·
Halpern et al, 2001 (8)	Carrying 10–30 pounds [b]	Duration: 6-point, percentage of worktime					
Patients group 1			0.63	·	·	·	·
Supervisor group 3			0.83	·	·	·	·
Worker group 4			0.59	·	·	·	·
Halpern et al, 2001 (8)	Lift 10–30 pounds [b]	Frequency: 5-point, times/hour					
Patients group 1			0.63	·	·	·	·
Supervisor group 3			0.61	·	·	·	·
Worker group 4			0.72	·	·	·	·
Torgén et al, 1997 (9)	Lifting/carrying >15 kg	Frequency: 5-point, days/week or month	·	0.83	·	·	·
Halpern et al, 2001 (8)	Carrying >30 pounds [b]	Duration: 6-point, percentage of worktime					
Patients group 1			0.59	·	·	·	·
Supervisor group 3			0.81	·	·	·	·
Worker group 4			0.71	·	·	·	·
Halpern et al, 2001 (8)	Lift > 30 pounds [b]	Frequency: 5-point, times/hour					
Patients group 1			0.68	·	·	·	·
Supervisor group 3			0.69	·	·	·	·
Worker group 4			0.74	·	·	·	·

(continued)

Table 2. Continued.

Type of demand	Exposure variables	Response scales	Reliability statistics/results				
			K_W	ICC	r_S	Agreement (%)	r
Balogh et al, 2001 (5)	Lifting or carrying 16–45 kg	Quantity: 3-point [a]	0.75	·	·	·	·
Wiktorin et al, 1996 (14)	Lifting or carrying 16–45 kg	Frequency: 5-point, times/hour	·	0.63	·	·	·
Wiktorin et al, 1996 (14)	Carrying, pushing or pulling 16–45 kg	Duration: 6-point, fraction of worktime	·	0.59	·	·	·
Balogh et al, 2001 (5)	Lifting or carrying >45 kg	Quantity: 3-point [a]	0.64	·	·	·	·
Wiktorin et al, 1996 (14)	Lifting or carrying >45 kg	Frequency: 5-point, times/hour	·	0.53	·	·	·
Wiktorin et al, 1996 (14)	Carrying, pushing or pulling >45 kg	Duration: 6-point, fraction of worktime	·	0.71	·	·	·
Balogh et al, 2001 (5)	Unexpected great loads	Quantity: 3-point [a]	0.71	·	·	·	·
Level of physical effort at work							
Leijon et al, 2002 (13)	General physical activity during the last year	Intensity: 4-point	0.85	·	0.86	77	·
Balogh et al, 2001 (5)	Degree of physical activity required in work	Quantity: 3-point [a]	0.80	·	·	·	·
Wiktorin et al, 1996 (14)	Degree of physical activity required in work	Intensity: 5-point	·	0.87	·	·	·
Torgén et al, 1997 (9)	Perceived general exertion in current job	Borg scale, 15-point	·	0.67	·	·	·
Wiktorin et al, 1996 (14)	Perceived general exertion in current job	Borg scale, 15-point	·	0.82	·	·	·
Physical environment							
Whole-body vibration							
Balogh et al, 2001 (5)	Work on jolting surfaces (eg, vibrating floor, vehicle seat, ship floor)	Quantity: 3-point [a]	0.75	·	·	·	·
Wiktorin et al, 1996 (14)	Work on jolting surfaces (eg, vibrating floor, vehicle seat, ship floor)	Duration: 6-point, fraction of worktime	·	0.70	·	·	·
Torgén et al, 1997 (9)	Whole-body vibration	VAS: duration (anchors: not at all – all the time)	·	0.95	·	·	·
Handheld vibrating tools							
Balogh et al, 2001 (5)	Work with handheld tools that vibrate or give impact	Quantity: 3-point [a]	0.73	·	·	·	·
Torgén et al, 1997 (9)	Work with handheld tools that vibrate or give impact	VAS: duration (anchors: not at all – all the time)	·	0.85	·	·	·
Wiktorin et al, 1996 (14)	Work with handheld tools that vibrate or give impact	Duration: 6-point, fraction of worktime	·	0.84	·	·	·
Halpern et al, 2001 (8)	Operating powered hand tools (drills, saws, jack hammers)	Duration: 6-point, percentage of worktime					
Patients group 1			0.83	·	·	·	·
Supervisor group 3			0.93	·	·	·	·
Worker group 4			0.65	·	·	·	·
Other exposures							
Halpern et al, 2001 (8)	Working on slippery or uneven surfaces	Duration: 6-point % worktime					
Patients group 1			0.51	·	·	·	·
Supervisor group 3			0.56	·	·	·	·
Worker group 4			0.54	·	·	·	·
Visual display terminal use							
Wiktorin et al, 1996 (14)	Work involving visual display unit (computer)	Duration: 6-point fraction of worktime	·	0.95	·	·	·
Torgén et al, 1997 (9)	Proportion of day using VDT	Duration: VAS (anchors: not at all – all the time)	·	0.93	·	·	·
Karlqvist et al, 1996 (16)	Percentage of total work time on VDT work	Duration: open question (0–100%)	·	·	·	·	0.90
Karlqvist et al, 1996 (16)	Percentage of mouse use in VDU work	Duration: open question (0–100%)	·	·	·	·	0.75
Karlqvist et al, 1996 (16)	Location of keyboard (based on a diagram with a grid of 14–28 squares)	Dichotomized responses: optimal—nonoptimal	·	·	0.79	·	·
Karlqvist et al, 1996 (16)	Location of mouse (based on a diagram with a grid of 14–28 squares)	Dichotomized responses: optimal—nonoptimal	·	·	0.81	·	·
Karlqvist et al, 1996 (16)	Distance between elbows and keyboard	Dichotomized responses (distance in cm:–4 to 0; >0 to 4)	·	·	0.39	·	·

[a] Three-level scale: nothing or not at all, somewhat and a great deal (scores 1, 2 and 3, respectively).
[b] 1 pound = 0.45 kilograms.
[c] 1 foot = 30.48 centimeters.

the less-specific three-point scale (5) and just fair when based on a five-point frequency scale (14), these findings suggesting that workers may be unable to estimate reliably how many times per hour they jump or climb at work.

Postures of specific body regions. Overall, the reproducibility of the questions on the duration or frequency of postures involving specific body regions was not as good as that of questions on general body postures. Neck posture was tested in three studies (5, 14, 15). For the multilevel duration or frequency response scales (14, 15), poor reproducibility was observed (ICC 0.42–0.63, kappa 0.29). Moderate reproducibility (kappa 0.49–0.60) was observed for the less-specific three-point scale and questions about head bent backwards or forwards (5).

Questions on forward bending of the trunk were tested in five different studies. The results ranged from acceptable to good in three studies using four- or six-point duration scales or the less-specific three-point scale (5, 8, 13) and were poor in two studies (9, 14). In one of these studies with lower results (ICC 0.61), the question was formulated as "hands below knee >30 minutes/day" using a five-point frequency scale measuring number of days per week or per month (9). The other study that had lower ICC (0.52–0.59) used a question similar in wording and a six-point duration scale (14). The question on bent trunk in the Leijon et al study (13) had a relatively high weighted kappa but a relatively lower Spearman's rank correlation coefficient (r_S 0.69). Using the dichotomized version of their response scale for this question, they also found a statistically significant difference between the proportion identified as exposed to trunk flexion on the first administration when compared with the proportion exposed on the second administration of the same question, the result suggesting poor reproducibility.

Twisted trunk posture was tested in one study and showed good reproducibility with a three-point scale (5). Combined trunk posture was tested in three studies, trunk twisting combined with lateral bending using a six-point duration scale (% worktime) showing moderate-to-good reproducibility (8) and bent and twisted body several times per hour (5-point scale) showing good and fair reproducibility in two studies [K_w 0.83, r_S 0.78 (13); ICC 0.74 (9)].

The reproducibility of questions about arm or shoulder postures varied somewhat. Questions about work with hands above shoulder level had very good reproducibility, with a duration scale expressed as the percentage of the workday [K_w 0.79 (13)] and a fair result with a frequency scale [ICC 0.75 (9)]. Balogh et al (5) asked about arm elevated or stretched forward with a three-point scale, the result being a K_w of 0.60. The worst results were found for people with a low education level [ICC 0.20 (14)] or using a visual analogue scale for duration [ICC 0.24 (15)]. Generally, reproducibility results using the formulation "raise your arms during each activity" (15) were considerably worse than those of questions specifying work with "hands above shoulder".

Hand or wrist posture (duration and frequency of wrist bending) was tested in only one study, using both a 10-cm visual analogue scale for duration showing poor reproducibility (ICC 0.62) and a seven-point frequency scale showing fair reproducibility (K_w 0.40) (15).

Repetitive movements. Four studies tested the reproducibility of questions concerning repetitive movements. Very good performance was observed in the studies by Leijon et al (13) and Balogh et al (5) with the formulation "repetitive movements several times/hour". Wiktorin et al (14), using a six-point duration scale measuring fraction of worktime and a formulation asking about "repetitive hand or finger movements several times/minute", and Torgén et al (9), using the formulation "repetitive hand or finger movements several times/minute >2 hours/day", with a five-point frequency scale, measuring days/week or month had poorer results (ICC 0.64–0.71).

Hand use. Questions on hand use (precision work, grasping, and hand or wrist exertion) were tested in five studies. The best result was K_w 0.68, using the less specific three-point scale, measuring the presence of precise movements (5). The results using five- or seven-point scales measuring the frequency of precision work >2 hours/day [ICC 0.36 (9)] and grasping duration, frequency, and intensity were relatively poor [ICC 0.59–0.69, K_w 0.36 (15)].

Material handling. Material handling was tested in five studies, and the reproducibility results were fair to excellent in four studies (5, 8, 9, 13). Reproducibility was better using a single question about material handling of >10 kg or using two items (handling 5–15 kg or >15 kg) (9, 13) than using more-detailed questions, with a wide range of weights and narrow intervals (14). The only study that showed poor agreement was that of Wiktorin et al (14), in which they asked about carrying, pushing, or pulling specific weights (1–5, 6–15, 16–45, >45 kg) using a six-point scale for duration. The agreement was particularly low in this study for workers with a low education level (ICC 0.30).

Physical effort. Questions about the level of physical effort at work showed very good reproducibility in three of four studies. The results were very good using a four or five-point intensity scale [K_w 0.85 (13), ICC 0.87 (14)]. Using the same phrasing and the same 15-point Borg scale, two studies found different reproducibility results [ICC 0.82 (14) and ICC 0.67 (9)]. It is possible that the higher ICC found in the study by Wiktorin et al (14) may have been due to lower variability in the distribution of exposure responses.

Vibration. Overall, questions on whole-body vibration showed good reproducibility [ICC 0.95 (9), K_w 0.75 (5), ICC 0.70 (14)]. Similarly, the results on questions concerning the use of handheld vibrating tools were good to excellent (5, 8, 9, 14).

Use of video display terminals. The question on duration of the use of video display terminals showed excellent reproducibility (ICC 0.95 and 0.90), in three studies, using different response scales, a five-point duration scale (14), a 10-cm visual analogue scale (9) and an open question [0-100% of worktime (9)]. Reducing responses to a question on "location of mouse" to a dichotomized scale (optimal-nonoptimal), Karlqvist et al (16) found a good result (r_S 0.81), but the questions on "proportion of the worktime using a mouse" and on "location of the keyboard" showed fair reproducibility. A question on the distance between the elbow height and keyboard was not reliably answered by workers (r_S 0.39).

Confounders and effect modifiers. Two studies took account of gender when analyzing test-retest reproducibility. No substantial differences were found between men and women (14, 15), although men tended to answer more consistently (14). The influence of age was studied in one study and was not found to influence the results (14). Education level was found to influence the reproducibility of questions, especially those concerning specific work postures (eg, head rotation, work above shoulder level), and manual materials handling (14). People with only a grade school education achieved a lower ICC than those with vocational or a senior high school education or college or university education (14).

The influence of the presence of musculoskeletal symptoms on the reproducibility of responses to physical work demands was analyzed in one study (14). The magnitude of the influence of low-back pain on the reproducibility of questions was generally small, although the participants with pain tended to give slightly less consistent answers.

Validity studies

All 10 of the studies of validity included in this review evaluated the convergent validity of single questions by comparing the worker's self-reports of exposure to measures obtained with a structured interview (13, 17), observation (18–22), or direct measurement (23, 24).

Table 3 describes the studies on the validity of work demand questions, and table 4 summarizes the main findings of each study.

General body postures. Overall, the agreement between the reference methods and self-reports was good for questions on sitting posture (13, 17, 20, 22), with the exception of one study by Burdorf & Laan (19), which found a statistically significant difference between the worker questionnaire results of the mean percentage of daily worktime sitting and that extrapolated from 60 observations every 20 seconds over two 10-minute periods by the OWAS (Ovako working-posture analysis system) observation technique. In the study by Wiktorin et al (17), workers tended to underestimate the proportion of the workday in a sitting posture. Using collapsed results into a dichotomized scale, sensitivity increased (from 0.44 to 0.96) when the cut-off point for exposure was defined as being higher than a quarter of the workday in the posture, instead of higher than three-quarters of the workday.

Table 3. Summary of the studies on the validity of questions measuring physical work demands. (OWAS = Ovako working-posture analysis system, ANOVA = analysis of variance, VDU = visual display units, PEO = postural ergonomic observation, χ^2 = chi square)

Study	Population	Physical work demands and questionnaire	Reference method	Statistics	Methodological issues
Andrews et al, 1998 (18)	201 workers (16 women, 185 men) from a Canadian automobile plant who were participants in a case-control study; the case participants reporting back pain and the random control participants came from different job titles: production operators (jobs consisted of several short cycle, repetitive assembly line tasks) and utility relief, maintenance and support personnel (more varied and longer cycle times)	9 self-administered questions answered at home, prior to on-site data collection: trunk posture (5 items), arm overhead (1 item), heavy lifting (1 item), squatting (1 item), trunk static posture (1 item) asking about the presence of exposure per task	Checklist filled out by 1 of 3 trained observers, all of them specialized in occupational biomechanics, physical ergonomics or kinesiology; the checklist was completed during the on-site visit; observations took 2–4 hours, depending on the variation of the task; the workers were observed during typical periods of work, with dynamic and static efforts and during periods of waiting; the frequency of tasks per workshift was determined by observer from the quantity of units or asking the area supervisor	Pearson correlation between methods on the estimates per workshift; χ^2 for nominal responses about trunk static posture; paired t-tests for the mean difference between the methods	Variability of distribution of exposure: not stated; confounders/effect modifiers: not described; validity of reference methods: not described, but extensive observer training including instruction and practice in laboratory and workplace, lack of comparability between exposures measured by questionnaire and observation; interval between questionnaire and reference method: not described; other: unclear whether the lack of concordance was due to the estimation of the number of times each task was performed per workshift or the exposure presence per task; other: response rate to individual questions ranged from 77% to 83%

(continued)

Table 3. Continued.

Study	Population	Physical work demands and questionnaire	Reference method	Statistics	Methodological issues
Burdorf & Laan, 1991 (19)	35 workers from a maintenance department of a Dutch steel factory, from 4 different job titles (pipe fitters, mechanical repair, constructional fitters and benchmen)	9 questions administered by workplace physician during an on-site medical examination: general body postures, trunk postures, lifting	OWAS; time-sampling approach: 2 periods of 10 minutes for each worker, consisting of 60 observations at 20-second intervals	For continuous variables: t-test on means; for categorical variables: ANOVA; multiple linear regression analysis using observation data as dependent variable and questionnaire responses as independent variable, (controlling for age): R^2 (proportion of variance explained) used to describe strength of correlation between the 2 methods	Variability of distribution of exposure: not described; confounders/effect modifiers: age, height, weight, duration of employment, task group, musculoskeletal complaints; validity of reference method: criteria for each question explicitly stated, very short observation period, interrater reliability of OWAS method not described; interval between questionnaire and reference method: observations took place 2 weeks or less after workers answered questionnaire; other: small sample size
Hansson et al, 2001 (24)	82 Swedish participants: 41 (24 women, 17 men) office workers from 2 municipal offices and 1 company, 45 years old or over, stratified by gender and musculoskeletal complaints; 41 participants selected from a sample of 218 female cleaners from 4 hospitals, stratified by age (<45 or ≥45) and musculoskeletal complaints	Postal questionnaire (from Malmö Shoulder-Neck Study) with 7 questions: posture of head, back and arms, repetitive movements of wrist and arms; response scale: 3-point scale (very little, somewhat and much)	Direct measurements of posture and movements by inclinometer and goniometers, continuously recorded during a full workday (4–8 hours); task diary filled out on same day as direct record and on 9 additional days; based on measurements during the specified worktasks and task diary (duration of each task), time-weighted values of exposure were calculated	Kappa; percentage of full agreement; Mann-Whitney test for differences in time-weighted exposure between groups; Jonkheere-Terpstra test for trend for differences in responses between participants with and without neck-shoulder complaints and between gender in office workers	Variability of distribution of exposure: many responses were concentrated at the lowest score; confounders/effect modifiers: occupation, neck-shoulder complaints, gender, age; validity of reference methods: goniometers and inclinometers validated in previous published studies, inclinometers measure absolute, not relative angles and measures may be influenced by the dynamics of motion; validity of the time-weighted job exposure not assessed; on the average, 60.8 tasks were performed during the 9-day diaries, and 30.5 tasks were measured on the day of measurements; these latter tasks represented 78% (range 12–100%) of the worktask time performed during the 9-day diaries; interval between questionnaire and reference method: not described
Karlqvist et al, 1996 (16)	100 volunteer Swedish workers (64 women, 36 women) from 2 research institutes, working on VDU for an average of 40% of the workday	3 questions: location of keyboard, mouse, height of elbow in relation to keyboard	Direct measurements, using a rule to measure the location of keyboard and mouse, based on a diagram with grid of 14–28 squares and to measure height of elbow in relation to keyboard	Weighted kappa; sensitivity; specificity	Variability of distribution of exposure: not stated among VDU users; confounders/effect modifiers: gender; validity of reference method: not described; interval between questionnaire and reference method: same day
Leijon et al, 2002 (13)	203 employees (102 women, 101 men) sampled from 80 worksites in 5 Swedish counties, selected through a 4-step process to ensure a variation in type, size and location of employers and worktasks and to ensure a similar distribution of gender and a similar distribution by age, level of education and type of work for each gender; part of Modern Work and Living Conditions for Women and Men (MOA) study, 203 participants completed the questionnaire and 202 participants underwent reference method	7 self-administered questions from the Stockholm Public Health Questionnaire concerning general physical activity, repetitive movements, general body postures, specific awkward postures, manual materials handling	45-minute structured personal interview by 4 trained ergonomists about worktasks and activities over the preceding 2 months; by consensus, the ergonomists categorized the physical demands of interest according to the questionnaire response alternatives using explicit criteria and based on the interview and observation and direct measurements performed after the interviews, nature of observation and direct measurements not described	Weighted kappa (K_w); percentage of full agreement; Spearman rank correlation coefficient (r_S); sensitivity and specificity of dichotomized regrouping of responses (exposed versus unexposed); ratio of the prevalence odds ratios for the musculoskeletal complaints from the questionnaire responses to the prevalence odds ratios for the musculoskeletal complaints from the reference method	Variability of distribution of exposure: skewed distribution [ie, frequency of highly exposed participants low in four questions (work with hands above shoulder level, bent trunk, manual handling of loads ≥10 kg, general physical activity)]; reference method rated 44% to 76% of the participants in the lowest exposure category (ie, unexposed) for these same four questions; confounders/effect modifiers: gender, type of work and presence of musculoskeletal complaints; validity of reference method: reference method validated for 5 of the 7 physical work demands evaluated; reference method relies on structured interview, observation and direct measurements; nature of observation and direct measurements not described; interval between questionnaire and reference method: not described

(continued)

Table 3. Continued.

Study	Population	Physical work demands and questionnaire	Reference method	Statistics	Methodological issues
Nordstrom et al, 1998 (21)	61 American (Wisconsin) participants (31 women, 30 men), 28 with carpal tunnel syndrome and 33 controls, randomly sampled from participants of a case-control study, who had only 1 job the 12 months prior to the diagnosis of carpal tunnel syndrome or the control interview	11 phone interviewer-administered questions: forearm and hand movements, lifting, postures of trunk, forearms, wrists and hands	Approximately 1-hour observation (range: 15–150 minutes, median = 55 minutes) by an ergonomist, recording worker exposure to 11 physical activities included in the questionnaire; observer was blinded to case-control status and exposure based on self-report	Kappa; full agreement; Spearman correlation coefficient	Variability of distribution of exposure: not stated; confounders/effect modifiers: gender, presence of carpal tunnel syndrome; validity of reference method: ergonomist extrapolated daily duration of exposures based on 1 hour of observation, ergonomist's observation methods and their validity not described; interval between questionnaire and reference method: not described
Pope et al, 1998 (20)	123 (76 women, 47 men) fulltime employees randomly selected from 6 workplaces in Manchester, United Kingdom	14 self-administered questions: manual handling postures, repetitive movements; response scale: all responses on 2-point scale, characterizing exposed and unexposed only at a dichotomous level	1-hour observation by 2 observers (an ergonomist and a physiotherapist); at 30-second intervals, physical demands were recorded by 1 of 2 observers	Kappa; full agreement; sensitivity and specificity	Variability of distribution of exposure: not stated; confounders/effect modifiers: not described; validity of reference methods: the interobserver reliability of the reference methods was assessed, a time-sampling approach of observation was used, and the questionnaire asked about the same time period as the observation; interval between questionnaire and reference method: questionnaire followed observations—interval not stated but description suggests same day; other: response rate: 140 persons observed, but the questionnaire response rate was 88%; no difference between the respondents and nonrespondents in terms of age, gender and physical demands
Viikari-Juntura et al, 1996 (22)	36 participants selected from a sample of male forest industry workers in Finland, had previously answered the questionnaire; 18 in tasks known to include awkward constrained body postures or lifting (maintenance men, repairmen, carpenters and truck drivers) and 18 without such loading (foremen and process control workers); both subgroups included 9 men who had reported severe low-back pain; for each worker with back pain a referent with no pain in the same occupation with a similar set of tasks was selected; 35 had had their current tasks more than 2 years	10 self-administered questions: whole-body posture, trunk, and neck posture, hand above shoulder level, manual handling, repetitive movements (body postures demonstrated with illustrations)	PEO method, a real-time observation method, performed by occupational physiotherapist for measuring postures of the hand, neck and trunk, repetitive movements of the hands, lifting, kneeling or squatting; observation time of each worker ranged from 2 to 90 minutes (median 18 minutes); task analysis for measuring sitting posture; direct measurement using pedometer for measuring walking distance (number of steps x step length)	Spearman correlation	Variability of distribution of exposure: trunk rotation >45 degrees and manual handling >15 kg could not be validated because exposure was too rare; confounders/effect modifiers: job type, presence of low-back or neck pain; validity of reference method: PEO method previously validated but not for neck rotation, the pedometer may overestimate walking distance; interval between questionnaire and reference method: 6–9 months; other: small sample size
Wiktorin, 1993 (23)	97 Swedish workers (58 women, 39 men): 72 randomly selected from working population of the Stockholm area, 12 furniture removers and 13 medical secretaries, employed in their current job for at least 3 months	17 questions: postures (9 illustrated questions), material handling (8 questions)	Direct measurements using a posimeter for measuring sitting position, a pedometer for measuring walking distance (number of steps x step length), and a trunk flexion analyzer, a type of inclinometer; PEO observation method used for measuring 13 questions; each participant studied during an ordinary workday; average recording time of 5 hours and 48 minutes for the pedometer measurements, 5 hours and 48 minutes for the posimeter measurements, 3 hours and 36 minutes for the trunk flexion analyzer and 26 minutes for the systematic PEO with portable computers by experienced ergonomists	Kappa; Spearman correlation	Variability of distribution of exposure: in 12 of 17 questions, 80% of the participants selected the 2 lowest frequency or duration choices on the 6-point response scales suggesting very limited distribution of exposure, in the analyses responses were collapsed to test the ability of participants to discriminate between exposed or unexposed (dichotomized response); confounders/effect modifiers: presence of musculoskeletal complaints, differential misclassification according to presence of musculoskeletal complaints assessed not only by different agreement for the group with complaints and that with no complaints, but also by calculating risk estimate for musculoskeletal complaints on the basis of exposure by questionnaire and exposure by reference methods (musculoskeletal complaints concerned the last 12 months and not the day when the recordings were performed); validity of reference method: criteria for each question explicitly stated, validation of the reference methods published, duration of reference measurements varied from 26 minutes to 6 hours, which may have led to a memory bias for some participants for some questions, the pedometer may overestimate walking distance, authors mention trunk rotation, head rotation and head bent as difficult to estimate by the ergonomists; interval between questionnaire and reference method: immediately after the end of each recording period the participants answered the corresponding question

(continued)

Table 3. Continued.

Study	Population	Physical work demands and questionnaire	Reference method	Statistics	Methodological issues
Wiktorin et al, 1999 (17)	Cases: 1053 (632 women, 421 men) who sought care for low-back, neck or shoulder disorders; controls: 1423 (813 women, 610 men) selected as a stratified (by gender and age) random sample from the Norrtalje, Sweden, study base; the participants recruited for this study worked at least 17 hours/week and had been in their current job for at least 2 months; most of the participants were from the service sector or white-collar positions	5 self-administered questions: sitting posture, whole-body vibration, hand positions (above shoulder level and below knee level), VDU work	30- to 45-minute structured personal interview, concerning physical loads of each work-task in a typical workday, including duration of general body postures, postures of specific body regions, vehicle use and VDU use, carried out by 7 trained occupational physiotherapists	Spearman correlation coefficient; Pearson correlation coefficient; differential misclassification according to presence of musculoskeletal complaints: comparison of interview odds ratio to questionnaire odds ratio for 5 questions	Variability of distribution of exposure: skewed distribution of all exposures with few participants highly exposed (except to sitting); confounders/effect modifiers: presence of musculoskeletal complaints, gender, age; validity of reference method: reference method relied on participant self-reports (interview), validated previously; interval between questionnaire and reference method: questionnaires answered immediately after interviews

Table 4. Results of the studies on the validity of questions measuring physical work demands. (r = correlation coefficient, r_S = Spearman correlation coefficient, PEO = portable ergonomic observation, OWAS = Ovako working-posture analysis system, VAS = visual analogue scale)

Type of demand	Exposure variables	Response scales	Reference method	Agreement statistics					
				Sensitivity	Specificity	Kappa	r_S	Agreement (%)	r
General body posture									
Sitting/standing									
Leijon et al, 2002 (13)	Sitting work posture in current job	Duration: 5-point, percentage of workday	45-minute structured personal interview, task by task, by ergonomists	·	·	0.81	0.83	47	·
Leijon et al, 2002 (13)	Sitting work posture in current job	Duration: 5-point, percentage of workday, collapsed into 2 categories	45-minute structured personal interview, task by task, by ergonomists	0.71	0.91	·	·	·	·
Wiktorin et al, 1999 (17)	Proportion of day spent in sitting posture	Duration: VAS fraction of workday (anchors: not at all, whole time)	45-minute structured personal interview, task by task, by occupational physiotherapists	·	·	·	·	·	0.82
Wiktorin et al, 1999 (17)	Proportion of day spent in sitting posture	Duration: VAS fraction of workday, stratified into 2 categories (≤¾ of time; >¾ of time)	45-minute structured personal interview, task by task, by occupational physiotherapists	0.44	0.98	·	·	·	·
Wiktorin et al, 1999 (17)	Proportion of the day spent in sitting posture	Duration: VAS fraction of workday, stratified into 2 categories (<¼; ≥¼ of time)	45-minute structured personal interview, task by task, by occupational physiotherapists	0.96	0.57	·	·	·	·
Viikari-Juntura et al, 1996 (22)	Sitting in work	Duration: 3-point, hours/workday	Task analysis and observations by occupational physiotherapist for estimating sitting hours	·	·	·	0.86	·	·
Wiktorin et al, 1993 (23)	Sitting	Duration: 6-point, fraction of specified time period	Direct measurement using posimeter: sitting duration by inclinometry	·	·	0.35	0.85	0.48	
Wiktorin et al, 1993 (23)	Sitting	Duration: 6-point fraction of specified time period, collapsed into 3 categories (0–37%; 38–62%; 63–100%)	Direct measurement using posimeter: sitting duration by inclinometry	·	·	0.52	·	68	·
Wiktorin et al, 1993 (23)	Sitting	Duration: 6-point, fraction of specified time period, collapsed into 2 categories (0–37%; 38–100%)	Direct measurement using posimeter: sitting duration by inclinometry	·	·	0.77	·	89	·
Pope et al, 1998 (20)	Seated in one position for ≥30 minutes (in specified hour)	Presence: 2-point (yes; no)	1-hour observation by 2 observers (time sampling approach)	10.0	0.93	·	·	·	·
Pope et al, 1998 (20)	Standing in one position for ≥30 minutes (in specified hour)	Presence: 2-point (yes; no)	1-hour observation by 2 observers (time sampling approach)	0.73	0.80	·	·	·	·
Burdorf & Laan, 1991 (19) [a]	Sitting on an average workday	Duration: proportion of daily worktime	Direct observation using OWAS	·	·	·	·	·	·
Burdorf & Laan, 1991 (19) [a]	Standing on an average workday	Duration: number of hours/workday	Direct observation using OWAS	·	·	·	·	·	·

(continued)

Table 4. Continued.

Type of demand	Exposure variables	Response scales	Reference method	Agreement statistics: Sensitivity	Specificity	Kappa	r_S	Agreement (%)	r
Walking									
Viikari-Juntura et al, 1996 (22)	Distance walked on ordinary workday	Distance (km): 3-point (<1 km; 1–5 km; >5 km)	Direct measurement using pedometer: number of steps × step length= walking distance						
All participants				·	·	·	0.65	·	·
Workers with severe low-back pain				·	·	·	0.53	·	·
Workers with no low-back pain				·	·	·	0.77	·	·
Wiktorin et al, 1993 (23)	Walking distance	Distance (km): 5-point (in specified time period)	Direct measurement using pedometer: number of steps × step length= walking distance	·	·	0.17	0.59	41	·
Wiktorin et al, 1993 (23)	Walking distance	Distance: 5-point (in specified time period), collapsed into 2 categories (<1 km;≥1 km)	Direct measurement using pedometer: number of steps × step length= walking distance	·	·	–0.02	·	37	·
Wiktorin et al, 1993 (23)	Walking distance	Distance: 5-point (in specified time period), collapsed into 2 categories (≤2 km; >2 km)	Direct measurement using pedometer: number of steps × step length= walking distance	·	·	0.41	·	76	·
Burdorf & Laan, 1991 (19) [a]	Walking on an average workday	Duration: number of hours/ workday	Direct observation using OWAS						
Kneeling or squatting									
Viikari-Juntura et al, 1996 (22)	Kneeling or squatting at work	Duration: 4-point, hours/workday	Direct observation of kneeling or squatting: PEO method						
All participants				·	·	·	0.42	·	·
Workers with severe low-back pain				·	·	·	0.34	·	·
Workers with no low-back pain				·	·	·	0.50	·	·
Wiktorin et al, 1993 (23)	Kneeling or squatting	Duration: 6-point, fraction of specified time period, collapsed into 2 categories (not at all, >not at all)	Direct observation of kneeling or squatting: PEO method	·	·	0.76	·	88	·
Burdorf & Laan, 1991 (19) [b]	Kneeling or squatting on an average workday	Duration: number of hours/ workday	Direct observation using OWAS						
Andrews et al, 1998 (18)	Squat (hips below knees)	Frequency: times/shift	Checklist filled out by trained observers	·	·	·	·	·	0.16
Postures involving parts of body									
Neck posture									
Hansson et al, 2001 (24)	Work with head bent forward a little	Intensity: 3-point [c]	Direct measurement using inclinometer, measuring head angles with gravity, 90th percentile (degrees)						
All participants				·	·	0.27	·	61	·
Office workers						0.34	·	62	·
Cleaners				·	·	0.24	·	62	·
Hansson et al, 2001 (24)	Work with head bent forward a lot	Intensity: 3-point [c]	Direct inclinometer measure of head angles with gravity, 90th percentile (degrees)						
All participants				·	·	0.27		57	·
Office workers				·	·	–0.07	·	72	·
Cleaners				·	·	0.07	·	42	·
Hansson et al, 2001 (24)	Work with head bent backward	Intensity: 3-point [c]	Direct inclinometer measure of head angles with gravity, 90th percentile (degrees)						
All participants				·	·	0.23	·	67	·
Office workers				·	·	0.18	·	84	·
Cleaners				·	·	0.18	·	53	·
Viikari-Juntura et al, 1996 (22)	Neck bent forward	Duration: 3-point, hours/workday	Observation by physiotherapist using PEO method						
All participants				·	·	·	0.15	·	·
Workers with severe low-back pain				·	·	·	–0.22	·	·
Workers with no low-back pain				·	·	·	0.47	·	·

(continued)

Table 4. Continued.

Type of demand	Exposure variables	Response scales	Reference method	Agreement statistics					
				Sensitivity	Specificity	Kappa	r_S	Agreement (%)	r
Viikari-Juntura et al, 1996 (22)	Rotated neck	Duration 3-point: hours/workday	Direct observation of neck rotation >45 degrees: PEO method						
All participants				·	·	·	0.55	·	·
Workers with severe low-back pain				·	·	·	0.58	·	·
Workers with no low-back pain				·	·	·	0.51	·	·
Wiktorin et al, 1993 (23)	Head rotation	Duration 6-point: fraction of specified time period, collapsed into 2 categories (not at all; >not at all)	Direct observation of neck rotation >45 degrees: PEO method	·	·	0.17	·	58	·
Wiktorin et al, 1993 (23)	Head bent forward	Duration 6-point: fraction of a specified time period	Direct observation of neck flexion >20 degrees: PEO method	·	·	0.06	0.41	26	·
Wiktorin et al, 1993 (23)	Head bent forward	Duration: 6-point fraction of specified time period, collapsed into 3 categories (<¼; ≥¼; <¾; ≥¾)	Direct observation of neck flexion >20 degrees: PEO method	·	·	0.31	·	57	·
Wiktorin et al, 1993 (23)	Head bent forward	Duration: 6-point, fraction of specified time period, collapsed into 2 categories (<¼; ≥¼)	Direct observation of neck fle xion >20 degrees: PEO method	·	·	0.41	·	76	·
Trunk posture									
Leijon et al, 2002 (13)	Trunk bent in current job	Duration: 4-point, percentage of workday	45 minutes structured personal interview, task by task, by ergonomists	·	·	0.48	0.41	69	·
Leijon et al, 2002 (13)	Trunk bent in current job	Duration: 4-point, percentage of workday, collapsed to 2 categories	45-minute structured personal interview, task by task, by ergonomists	0.47	0.87	·	·	·	·
Hansson et al, 2001 (24)	Back bent forward a lot	Intensity: 3-point [c]	Direct inclinometer measure of upper back with gravity, 90th percentile (degrees)						
All participants				·	·	0.26	·	55	·
Office workers				·	·	–0.06	·	89	·
Cleaners				·	·	–0.12	·	35	·
Viikari-Juntura et al, 1996 (22)	Bending the trunk forward (standing or kneeling)	Duration: 4-point, hours/day	Direct observation of trunk flexion: PEO method						
All participants				·	·	·	0.42	·	·
Workers with severe low-back pain				·	·	·	0.62	·	·
Workers with no low-back pain				·	·	·	0.30	·	·
Wiktorin et al, 1993 (23)	Trunk bent forward >60 degrees	Duration: 6-point, fraction of specified time period, collapsed into 2 categories (not at all, >not at all)	Direct measurement using inclinometer: duration >54 degrees	·	·	0.43	·	88	·
Wiktorin et al, 1999 (17)	Hands below knee level >30 minutes/day	Frequency: 5-point, days/week or month	45-minute structured personal interview, task by task, by occupational physiotherapists	·	·	·	0.66	·	·
Wiktorin et al, 1999 (17)	Hands below knee level >30 minutes/day	Frequency: 5-point, days/week or month, collapsed into 2 points (never; >never)	45-minute structured personal interview, task by task, by occupational physiotherapists	0.79	0.73	·	·	·	·
Wiktorin et al, 1999 (17)	Hands below knee level >30 minutes/day	Frequency: 5-point days/week or month, collapsed into 2 points (<every day; every day)	45-minute structured personal interview, task by task, by occupational physiotherapists	0.38	0.94	·	·	·	·
Nordstrom et al, 1998 (21)	Repeated bending or twisting at the waist	Presence: 2-point (yes; no)	1-hour observation by a blinded ergonomist concerning case–control status						
Carpal tunnel syndrome				·	·	0.79	·	89	·
Control				·	·	0.28	·	64	·
Nordstrom et al, 1998 (21)	Repeated bending or twisting at the waist	Duration: 4-point, minutes or hour/day	1-hour observation by a blinded ergonomist concerning case–control status						
Carpal tunnel syndrome				·	·	·	0.67	·	·
Control				·	·	·	0.38	·	·
Wiktorin et al, 1993 (23)	Trunk bent forward 20–60 degrees	Duration: 6-point fraction of a specified time period	Direct measurement using inclinometer	·	·	0.12	0.10	31	·
Wiktorin et al, 1993 (23)	Trunk bent forward 20–60 degrees	Duration: 6-point, fraction of a specified time period, collapsed into 2 categories (0–17%; 18–100%)	Direct measurement using inclinometer	·	·	0.16	·	66	·

(continued)

Table 4. Continued.

Type of demand	Exposure variables	Response scales	Reference method	Agreement statistics					
				Sen-siti-vity	Spe-cifi city	Kappa	r_S	Agree-ment (%)	r
Andrews et al, 1998 (18)	Trunk twists >20 degrees	Frequency: times/shift	Checklist filled out by trained observers	·	·	·	·	·	0.17
Andrews et al, 1998 (18)	Trunk lateral bends >20 degrees	Frequency: times/shift	Checklist filled out by trained observers	·	·	·	·	·	0.40
Andrews et al, 1998 (18)	Moderate trunk flexion >15 degrees	Frequency: times/shift	Checklist filled out by trained observers	·	·	·	·	·	0.51
Andrews et al, 1998 (18)	Severe trunk flexion >45 degrees	Frequency: times/shift	Checklist filled out by trained observers	·	·	·	·	·	0.47
Andrews et al, 1998 (18)	Trunk extension >0 degrees	Frequency: times/shift	Checklist filled out by trained observers	·	·	·	·	·	0.01
Burdorf & Laan, 1991 (19) [a]	Bent or twisted trunk on an average day	Duration:percentage of daily worktime	Direct observation using OWAS						
Arm posture									
Leijon et al, 2002 (13)	Hands above shoulder in current job	Duration: 4-point, percentage of workday	45-minute structured personal interview, task by task, by ergonomists						
All participants				·	·	0.48	0.53	69	·
With musculo-skeletal com-plaints [d]				·	·	0.34	·	·	·
Without muscu-loskeletal complaints [d]				·	·	0.53	·	·	·
Leijon et al, 2002 (13)	Hands above shoulder in current job	Duration: 4-point, percentage of workday, collapsed to 2 categories	45-minute structured personal interview, task by task, by ergonomists	0.79	0.76	·	·	·	·
Hansson et al, 2001 (24)	Arms elevated or stretched forward	Intensity: 3-point [c]	Direct inclinometer measure of angles of upper arms with gravity, 90th percentile (degrees)						
All participants				·	·	0.27	·	53	·
Office workers				·	·	–0.08	·	59	·
Cleaners				·	·	0.26	·	59	·
Wiktorin et al, 1999 (17)	Hands above shoulder >30 minutes/day	Frequency: 5-point, days/week or month	45-minute structured personal interview, task by task, by occupational physiotherapists				0.63		
Wiktorin et al, 1999 (17)	Hands above shoulder >30 minutes/day	Frequency: 5-point, days/week or month, collapsed to 2 categories (never; >never)	45-minute structured personal interview, task by task, by occu-pational physiotherapists	0.86	0.74	·	·	·	·
Wiktorin et al, 1999 (17)	Hands above shoulder >30 minutes/day	Frequency: 5-point, days/week or month, collapsed into 2 categories (<every day; every day)	45-minute structured personal interview, task by task, by occu-pational physiotherapists	0.43	0.95	·	·	·	·
Pope et al, 1998 (20)	Work at or above shoulder level	Presence: 2-point (yes; no)	1-hour observation by 2 observers (time sampling approach)	0.71	0.73	·	·	·	·
Viikari-Juntura et al, 1996 (22)	Hands above shoulder	Duration: 4-point, hours/workday	Direct observation of hand above shoulder level: PEO method						
All participants				·	·	·	0.55	·	·
Workers with severe low-back pain				·	·	·	0.51	·	·
Workers with no low-back pain				·	·	·	0.71	·	·
Wiktorin et al, 1993 (23)	Hands above shoulder	Duration: 6-point, percentage of specified time period (not at all, >not at all)	Direct observation of hand above shoulder level: PEO method	·	·	0.17	·	57	·
Andrews et al, 1998 (18)	Arms overhead	Frequency: times/shift	Checklist filled out by trained observers	·	·	·	·	·	0.61
Forearm, hand, wrist posture									
Nordstrom et al, 1998 (21)	Bend or twist hands or wrists	Presence: 2-point, (yes; no)	1-hour observation by a blinded ergonomist concerning case–control status						
Carpal tunnel syndrome				·	·	0.26	·	64	·
Control				·	·	0.09	·	56	·

(continued)

Table 4. Continued.

Type of demand	Exposure variables	Response scales	Reference method	Agreement statistics					
				Sen-siti-vity	Spe-cifi city	Kappa	r_S	Agree-ment (%)	r
Nordstrom et al, 1998 (21)	Bend or twist hands or wrists	Duration: 4-point, minutes or hour/day	1-hour observation by a blinded ergonomist concerning case–control status						
Carpal tunnel syndrome				·	·		0.33	·	·
Control				·	·		0.01	·	·
Carpal tunnel syndrome				·	·	0.45		71	·
Control				·	·	–0.02		45	·
Nordstrom et al, 1998 (21)	Twist forearm	Duration: 4-point, minutes or hour/day	1-hour observation by a blinded ergonomist concerning case–control status						
Carpal tunnel syndrome				·	·		0.35	·	·
Control				·	·		0.05	·	·
Other postures									
Leijon et al, 2002 (13)	Bent/twisted body several times/hour	Frequency: 5-point, days/week or month	45-minute structured personal interview, task by task, by ergonomists	·	·	0.38	0.39	45	·
Leijon et al, 2002 (13)	Bent/twisted body several times/hour	Frequency: 5-point, days/week or month, collapsed into 2 categories	45-minute structured personal interview, task by task, by ergonomists	0.81	0.55	·	·	·	·
Andrews et al, 1998 (18)	Static posture	Presence: 2-point, (yes; no)	Checklist filled out by trained observers	·	·	·	·	·	0.11
Repetitive movements									
Leijon et al, 2002 (13)	Repetitive movements several times/hour	Duration: 5-point, percentage of workday	45-minute structured personal interview, task by task, by ergonomists						
All participants				·	·	0.39	0.40	32	·
With musculoskeletal complaints [d]				·	·	0.12	·	·	·
Without musculoskeletal complaints [d]				·	·	0.45	·	·	·
Leijon et al, 2002 (13)	Repetitive movements several times/hour	Duration: 5-point, percentage of workday, collapsed into 2 categories	45-minite structured personal interview, task by task, by ergonomists	0.51	0.74	·	·	·	·
Hansson et al, 2001 (24)	Same arm movements many times per minute	Intensity: 3-point [c]	Direct measurement using inclinometer, measuring upper-arn angular velocity (degrees/second)						
All participants				·	·	0.39	·	61	·
Office workers				·	·	–0.04	·	63	·
Cleaners				·	·	0.07	·	59	·
Hansson et al, 2001 (24)	Same wrist movements many times per minute	Intensity: 3-point [c]	Direct measurement using goniometer, measuring wrist angular velocity (degrees/second)						
All participants				·	·	0.18	·	54	·
Office workers				·	·	0.07	·	39	·
Cleaners				·	·	–0.12	·	59	·
Hansson et al, 2001 (24)	Same wrist movements many times per minute	Intensity: 3-point [c]	Direct measurement with goniometer, measuring wrist angular acceleration (degrees/second squared)						
All participants				·	·	0.16	·	53	·
Office workers				·	·	0.00	·	34	·
Cleaners				·	·	–0.12	·	59	·
Hansson et al, 2001 (24)	Same wrist movements many times per minute	Intensity: 3-point [c]	Direct measurement using goniometer, measuring wrist rest, velocity <1 degree/second (% time)						
All participants				·	·	0.16	·	53	·
Office workers				·	·	0.11	·	41	·
Cleaners				·	·	–0.12	·	59	·
Hansson et al, 2001 (24)	Same wrist movements many times per minute	Intensity: 3-point [c]	Direct measurement using goniometer, measuring wrist repetitiveness (Hz)						
All participants				·	·	0.23	·	56	·
Office workers				·	·	0.18	·	46	·
Cleaners				·	·	–0.19	·	56	·

(continued)

Table 4. Continued.

Type of demand	Exposure variables	Response scales	Reference method	Agreement statistics					
				Sensitivity	Specificity	Kappa	r_S	Agreement (%)	r
Pope et al, 1998 (20)	Repetitive wrist movements ≥10 minutes/specified hour	Presence: 2-point (yes; no)	1-hour observation by 2 observers (time sampling approach)	0.81	0.66	·	·	·	·
Pope et al, 1998 (20)	Repetitive arm movements ≥10 minutes/ specified hour	Presence: 2-point (yes; no)	1-hour observation by 2 observers (time sampling approach)	0.81	0.56	·	·	·	·
Viikari-Juntura et al, 1996 (22)	Repetitive wrist or finger movements (eg, keyboard work, screwing screws)	Duration: 4-point hours/workday	Direct observation of repetitive movements: PEO method						
All participants				·	·	·	0.26	·	·
Workers with severe low-back pain				·	·	·	0.18	·	·
Workers with no low-back pain				·	·	·	0.37	·	·
Hand or finger use									
Nordstrom et al, 1998 (21)	Pinch grip	Presence: 2-point (yes; no)	1-hour observation by an ergonomist blinded to case–control status						
Carpal tunnel syndrome				·	·	0.00	·	43	·
Control				·	·	–0.06	·	47	·
Nordstrom et al, 1998 (21)	Pinch grip	Duration: 4-point, minutes or hour/day	1-hour observation by an ergonomist blinded to case–control status						
Carpal tunnel syndrome				·	·	·	0.31	·	·
Control				·	·	·	0.24	·	·
Nordstrom et al, 1998 (21)	Press with finger	Presence: 2-point (yes; no)	1-hour observation by an ergonomist blinded to case–control status						
Carpal tunnel syndrome				·	·	0.00	·	0.50	·
Control				·	·	0.11	·	0.52	·
Nordstrom et al, 1998 (21)	Press with finger	Duration: 4-point, minutes or hour/day	1 hour observation by an ergonomist blinded to case–control status						
Carpal tunnel syndrome				·	·	·	0.16	·	·
Control				·	·	·	0.08	·	·
Manual materials handling									
Leijon et al, 2002 (13)	Handling ≥10 kg in current job	Frequency: 4-point, times/day	45-minute structured personal interview, task by task, by ergonomists						
All participants				·	·	0.54	0.54	60	·
Men				·	·	0.59	·	·	·
Women				·	·	0.41	·	·	·
Group working with "things"				·	·	0.60	·	·	·
Group working with "people"				·	·	0.44	·	·	·
Group working with "data"				·	·	0.33	·	·	·
Leijon et al, 2002 (13)	Handling ≥10 kg in current job	Frequency: 4-point, times/day, collapsed into 2 categories	45-minute structured personal interview, task by task, by ergonomists	0.67	0.88	·	·	·	·
Pope et al, 1998 (20)	Lift weights with one hand	Presence: 2-point (yes; no)	1-hour observation by 2 observers(time sampling approach)	0.63	0.78	·	·	·	·
Pope et al, 1998 (20)	Lift weights with both hands	Presence: 2-point (yes; no)	1-hour observation by 2 observers (time sampling approach)	0.87	0.48	·	·	·	·
Pope et al, 1998 (20)	Carrying weights with both hands	Presence: 2-point (yes; no)	1-hour observation by 2 observers (time sampling approach)	0.74	0.65	·	·	·	·
Pope et al, 1998 (20)	Carrying weights with one hand	Presence: 2-point (yes; no)	1-hour observation by 2 observers (time sampling approach)	0.43	0.73	·	·	·	·

(continued)

Table 4. Continued.

Type of demand	Exposure variables	Response scales	Reference method	Agreement statistics					
				Sensitivity	Specificity	Kappa	r_S	Agreement (%)	r
Pope et al, 1998 (20)	Lifting or carrying weights with both hands	Presence: 2-point (yes; no)	1-hour observation by 2 observers (time sampling approach)	0.81	10.0	·	·	·	·
Pope et al, 1998 (20)	Lifting or carrying weights with one hand	Presence: 2-point (yes; no)	1-hour observation by 2 observers (time sampling approach)	0.70	0.80	·	·	·	·
Pope et al, 1998 (20)	Carrying weights on one shoulder	Presence: 2-point (yes; no)	1-hour observation by 2 observers (time sampling approach)	0.75	0.97	·	·	·	·
Pope et al, 1998 (20)	Lifting weights above shoulder level	Presence: 2-point (yes; no)	1-hour observation by 2 observers (time sampling approach)	0.40	0.92	·	·	·	·
Pope et al, 1998 (20)	Pushing weights	Presence: 2-point (yes; no)	1-hour observation by 2 observers (time sampling approach)	0.74	0.76	·	·	·	·
Pope et al, 1998 (20)	Pulling weights	Presence: 2-point (yes; no)	1-hour observation by 2 observers (time sampling approach)	0.60	0.76	·	·	·	·
Nordstrom et al, 1998 (21)	Lifting, carrying, pushing or pulling objects >2 pounds	Presence: 2-point (yes;no)	1-hour observation by a blinded ergonomist concerning case-control status						
Carpal tunnel syndrome				·	·	0.41	·	75	·
Control				·	·	0.35	·	70	·
Nordstrom et al, 1998 (21)	Lifting, carrying, pushing or pulling objects >2 pounds	Duration: 4-point, minutes or hour/day	1-hour observation by a blinded ergonomist concerning case–control status						
Carpal tunnel syndrome				·	·	·	0.58	·	·
Control				·	·	·	0.41	·	·
Viikari-Juntura et al, 1996 (22)	Lifting, carrying, transferring 6–15 kg	Frequency: times/workday	Direct observation of lifting 6–15 kg: PEO method						
All participants				·	·	·	0.49	·	·
Workers with severe low-back pain				·	·	·	0.34	·	·
Workers with no low-back pain				·	·	·	0.55	·	·
Wiktorin et al, 1993 (23)	Carrying, pushing or pulling, using force corresponding to 1–5 kg	Duration: 6-point, fraction of specified time period, collapsed into 2 categories (not at all; >not at all)	Direct observation of manual handling 1–5 kg: PEO method	·	·	0.26	·	63	·
Wiktorin et al, 1993 (23)	Carrying, pushing or pulling, using force corresponding to 6–15 kg	Duration: 6-point, fraction of specified time period, collapsed into 2 categories (not at all; >not at all)	Direct observation of manual handling 6–15 kg: PEO method	·	·	0.50	·	79	·
Wiktorin et al, 1993 (23)	Carrying, pushing or pulling, using force corresponding to 16–45 kg	Duration: 6-point, fraction of specified time period, collapsed into 2 categories (not at all; >not at all)	Direct observation of manual handling 16–45 kg: PEO method	·	·	0.64	·	90	·
Wiktorin et al, 1993 (23)	Carrying, pushing or pulling, using force corresponding to >45 kg	Duration: 6-point, fraction of specified time period, collapsed into 2 categories (not at all; >not at all)	Direct observation of manual handling >45 kg: PEO method	·	·	–	·	90	·
Wiktorin et al, 1993 (23)	Lifting weight 1–5 kg	Frequency: 4-point times/hour	Direct observation of lift 1–5 kg: PEO method	·	·	0.12	0.63	35	·
Wiktorin et al, 1993 (23)	Lifting 1–5 kg	Frequency: 4-point, times/hour, collapsed into 2 categories (<1/hour; ≥1/hour)	Direct observation of lift 1–5 kg: PEO method	·	·	0.32	·	·	·
Wiktorin et al, 1993 (23)	Lifting 1–5 kg	Frequency: 4-point, times/hour, collapsed into 2 categories (≤30/hour; >30/hour)	Direct observation of lift 1–5 kg: PEO method	·	·	0.38	·	88	·
Wiktorin et al, 1993 (23)	Lifting 6–15 kg	Frequency: 4-point, times/hour, collapsed into 2 categories (<1/hour; ≥1/hour)	Direct observation of lift 6–15 kg:PEO method	·	·	0.66	·	86	·
Wiktorin et al, 1993 (23)	Lifting 16–45 kg	Frequency: 4-point, times/hour, collapsed into 2 categories (<1/hour; ≥1/hour)	Direct observation of lift 16–45 kg: PEO method	·	·	0.65	·	93	·
Wiktorin et al, 1993 (23)	Lifting >45 kg	Frequency: 4-point, times/hour, collapsed into 2 categories (<1/hour; ≥1/hour)	Direct observation of lift >45 kg: PEO method	·	·	–	·	90	·
Andrews et al, 1998 (18)	Heavy lifts	Frequency: times/shift	Checklist filled out by trained observers	·	·	·	·	·	0.06
Burdorf & Laan, 1991 (19) [a]	Lifting or carrying loads	Duration:percentage of daily worktime	Direct observation using OWAS						

(continued)

Table 4. Continued.

Type of demand	Exposure variables	Response scales	Reference method	Agreement statistics					
				Sen-siti-vity	Spe-cifi city	Kappa	r_S	Agree-ment (%)	r
Level of physical effort at work									
Leijon et al, 2002 (13)	General physical activity during the last year	Intensity: 4-point [e]	45-minute structured personal interview, task by task, by ergonomists	·	·	0.66	0.71	50	·
Leijon et al, 2002 (13)	General physical activity during the last year	Intensity: 4-point, collapsed into 2 categories	45-minute structured personal interview, task by task, by ergonomists	0.88	0.80	·	·	·	·
Vibration									
Whole-body vibration									
Wiktorin et al, 1999 (17)	Working on vibrating floor or seat (whole-body vibration)	Duration: VAS fraction of workday (anchors: not at all; whole time)	45-minute structured personal interview, task by task, by occupational physiotherapists	·	·	·	·	·	0.80
Wiktorin et al, 1999 (17)	Working on vibrating floor or seat (whole-body vibration)	Duration: VAS fraction of workday, stratified into 2 categories (≤¾ of time;>¾ of time)	45-minute structured personal interview, task by task, by occupational physiotherapists	0.52	0.99	·	·	·	·
Wiktorin et al, 1999 (17)	Working on vibrating floor or seat (whole-body vibration)	Duration: VAS fraction of workday, stratified into 2 categories (< half time; ≥half time)	45-minute structured personal interview, task by task, by occupational physiotherapists	0.93	0.96	·	·	·	·
Wiktorin et al, 1999 (17)	Working on vibrating floor or seat (whole-body vibration)	Duration: VAS fraction of workday, stratified into 2 categories (<¼; ≥¼ of time)	45-minute structured personal interview, task by task, by occupational physiotherapists	0.97	0.89	·	·	·	·
Handheld vibrating tools									
Nordstrom et al, 1998 (21)	Work with handheld or hand-operated power tools	Presence: 2-point (yes; no)	1-hour observation by a blinded ergonomist concerning case–control status						
Carpal tunnel syndrome				·	·	0.28	·	68	·
Control				·	·	0.02	·	67	·
Nordstrom et al, 1998 (21)	Work with handheld or hand -operated power tools	Duration: 4-point, minutes or hour/day	1-hour observation by a blinded ergonomist concerning case–control status						
Carpal tunnel syndrome				·	·	·	0.45	·	·
Control				·	·	·	0.03	·	·
Visual display terminal use									
Wiktorin et al, 1999 (17)	Working with VDU	Duration: VAS fraction of workday (anchors: not at all; whole time)	45-minute structured personal interview, task by task, by occupational physiotherapists	·	·	·	·	·	0.87
Wiktorin et al, 1999 (17)	Working at VDU	Duration: VAS fraction of workday, stratified into 2 categories (≤¾ of time;>¾ of time)	45-minute structured personal interview, task by task, by occupational physiotherapists	0.46	0.99	·	·	·	·
Wiktorin et al, 1999 (17)	Working with VDU	Duration: VAS fraction of workday, stratified into 2 categories (<half time; ≥half time)	45-minute structured personal interview, task by task, by occupational physiotherapists	0.92	0.94	·	·	·	·
Wiktorin et al, 1999 (17)	Working with VDU	Duration: VAS fraction of workday, stratified into 2 categories (<¼; ≥¼ of time)	45-minute structured personal interview, task by task, by occupational physiotherapists	10.0	0.86	·	·	·	·
Karlqvist et al, 1996 (16)	Location of the keyboard (based on a diagram with a grid of 14–28 squares)	Dichotomized responses (optimal; not optimal)	Direct measurement	0.57	0.98	0.59	·	·	·
Karlqvist et al, 1996 (16)	Location of the mouse (based on a diagram with a grid of 14–28 squares)	Dichotomized responses (optimal; not optimal)	Direct measurement	0.91	0.87	0.68	·	·	·
Karlqvist et al, 1996 (16)	Distance between elbows and keyboard	Dichotomized responses (optimal; optimal)	Direct measurement	0.19	0.93	0.14	·	·	·

[a] Statistically significant difference found between self-report question and the OWAS reference method in the Burdorf & Lann study, 1991 (19).
[b] No statistically significant difference found between self-report question and the OWAS reference method in the Burdorf & Lann study, 1991 (19).
[c] Intensity 3-point scale: very little or not at all, somewhat, and much for all of the results by Hansson et al, 2001 (24).
[d] Ratio of the prevalence odds ratios, from the questionnaire and from the interviews did not differ from unity.
[e] 1. sedentary, light work; 2. light, somewhat mobile work; 3. mobile, fairly heavy work; 4. heavy work.

The validity for standing posture was tested in two studies, and acceptable concordance between participant reports and the reference method was found in one study (20). In the other study, Burdorf & Laan (19) found a statistically significant difference between the worker questionnaire results for the mean percentage of daily worktime standing and that extrapolated from OWAS observations over two 10-minute periods. These results may reflect the limits of the very short observation time of the reference method used in this study. Alternatively, the low results may reflect the difficulty workers have reporting the duration of the workday in a standing posture, as this was the only study that used questions with a duration response scale.

The validity of questions on distance walked at work compared with that of direct measurements using a pedometer or observation was poor when a five-point distance scale was used, and it remained poor when this distance scale was collapsed into two categories of less than or greater than 1 km. However, it was possible for workers to indicate that they walked less than or more than 2 km in a specified time period (kappa 0.41, 76% agreement) when the scale was collapsed into these two categories (23). Viikari-Juntura et al (22) had good correlations (r_S 0.65) using a three-point distance scale distinguishing between <1 km, 1–5 km, and >5 km. Burdorf & Laan (19) had poor results when workers were asked how many hours per day they walk at work.

Questions on kneeling or squatting at work were tested in four studies, and agreement with observations was very poor when workers were asked how many hours per day or how many times per day they squat or kneel at work (18, 19). A four-point duration scale had moderate agreement with the observations [r_S 0.42 (22)]. However, when a six-point duration scale was collapsed into a dichotomous scale (not at all, >not at all), agreement was very high [K_w 0.76 (23)].

Postures of specific body regions

The validity of questions on neck posture was tested in three studies (22, 23, 24), and all of them found poor agreement between self-report and reference methods. There were two exceptions, however, one for a question about "head bent forward", using a six-point duration scale, collapsed into two categories [K_w 0.41 (23)], and the other for a question on "rotated neck" using a three-point duration scale [r_S 0.55 (22)].

The validity of questions on trunk posture was tested in eight different studies. Agreement between self-reports of the duration or frequency of trunk flexion was moderate when compared with that for structured interviews (13, 17), but generally poor when compared with that of direct measures with an inclinometer (23, 24). Better agreement was observed by Wiktorin et al (23) when they collapsed and dichotomized their response scales for questions on trunk bent >60 degrees (kappa 0.43, 88% agreement) and hands below knee level (sensitivity 0.79, specificity 0.73). Comparisons between self-reports and observations generally gave moderate or better results for questions on trunk flexion (18, 21, 22), with the exceptions of Burdorf & Laan (19), who observed very poor correlations between short OWAS observations and the self-report of the number of hours per workday spent with the trunk bent or twisted. As well, poor correlations between self-reports and trained observers' responses on a checklist for questions on trunk rotation and trunk extension were found in the study by Andrews et al (18).

The validity of arm posture (eg, "hands above shoulders" or "arms overhead") was tested in seven studies, and five showed moderate or better agreement when compared with that of observations or structured interview results. One study using relatively short observation periods had poor agreement (23), as did the study by Hansson et al (24), which compared a question about "arms elevated or stretched forward" on a three-point scale to inclinometer measures.

Questions on hand or wrist postures were tested in one study (21), which showed fair agreement, with the exception of a single item on twisting of the forearm. This question had moderate agreement for persons with carpal tunnel syndrome (kappa 0.45, 71% agreement) when compared with the results of observation. The controls (without carpal tunnel syndrome) showed very low agreement.

Repetitive movements

The validity of questions on repetitive movements was evaluated in four studies, and it was found to be poor to fair.

Hand use

Questions on hand or finger use were tested in one study with fair-to-poor agreement between self-reports and the reference method (21).

Material handling

Questions on material handling were tested in seven studies, and, overall, the results suggested that there was moderate agreement between observations and responses to questions about frequency or duration of specific loads handled (13, 20–23). Most of these studies used dichotomous response scales (20, 21) or presented results of four- to six-point response scales collapsed into dichotomous scales (23). Very good results (K_w 0.54) were also found in a comparison of a question on "handling ≥10 kg in present job", on a four-point frequency scale measuring times per day to structured interview responses (13). Studies using more nonspecific question formulations and continuous response scales had much poorer results than observations (ie, a question about the number of times per shift of "heavy lifts") (18) and one asking the number of hours per workday "lifting or carrying loads" (19). The results of Wiktorin et al (23) and Leijon et al (13) suggest that self-reports are more accurate for questions about the duration or frequency of manipulating higher weights (eg, 6–15 kg, >10 kg, 16–45 kg) than for low weights (≤5 kg).

Physical effort

The level of physical effort at work showed good validity in a study using a four-point intensity scale for "general physical activity during the last year" in comparison with a structured interview [K_w 0.66 (13)].

Vibration

A question on whole-body vibration showed good validity, on a visual analogue scale for duration when compared with that of a structured interview (r 0.80). When the scale was collapsed into two categories, very high sensitivity (0.93–0.97) and specificity (0.89–0.96) were found when work on a vibrating floor or seat for more than or less than half the workday was compared and also for more than or less than one-fourth of the workday (17). Questions on work with handheld vibrating tools was tested in one study that showed fair agreement (kappa 0.28) with observations when a dichotomous scale indicating presence or absence of exposure was used and moderate agreement (r_S 0.45) when a four-point duration scale measuring number of minutes or hours per day was used (21).

Use of visual display terminals

A self-administered question on the use of visual display terminals by Wiktorin et al (17) showed good validity (r 0.87) when compared with a structured interview by an ergonomist. The sensitivity and specificity were excellent when exposure was defined as higher than one-fourth or one-half of a workday, but sensitivity was considerably lower when exposure was defined as higher than three-fourths of a workday. Karlqvist et al (16) found very acceptable kappa testing questions on the location of the keyboard (kappa 0.59) and the location of the mouse (kappa 0.68) when compared with direct measurements. The question on "distance between elbow height and keyboard" had poor agreement (kappa 0.14).

Confounders and effect modifiers

No consistent difference according to gender was found in the validity studies. When the kappa differed between the men and women, the confidence intervals were wide and overlapping or the variability of the distribution of responses differed (13). Among office workers, Hansson et al (24) observed gender differences in self-assessed exposure to arm elevation, head bent forward a little, and repetitive arm movements. The men rated their exposure lower than the women when their measured exposure was similar to or higher than the women's. No difference according to age was found by Hansson et al (24) among cleaners. With respect to the influence of types of occupation, no substantial difference was found by Leijon et al (13) after the variability of the distribution of exposure was taken into consideration, but Hansson et al (24) observed that, for the same response, cleaners had a higher measured exposure than that of office workers.

Several studies examined the influence of the presence of musculoskeletal symptoms or disorders on the reporting of physical work demands. Two studies that found a differential misclassification of exposure in association with the presence of musculoskeletal complaints concluded that the misclassification was not significant because the difference was too small to have an effect on the risk estimate (ie, the ratio between the odds ratio for musculoskeletal complaints from the questionnaire compared with the odds ratio from the structured interview did not differ from unity) (13, 17). Hansson et al (24) reported that people with neck–shoulder complaints rated their exposure higher than did healthy ones when asking about arm movements, although their measured exposure was similar. Viikari-Juntura et al (22) found that workers with severe low-back pain had substantially lower correlations between self-reported neck flexion, hand posture above the shoulders, repetitive movements of the wrist and fingers, or lifting or carrying 6–15 kg than that observed by the reference method than those without low-back pain. Wiktorin et al (23) found that the presence of shoulder or neck complaints was associated with a differential misclassification of exposure to material handling. But these results are problematic. Surprisingly, the risk estimates for low-back pain based on exposure by the reference method was 0.5 for the question about lifting 16–45 kg, the result suggesting that lifting these loads was protective for back pain. The authors offered some alternative hypotheses to explain their results, including that those without low-back complaints could have underreported lifting activities and that the reference interviews may have also underestimated their exposure. In contrast to these studies, Nordstrom et al (21) found that persons with carpal tunnel syndrome, when compared with controls, had higher correlations between ergonomist's observations and their self-reported bending of the trunk, bending and twisting of the hands or wrists, and twisting of the forearm.

Discussion

A review of the methodological limits of the studies included in this review suggests that the measurement of work demands by observation or direct measurement poses as great a challenge as the design of reliable, valid self-report questions. Indeed, drawing firm conclusions from the results of this review is problematic because of the methodological limits of many of these studies. Nonetheless, this section summarizes the main results, discusses the main difficulties in interpreting them, and suggests future directions.

Questions that performed well in both sets of studies and were evaluated for both reproducibility and validity included those on the duration or presence of sitting posture, sitting–standing posture, the presence of walking, kneeling or squatting postures, the duration or frequency of hands above the shoulders, manual handling more than or less than 10 kg, general level of physical effort, presence and duration of whole-body vibration, and duration of visual display terminal use. Most of these questions measure the presence or absence of an exposure or provide only crude or limited quantification of the intensity, duration, or frequency of these physical workload exposures. Generally, they permit relative ranking of exposure rather than absolute quantification.

A wider range of questions was evaluated for reproducibility only, and good reproducibility was demonstrated. Questions on general body postures (eg, sitting, standing, and walking postures) identifying their presence, frequency, or duration generally showed good-to-excellent reproducibility. Questions on the duration or frequency of kneeling or squatting at work and jumping or climbing had relatively poor reproducibility, while questions that asked about the presence or absence of these postures (not at all, somewhat, a great deal) had a high kappa. Generally, the reproducibility of questions on postures involving specific body regions (eg, neck, shoulder, wrist, trunk) was not as good as questions on general body postures. For questions on material handling, reproducibility was better for broad categories of weights handled than for more-detailed questions with a wide range of weights and narrow intervals. Overall, questions on the level of physical effort at work, vibration exposure, and the duration of the use of visual display terminals showed good-to-excellent reproducibility.

The results of validity studies comparing self-reports with reference methods (structured interview, observation, or direct measurement) were mixed. Questions on the presence or duration of sitting posture and on the presence of standing posture generally showed high agreement. Questions on the presence or duration of sitting–standing posture generally showed high agreement, except for one study on the duration of sitting and standing (19). Questions on the duration of kneeling or squatting postures and on the distance walked at work correlated more poorly with the reference methods, unless the scales were collapsed to two or three points with broad distance categories or the questions only asked about the presence or absence of these postures. Overall, the agreement between self-reports and reference methods for questions on postures involving specific body regions (eg, neck, shoulder, trunk) was lower than for questions on sitting or standing, particularly when precision was sought about the duration or frequency. Agreement was fairly low in most of the studies on questions concerning repetitive movements. The results for questions on material handling suggest that formulations like handling loads less than or greater than 10 kg could be accurately answered, but more-detailed formulations with a wider range of weights and narrow intervals had lower levels of agreement with reference methods. Overall, questions on the level of physical effort at work, whole-body vibration, and use of visual display terminals corresponded well with the reference methods used.

The interpretation of these validity results warrants serious reflection. Several possibilities must be considered when the lack of concordance found between self-report questions on physical work demand and reference methods in some of the studies is accounted for. Possible explanations include the capacity of workers to judge physical work demands, question formulation, the choice of response scales, the time interval between the questionnaire and reference method, the variability in the distribution of exposure in the study population, and the methodological limitations of the reference methods. Most of the validity studies presented significant methodological limits, many of which are identified in table 3. The design and limits of the studies that showed poor agreement in this review do not allow us to draw conclusions about which of these explanations is the most likely.

Most of the authors of the original studies assumed that the lack of concordance represented a lack of validity; they assumed that their reference method represented a "gold standard", and thus the true exposure, and that low agreement meant that the workers' self-reports were an inaccurate measure of exposure. In some cases, this may well have been the case and, for some physical demands, workers may not be able to accurately describe their exposure on a questionnaire. Few studies have explored in depth workers' capacity to describe and judge the level of their exposure to different physical work demands or which formulations of questions or response scales allow workers to estimate exposure more accurately. It may be very difficult for workers to recall the frequency or duration of quick movements over very short periods that occur from once to many times in a workday, often at different moments of the day, such as occurs, for example, with kneeling, squatting, jumping, or climbing. Moreover, workers' mental representation of their body movements during work may differ from the representation of ergonomists or epidemiologists designing physical demand questions. Thus workers may not be aware of neck flexion or wrist positions during a task and seem to have particular difficulty describing details of the frequency or duration of many specific work postures or movements. The lack of awareness of specific body postures may be heightened by simultaneous cognitive demands or temporal pressures.

In some cases, the problem may lie with the formulation of the question or response scale rather than with the capacity of workers to report their work exposure. More work is needed to formulate response scales that quantify exposure in scales that take into account workers' mental representation of their work demands and provide meaningful measures that can be validated.

In other cases, the lack of concordance may reflect true differences in exposure because the two measures were taken at different times and the workers' exposure changed in the interim. This is particularly a possibility when there are lengthy periods between the two measures (22). Only 4 of 10 validity studies reported the time interval between the questionnaire and the reference method assessment.

In some studies, the problem may lie with the very limited variability in the distribution of exposure in the study population. Several studies had very few persons who had significant exposure to some of the physical work demands studied despite the inclusion of persons from many different job titles (8, 9, 13, 14, 17, 23). A skewed and narrow distribution of exposure can lead to a falsely low kappa despite a high percentage of agreement, sometimes referred to as the prevalence effect of the kappa. An analogous but opposite effect can be seen with estimates of ICC values when a falsely high ICC can be due to the lower variability of the distribution of responses. In studies in which few people have high exposure and most are distributed at the little or no exposure end of the exposure scale, there may be a very high correlation between self-report and the reference method for those with little or no exposure. Such data do not provide useful information about the correlation between self-report and the reference method (or test re-test reliability) for those with higher levels of exposure.

Another important potential explanation is that the differences in agreement may be due to limitations or inaccuracies inherent in the reference method selected to validate the self-reports. Overall, the results on validity were better in comparing self-reports with structured interviews than in comparing self-reports with observations, and the worst results were found when self-reports were compared with direct measurements. These findings may reflect the fact that both the structured interview of a worker and the worker's responses on a standardized questionnaire are based on that worker's self-reports. Leijon et al (13) argued against this interpretation and discussed the advantages of structured interviews that review exposure task by task and are carried out by trained professionals. The responses to structured interviews may indeed differ from the responses to self-administered standardized questionnaires, and the agreement between these responses can vary according to the type of information sought. An analysis of the variation in agreement between these responses may provide information about workers' mental representation of their physical work demands and what type of information is easier or harder to recall.

Another serious limitation of many of the reference methods used in these studies is that very few of them had previously demonstrated reproducibility and validity. Only recently have reproducibility studies been carried out on direct-measurement parameters used to characterize mechanical exposure from direct measurement (25–30). Knowledge of the reproducibility of the different parameters used to characterize mechanical exposure is essential for determining the best measurement strategy (sampling method) to capture the variance (within and between days, within and between persons) related to a specific work condition. No such analysis was performed with the reference methods used in the reviewed validity studies. Such knowledge is still lacking, and one has to be careful when interpreting results from direct exposure measurements performed once, especially for a short duration.

Problems with the generalizability of reference sampling methods may have affected several studies, particularly those with very short sampling frames that draw conclusions about daily exposure based on less than 10 or 20 minutes of observation or direct measurement. This approach can lead to an imprecise or unreliable characterization of exposure, particularly when there is substantial variability in the work from day to day, week to week, or season to season, or from one product model to another (31). Increasingly, work teams may carry out a wide range of changing tasks in the context of modular and just-in-time production methods that allow small quantities of different models to be produced.

In addition, in some workplaces, workers alternate jobs in an unpredictable pattern and are assigned to different production or assembly lines daily, or weekly depending on production needs. Other problems with the generalizability of reference sampling methods were identified by Viikari-Juntura et al (22). It may not be appropriate to generalize physical exposure observations made on one or two workers to all the others doing the same job because of interworker variation in the performance of the tasks and due to observation bias. Viikari-Juntura et al (22) concluded that, to reduce the measurement error associated with interworker variation, it is necessary to make several observations for each task. Pope et al (20) also commented on the limits of using a time-sample approach for observations and how it is possible that some tasks reported by workers may be missed by observers and lead to erroneous conclusions that workers overestimate exposure in self-reports.

Only a few of the observation methods or standardized interview methods had previously been validated. Even when an observation method had been previously validated, it was not always validated for all the postures for which it was used in the validity studies. For example, in the study by Viikari-Juntura et al (22), although the use of portable ergonomic observation had been validated for some postures, it is not clear whether the poor validity found for the question on neck posture was due to the fact that the method using portable ergonomic observation had not previously been validated for neck posture or due to the workers' difficulty to answer a structured question on neck posture.

In other cases, a lack of comparability between the exposure measured by a questionnaire and the specific variable measured by a reference method may have influenced the results. Such a situation occurs when the specific variable used to summarize the exposure parameter of interest is not comparable to the question asked of the worker, for example, measures of material handling in the Andrews et al (18) study in which workers were asked about tasks that involved "heavy lifting" without defining the precise loads, but for the observers the same variable was defined as "heavy lifting involved loads more than 18 kg" or in the study by Hansson et al (24), in which the 90th percentile of head angle distribution was compared with a question on "head bent forward a little". Selecting the appropriate parameter for summarizing a 20-minute or a whole workday of measurement is an important difficulty when direct measurement is used, and often little is known about which parameter is the most appropriate summary measure of exposure.

The choice of the best available measuring device can also be a problem in the context of validation studies. For instance, inclinometers such as that used by Hansson et al (24) measure trunk, arm, or head angle with respect to the line of gravity (ie, absolute angles instead of relative angles). Bending the trunk forward with the arms aligned with the trunk is identical to arm extension for the inclinometer, even though for the person studied it represents no arm motion. Thus there is likely to be disagreement between the direct measure results and the worker's self-report.

Another issue that could potentially influence reproducibility or validity results may be the presence of confounding or effect-modifying factors such as gender, age, education, type of work, presence of musculoskeletal symptoms, and the like. No substantial differences were found between men and women in the reproducibility of questions (14, 15). No gender influence was observed in one validity study (13), but another study did demonstrate some influence of gender among office workers (24), men showing a tendency to underestimate their exposure. Age did not appear to influence the results on reproducibility (14) or validity (24).

Wiktorin et al (14) found that workers with low education levels had lower reproducibility scores on questions about the frequency and duration of arm postures and on a question concerning material handling. This result may be due to their literacy level, or it may be that workers with low education perform more physically demanding jobs with a larger diversity of tasks and range of physical exposures. The more highly educated workers may have had little or no exposure to these physical demands, and their higher reproducibility scores may reflect greater agreement between self-reports and a reference method in the absence of exposure. No substantial differences in validity were found among different types of occupations by Leijon et al (13), but Hansson et al (24) observed that, for the same self-reported response, cleaners had a higher measured exposure than office workers.

The magnitude of the influence of the presence of musculoskeletal symptoms on reproducibility was generally small (14). However, the findings are not consistent about the influence of the presence of symptoms on the validity of questions. Generally, there was a tendency for those with symptoms to report higher exposure than that found by the reference method (13, 22). Wiktorin et al (23) suggested that people with symptoms tend to estimate their exposure less accurately, but their results did not demonstrate substantial impact on the risk estimates for musculoskeletal disorders. Wiktorin et al (23) found that people without low-back complaints had a tendency to underreport lifting activities and hypothesized that a lift of short duration may easily be overlooked, especially if the activity does not cause any harm to the person. Hays et al (32) suggested that differential misclassification may occur because workers with health problems may report greater work demands more often, either because they consider that these conditions have caused their health problems or because the health problem leads to the perception of greater work demands. Gamberale (33) suggested that the perception of exertion during physical work reflects, among other things, the interplay between the requirements of the job and the capacity of the person. The capacity to measure the differences in the accuracy of exposure measures between those with symptoms and those who are asymptomatic may be limited when symptom questionnaires that refer to pain in the past 12 months are used rather than current symptoms.

There remains a very real need for practical, reliable, and valid self-reports obtained with questionnaires on physical work demands. For example, they are needed in large population-based studies measuring trends in work conditions over time. Such studies may be helpful in identifying trends in exposure that may influence health, in identifying specific populations at risk and potentially requiring public health intervention, and in making international comparisons of work conditions. Some European studies have had this objective (34–42). These studies have influenced public policy in several European countries (2). Similar surveys have been proposed in North America. The challenge for larger population-based studies is to develop practical questionnaires that are easily understood by workers and interpreted in the same way by many workers in a wide range of different workplaces and situations.

Validated questions on physical work demands are also needed in etiologic and intervention studies of work-related musculoskeletal disorders to increase our knowledge about the multifactorial origin of these disorders and the complexity of interactions among physical, psychosocial, and personal contributing factors. Such questions may facilitate the synthesis of knowledge gained from several ergonomic intervention case studies and may permit the comparison of results from studies with small populations with the results from large surveys.

Given that it is generally not possible to measure the full diversity and complexity of work situations in a single study using direct measures and observations, questionnaires are needed to complement these measures and allow a better understanding of the variability of work exposure. It may also be useful to develop and validate work demand questions that are specific to certain industries or exposures. It may be possible to design questions and response scales for industry or occupation-specific studies that quantify exposure more precisely using terms that are meaningful to workers and which they can answer more accurately than the current set of questions tested. Most questions on physical work demands concern specific components of work activity such as posture, force, or repetition, but few address the work conditions that modulate the degree of risk associated with this activity. For example, Halpern et al (8) have a question about slippery or uneven surfaces because it may influence the risk associated with material handling. Messing and his colleagues have developed questions on sitting and standing postures that specify the degree of constraint (eg, the possibility of sitting or standing at will) and mobility associated with these postures (43, 44). In France, Leclerc (45) described the influence of ergonomists on the selection of exposure measures in the national epidemiologic survey of upper-extremity musculoskeletal disorders; they recommended taking into account dimensions of physical load associated with the latitude for changing body position such as workpace imposed by a machine or not, the possibility of taking breaks at will, and the like. Several other authors have also described the advantages of close collaboration between epidemiologists and ergonomists (2, 31, 46).

The need for practical, questionnaire-based measures for epidemiologic studies suggests that effort should be made to improve the design of individual questions and their response scales. We suggest that qualitative interdisciplinary studies be undertaken to pretest potential questions and document the mental representation that workers have of the physical demands of their work, also referred to as cognitive testing of questions (47). Workers in different settings (eg, manufacturing, health care, office work, services) should be asked to fill out a questionnaire; then the questions and responses should be reviewed with the respondent to identify what the worker understood each question to mean, what wording was unclear or confusing, which response scales are meaningful and easier to respond to, and how the worker perceives the various types of work demands. Attention should be paid to differences in the perceptions of questions with respect to type of work, education level, gender, age, and even ethnic origin. Recent immigrants from, for example, Asia, Africa, or the Middle East may have a different mental representation of physical work demands than other North American or European workers.

It has been suggested that, for some physical demands, it may be possible to calibrate points on a subjective response scale of a self-reported question to specific objective measures of that parameter of physical load (7, 48). Interesting results have been reported in studies using such calibration (49, 50). The development of new questionnaires should consider this possibility.

Indices combining several work demand questions can have better psychometric properties than individual items. This review focused on individual questions and did not address the indices that have previously been developed (5, 50). Nonetheless, we believe that physical demand indices may potentially be more reliable and capture greater variability of exposure than individual items and the development of such questionnaires should be pursued.

Future studies to test the reproducibility or validity of questions on physical work demands should ensure that an adequate sample size is used. Several methodologists have proposed criteria for estimating appropriate sample size in reproducibility studies (51–53). It is also important that the study population chosen for reproducibility or validity studies have a large range of variability of exposure and that adequate numbers of persons with high, medium, and low exposure for each physical demand under study be included. In addition, consideration should be given to the time interval between the reference method observations and the administration of the self-report questions.

The limits of the reference methods should be more thoroughly recognized. It is essential that reliable and valid reference methods be used to validate questions on physical work demands or indices that measure comparable exposures and that consideration be given to determining adequate sampling methods to capture the variability of exposure. More research is needed to validate exposure observation and direct measurement techniques. While there is no single method with demonstrated reproducibility and validity that can be used as a true gold standard, and there probably may never be, continued improvements are possible.

As Burdorf & van der Beek (7) have suggested, there is no perfect instrument for measuring all relevant dimensions of physical load simultaneously. It is probably necessary to use creative approaches that combine questionnaires with different observational methods and direct measurement techniques. An interdisciplinary approach to measuring workload exposure is particularly relevant to specific industry-based studies or studies of individual workplaces aimed at identifying risk factors in order to intervene and transform work conditions. Such a field-based approach may also lead to the development and validation of new and more-specific questions.

Acknowledgments

Rita Fernandes was the recipient of a PhD student scholarship from the Brazilian Public Foundation for Graduate Students (CAPES) that allowed her to participate in this work as a visiting research scholar at McGill University under the supervision of Dr Susan Stock. This study is part of a series of studies conducted by the Scientific Working Group on Work-related Musculoskeletal Disorders of the Québec National Institute of Public Health.

References

1. National Research Council (NRC), Institute of Medicine (IM). Musculoskeletal disorders and the workplace: low back and upper extremities: panel on musculoskeletal disorders and the workplace. Washington (DC): Commission on behavioral and social sciences and education, National Academy Press; 2001.
2. Volkoff S. Des comptes à rendre : usages des analyses quantitatives en santé au travail pour l'ergonomie [Being accountable: uses of quantitative analyses in occupational health for ergonomics]. In: Volkoff S, Cristofari MF, Gadbois C, Laville A, Molinié AF, Prunier-Poulmaire S, et al, editors. L'ergonomie et les chiffres de la santé au travail: ressources, tensions et pièges [Ergonomics and occupational health numbers: resources, tensions and pitfalls]. Toulouse (France): Octarès Éditions; 2005.
3. Vézina N, Stock S, Simard M. Problèmes musculo-squelettiques et organisation modulaire du travail dans une usine de fabrication de bottes, phase 2 : étude de l'implantation des recommandations [Musculoskeletal problems and modular work organization in a boot factory, phase 2: study of the implementation of recommendations]. Montreal: Institut de recherche Robert Sauvé en santé et en sécurité du travail (IRSST); 2003. Études et Recherches, R-345.
4. Guérin F, Laville A, Daniellou F, Duraffourg J, Kerguelen A. Comprendre le travail pour le transformer [Understanding work in order to transform it]. 2nd ed. Paris: Agence Nationale pour l'Amélioration des Conditions du Travail (ANACT); 1997.
5. Balogh I, Orbaek P, Winkel J, Nordander C, Ohlsson K, Ektor-Andersen J, et al. Questionnaire-based mechanical exposure indices for large population studies¾reliability, internal consistency and predictive validity. Scand J Work Environ Health 2001;27(1):41–8.
6. 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.
7. Burdorf A, van Der Beek AJ. In musculoskeletal epidemiology are we asking the unanswerable in questionnaires on physical load? [editorial]. Scand J Work Environ Health 1999; 25(2):81–3.
8. Halpern M, Hiebert R, Nordin M, Goldsheyder D, Crane M. The test-retest reliability of a new occupational risk factor questionnaire for outcome studies of low back pain. Appl Ergon 2001;32:39–46.
9. Torgén M, Alfredsson L, Köster M, Wiktorin C, Smith KF, Kilbom Å. Reproducibility of a questionnaire for assessment of present and past physical activities. Int Arch Occup Environ Health 1997;70:107–18.
10. Landis JR, Koch GG. The measurement of observer agreement for categorical data. Biometrics 1997;33(1):159–74.
11. Altman DG. Practical statistics for medical research. London: Chapman and Hall; 1991.
12. Currier DP. Elements of research in physical therapy. 2nd ed. London: Williams & Wilkins; 1984. p 155-65.
13. Leijon O, Wiktorin C, Harenstam A, Karlqvist L, MOA Research Group. Validity of a self-administered questionnaire for assessing physical workloads in a general population. J Occup Environ Med 2002;44(8):724–35.
14. Wiktorin C, Wigaeus-Hjelm E, Winkel J, Köster M Stockholm Music I Study Group. Reproducibility of a questionnaire for assessment of physical load during work and leisure time. J Occup Environ Med 1996;38(2):190–201.

15. Spielholz P, Silverstein BA, Stuart M. Reproducibility of a self-report questionnaire for upper extremity musculoskeletal disorder risk factors. Appl Ergon 1999;30:429–33.
16. Karlqvist LK, Hagberg M, Köster M, Wenemark M, Anell R. Musculoskeletal symptoms among computer-assisted design (CAD) operators and evaluation of a self-assessment questionnaire. Int J Occup Environ Health 1996;2:185–94.
17. Wiktorin C, Vingård E, Mortimer M, Pernold G, Wigaeus-Hjelm E, Kilbom Å, et al. Interview versus questionnaire for assessing physical loads in the population-based MUSIC-Norrtalje Study. Am J Ind Med 1999;35:441–55.
18. Andrews DM, Norman RW, Wells RP, Neumann P. Comparison of self-report and observer methods for repetitive posture and load assessment. Occup Ergon 1998;1(3):211–22.
19. Burdorf A, Laan J. Comparison of methods for the assessment of postural load on the back. Scand J Work Environ Health 1991;17:425–29.
20. Pope DP, Silman AJ, Cherry NM, Pritchard C, Macfarlane GJ. Validity of self-completed questionnaire measuring the physical demands of work. Scand J Work Environ Health 1998;24(5):376–85.
21. Nordstrom DL, Vierkant RA, Layde PM, Smith MJ. Comparison of self-reported and expert-observed physical activities at work in a general population. Am J Ind Med 1998;34(1):29–5.
22. Viikari-Juntura E, Rauas S, Martikainen R, Kuosma E, Riihimäki H, Takala E-P, et al. Validity of self-reported physical workload in epidemiologic studies on musculoskeletal disorders. Scand J Work Environ Health 1996;22(4):251–9.
23. Wiktorin C, Karlqvist L, Winkel J, Stockholm Music I Study Group. Validity of self-reported exposures to work postures and manual material handling. Scand J Work Environ Health 1993;19:208–14.
24. Hansson G-Å, Balogh I, Byström JU, Ohlsson K, Nordander C, Asterland P, et al. Questionnaire versus direct technical measurements in assessing postures and movements of the head, upper back, arms and hands. Scand J Work Environ Health 2001;27(1):30–40.
25. Burdorf A. Reducing random measurement error in assessing postural load on the back in epidemiologic surveys. Scand J Work Environ Health 1995;21(1):15–23.
26. Burdorf A and van Riel M. Design of strategies to assess lumbar posture during work. Int J Ind Ergon 1996;18:239–49.
27. Delisle A, Larivière C, Imbeau D, Durand, MJ. Physical exposure of sign language interpreters: baseline measures and reliability analysis. Eur J Appl Physiol 2005; 94:448–60.
28. Nordander C, Balogh I, Mathiassen SE, Ohlsson K, Unge J, Skerfving S, et al. Precision of measurements of physical workload during standardised manual handling, part I: surface electromyography of m. trapezius, m. infraspinatus and the forearm extensors. J Electromyogr Kinesiol 2004;14(4):443–54.
29. Mathiassen SE, Burdorf A, van Der Beek AJ, Hansson GÅ. Efficient one-day sampling of mechanical job exposure data-a study based on upper trapezius activity in cleaners and office workers. AIHAJ 2003;64(2):196–211.
30. Mathiassen SE, Moller T, Forsman M. Variability in mechanical exposure within and between individuals performing a highly constrained industrial work task. Ergonomics 2003;46(8):800–24.
31. Vézina N, Stock SR. Collaboration Interdisciplinaire dans le cas d'une intervention ergonomique [Interdisciplinary collaboration in an ergonomic intervention study]. In: Volkoff S, Cristofari MF, Gadbois C, Laville A, Molinié AF, Prunier-Poulmaire S, et al, editors. L'ergonomie et les chiffres de la santé au travail: ressources, tensions et pièges [Ergonomics and occupational health numbers: resources, tensions and pitfalls]. Toulouse (France): Octarès Éditions; 2005.
32. Hays M, Saurel-Cubizolles MJ, Bourgine M, Touranchet A, Verge C, Kaminski M. Conformity of workers' and occupational health physicians' descriptions of working conditions. Int J Occup Environ Health 1996;2:10–17.
33. Gamberale F. The perception of exertion. Ergonomics 1985;28(1):299-308.
34. Houtman I, Andries F, van den Berg R, Dhont S. Sectoral profiles of working conditions. Dublin: European Foundation for the Improvement of Living and Working Conditions; 2002. Available at: http://www.eurofound.eu.int/publications/files/EF0275EN.pdf.
35. Weiler A. Annual review of working conditions in the EU: 2003–2004. Dublin: European Foundation for the Improvement of Living and Working Conditions; 2004. Available at: http://www.eurofound.eu.int/publications/files/EF0494EN.pdf.
36. Paoli P, Parent-Thirion A. Working conditions in the acceding and candidate countries. Dublin: European Foundation for the Improvement of Living and Working Conditions; 2003. 122 p. Available at: http://www.eurofound.eu.int/publications/files/EF0306EN.pdf.
37. Pederson HS, Mahler S, Hansen CB. Danish Working Environment Cohort Study, 2000 (DWECS). Dublin: European Foundation for the Improvement of Living and Working Conditions; 2003. Available at: http://www.eurofound.eu.int/publications/files/EF03120EN.pdf.
38. Volkoff S. L'évolution du travail depuis deux décennies: intensité et fragilité [The evolution of work over the past two decades: intensity and fragility]. In: Daniellou F, editor. Les journées de Bordeaux sur la pratique de l'ergonomie [Bordeaux symposium on ergonomic practice]. Bordeaux (France): Collection Actes, Laboratoire d'ergonomie des systèmes complexes (LESC), Université de Bordeaux; 2003. p 25–41.
39. Houtman I. Work and health statistics in the Netherlands. Dublin: European Foundation for the Improvement of Living and Working Conditions; 2004. Available at: http://www.eurofound.eu.int/publications/files/EF0445EN.pdf
40. Fagan C, Burchell B. Gender jobs and working conditions in the EU. Dublin: European Foundation for the Improvement of Living and Working Conditions; 2002. Available at: http://www.eurofound.eu.int/publications/files/EF0249EN.pdf.
41. Gollac M, Volkoff S. Les conditions de travail [Conditions of work]. Paris: Éditions La Découverte; 2000. Les Éditions du travail, collection Repères.
42. Derriennic F, Touranchet A, Volkoff S. Âge, travail, santé: études sur les salariés de 37 à 52 ans: résultats de l'enquête ESTEV (Enquête santé travail et vieillissement). [Age, work, health: studies of employees aged 37-52: results of the Health, Work and Aging Survey]. Paris: Éditions de l'Institut national de la santé et de la recherche médicale (l'INSERM); 1996.
43. Tissot F, Messing K, Stock S. Standing, sitting and associated working conditions in the Quebec population in 1998. Ergonomics 2005;38(3):249–69.
44. Laperrière E, Couture V, Stock S, Messing K. Validation of questions on working posture among standing workers in Quebec. Int J Ind Ergon 2005;35(4):371–8.
45. Leclerc A. Les facteurs psychosociaux dans l'enquête sur les affections péri-articulaires et le travail répétitif [Psychosocial factors in the survey on joint disorders and repetitive work]. In: Bourgeois F, editor. TMS et évolution des conditions de travail: études et documents [Musculoskeletal disorders and

the evolution of working conditions: studies and documents. Paris: Agence Nationale pour l'Amélioration des Conditions de Travail (ANACT); 1998. 77–84.
46. Bourgeois F, Lemarchand C, Hubault F, Brun C, Polin A, Faucheux JM. Troubles musculo-Squelettiques et travail—quand la santé interroge l'organisation [Musculoskeletal disorders and work—when health challenges the organisation]. Paris: Éditions De L'Agence nationale pour l'Amélioration des Conditions de Travail (ANACT); 2000.
47. Collins D. Pretesting survey instruments: an overview of cognitive methods. Q Life Res 2003;12:229–38.
48. Spielholz P, Silverstein B, Morgan M, Checkway H, Kaufman J. Comparison of self-report, video observation and direct measurement methods for upper extremity musculoskeletal disorder physical risk factors. Ergonomics 2001;44(6):588–613.
49. Andrews DM, Norman RW, Wells RP, Neumann P. The accuracy of self-report and trained observer methods for obtaining estimates of peak load information during industrial work. Int J Ind Ergon 1997;19(6):445–55.
50. Hollmann S, Klimmer F, Schmidt K-H, Kylian H. Validation of a questionnaire for assessing physical work load. Scand J Work Environ Health 1999;25(2):105–14.
51. Walter SD, Eliasziw M, Donner A. Sample size and optimal designs for reliability studies. Stat Med 1998;17:101–10.
52. Streiner DL, Norman GR. Health measurement scales: a practical guide to their development and use. 2nd ed. London: Oxford Medical Publications; 1995.
53. Donner A, Eliasziw M. Sample size requirements for reliability studies. Stat Med 1987;6:441–8.

Received for publication: 29 March 2005