



## Review

Scand J Work Environ Health 2009;35(4):245-260

doi:10.5271/sjweh.1335

### Validity of self-reported mechanical demands for occupational epidemiologic research of musculoskeletal disorders

by [Barrero LH](#), [Katz JN](#), [Dennerlein JT](#)

**Affiliation:** Pontificia Universidad Javeriana, Carrera 7 # 40-62, Facultad de Ingeniería, Piso 3, Bogotá, DC, Colombia. [lopehugo@javeriana.edu.co](mailto:lopehugo@javeriana.edu.co)

Refers to the following texts of the Journal: [1996;22\(4\):251-259](#)  
[2001;27\(1\):30-40](#) [1998;24\(5\):376-385](#) [1999;25\(5\):404-409](#)  
[2005;31\(6\):409-437](#) [1999;25\(2\):81-83](#) [1999;25\(3\):246-254](#)  
[1999;25\(6\):505-510](#)

The following articles refer to this text: [2010;36\(1\):3-24](#);  
[2010;36\(1\):1-2](#); [2011;37\(6\):502-511](#); [2018;44\(2\):156-162](#)

**Key terms:** [epidemiologic research](#); [inter-method reliability](#); [mechanical demand](#); [mechanical exposure](#); [MSD](#); [musculoskeletal disorder](#); [occupational epidemiologic research](#); [occupational research](#); [review](#); [self-assessment](#); [self-report](#); [self-reported information](#); [validity](#)

This article in PubMed: [www.ncbi.nlm.nih.gov/pubmed/19562235](http://www.ncbi.nlm.nih.gov/pubmed/19562235)

### Additional material

Please note that there is additional material available belonging to this article on the [Scandinavian Journal of Work, Environment & Health -website](#).



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

## Validity of self-reported mechanical demands for occupational epidemiologic research of musculoskeletal disorders

by Lope H Barrero, ScD,<sup>1</sup> Jeffrey N Katz, MD, MSc,<sup>2,3</sup> Jack T Dennerlein, PhD<sup>2</sup>

Barrero LH, Katz JN, Dennerlein JT. Validity of self-reported mechanical demands for occupational epidemiologic research of musculoskeletal disorders. *Scand J Work Environ Health*. 2009;35(4):245–260.

**Objectives** To describe the relation of the measured validity of self-reported mechanical demands (self-reports) with the quality of validity assessments and the variability of the assessed exposure in the study population.

**Methods** We searched for original articles, published between 1990 and 2008, reporting the validity of self-reports in three major databases: EBSCOhost, Web of Science, and PubMed. Identified assessments were classified by methodological characteristics (eg, type of self-report and reference method) and exposure dimension was measured. We also classified assessments by the degree of comparability between the self-report and the employed reference method, and the variability of the assessed exposure in the study population. Finally, we examined the association of the published validity (*r*) with this degree of comparability, as well as with the variability of the exposure variable in the study population.

**Results** Of the 490 assessments identified, 75% used observation-based reference measures and 55% tested self-reports of posture duration and movement frequency. Frequently, validity studies did not report demographic information (eg, education, age, and gender distribution). Among assessments reporting correlations as a measure of validity, studies with a better match between the self-report and the reference method, and studies conducted in more heterogeneous populations tended to report higher correlations [odds ratio (OR) 2.03, 95% confidence interval (95% CI) 0.89–4.65 and OR 1.60, 95% CI 0.96–2.61, respectively].

**Conclusions** The reported data support the hypothesis that validity depends on study-specific factors often not examined. Experimentally manipulating the testing setting could lead to a better understanding of the capabilities and limitations of self-reported information.

**Key terms** self-reported information; mechanical exposure; inter-method reliability

Instruments based on self-report (referred to as “self-reports” hereafter) are used commonly for the assessment of mechanical exposures in epidemiologic research of musculoskeletal disorders (1). Self-reports have unique advantages for a number of applications: (i) they can be less expensive than observation-based and direct measurement instruments to assess large populations (2); and (ii) they constitute a feasible method to assess exposures that occur with highly irregular patterns, for example exposures that change seasonally (3), in the past (4), or under conditions where research space for interviewing is limited or privacy must be maintained (5).

Numerous studies have warned researchers about the lack of validity of self-reports, based on the low

level of association that is frequently reported between this type of instrument and observation-based or direct measurement methods for the assessment of mechanical demands (6–8). It has been argued that while self-reports effectively convey relative differences in exposures of heterogeneous populations, they are imprecise measures of the absolute levels of the exposure (9). Others, in contrast, have stated that the validity of self-reports for the assessment of mechanical exposures cannot be appropriately established with the information currently available (10, 11). These reviews have coincided in that the reported agreement between self-reports and other more objective measures of exposure assessment may be due to the methodological characteristics of the studies

<sup>1</sup> Department of Industrial Engineering, Pontificia Universidad Javeriana, Bogotá, Colombia.

<sup>2</sup> Department of Environmental Health, Harvard School of Public Health, Boston, MA, USA.

<sup>3</sup> Orthopaedic and Arthritis Center for Outcomes Research, Brigham and Women’s Hospital, Harvard Medical School, Boston, MA, USA.

Correspondence to: Lope H. Barrero, Pontificia Universidad Javeriana, Carrera 7 # 40-62, Facultad de Ingeniería, Piso 3, Bogotá, DC, Colombia. [E-mail: lopehugo@javeriana.edu.co]

testing such agreement and not the true capability of self-reports to measure the exposure of a population.

In their review, Stock et al (10) systematically explored the effects of question, response scale formulation, and criterion methods on the validity of questionnaires and interviews used to assess occupational mechanical exposures. The authors found that low measured validity may often be due to the poor formulation of questions, which limits the ability of the study population to accurately report their exposures. Also, the authors noted that many of the studies included in the review used reference methods that were untested, not necessarily able to capture the variability of work (which may have been captured by self-reports), or not comparable to the questions being tested. Stock et al's review found it difficult to establish definitive conclusions about the validity of these self-reports due to the methodological limitations in many of the studies that were reviewed.

Another recent review by Barriera-Viruet et al (11) assessed the quality of studies that tested the validity of self-reports, and the association between study quality and the measured validity. To assess the quality of each study, the authors considered the reported data, the validity of the criterion method, and the employed statistics. The authors reported that validity studies using direct measurements as the criterion method tended to have better quality scores than those based on observations. Barriera-Viruet et al's study, however, did not find a relation between the overall quality of the study and the reported agreement between the self-report and the criterion method. This finding was used to suggest that what drives the reported validity of self-reports is not the quality of the study, but rather the fact that self-assessments can be considered fundamentally a psychophysical measure of exposure which may not be lineally related to physical stimuli. Similarly, self-reports may reflect the subject's response to a variety of stimuli simultaneously, some of which may not be controlled or measured.

Our study aims to advance this discussion by quantifying the effect of key characteristics of validity assessment methods on the measured validity of self-reported mechanical demands currently published in the literature.

Specifically, we explored the hypothesis that the use of non-comparable reference methods explains, at least partially, the low-to-moderate measured validity of self-reports. Also, in recognition of the fact that the correlation, a commonly used statistical measure of validity, is affected by the amount of variability present in the variables being compared, we explored the relation between the heterogeneity of the population with regard to the exposure of interest and the reported validity of self-reported ergonomic demands. To do this, we classified all validity assessments of self-reports available in the recent literature by: (i) methodological characteristics related to study population, (ii) reference method, (iii) statistical measure of validity employed, and (iv) exposure and exposure dimension measured by the tested self-report (12). Notably, we documented whether demographic features of the study population such as age, gender, and education level were reported in the studies. This is pertinent as demographic features may be related to the accuracy of self-reported data (13) for example, via differences in the physical capability to execute a given task (14) or in the capability to understand and process information in response to a question (15). We used the information about validity assessments both to conduct the tests mentioned above and describe the trends in validity assessment research of self-reported mechanical demands.

## Materials and methods

### Manuscript search

We aimed to identify original peer-reviewed articles, published in English between January 1990 and March 2008, testing the inter-method reliability of self-reported mechanical demands occurring at the workplace. Mechanical demands of interest included: assuming awkward postures, applying force or executing forceful movements, doing repetitive movements, and measures of physical exertion. Self-report-based instruments of interest included: self-administered questionnaires,

Table 1. Search terms.<sup>a</sup>

Group 1: Self-reported information	Group 2: Ergonomic demands of interest	Group 3: Occupational settings
Diary, interview, log, logbook, "log-book", questionnaire, self-administered, self-administration, self-assessment, self-completed, self-estimation, self-report, survey, "self report", self-reported	Ergonomic, exertion, force, lifting, "manual materials handling", "mechanical exposure", "mechanical factor", "mechanical load", "mechanical risk factor", motion, movement, "physical demand", "physical exertion", "physical exposure", "physical load", "physical risk factor", "postural demand", "postural exertion", "postural exposure", "postural factor", "postural load", "postural risk factor", "repetitive load", "repetitive motion", "repetitive movement", "repetitive work", vibration, "work demand", "work exertion", "work load", "work pace", "working load", "working pace", posture, workload	Employee, industrial, industry, job, occupation, occupational, professional, work, worker, working, workplace, workstation, ergonomic
NOT (child or children or cancer)		

<sup>a</sup> Terms within each column are connected by the operator *or*; terms between columns are connected by the operator *and*.

interviews, and diaries. Our study also included self-reports that were used to help estimate other measures of exposures (eg, self-reported time can be used to estimate time-weighted biomechanical exposures). Only single questions that were tested or compared with other exposure assessment methods were included; that is, indices based on self-reported information were excluded.

We searched three major databases: EBSCOhost (Academic Search Premier and PsycINFO), Web of Science (Science Citation Index Expanded, Social Sciences Citation Index, and Arts and Humanities Citation Index) and PubMed. Three groups of search terms were used: (i) self-reported information, (ii) ergonomic demands of interest, and (iii) occupational settings (table 1). We also checked the references of the selected articles for additional manuscripts.

### Manuscript selection

One of the authors conducted a first screening of all identified articles based on titles and abstracts. This stage aimed to exclude any manuscripts that tested the validity of instruments for health outcomes evaluation, scale-based self-reports for evaluation of ergonomics demands (eg, self-reports that sum-up the results of various questions to estimate a total level of exposure), or studies testing self-reports with methods different from the reference or criterion validity method (eg, studies using construct validity, face validity). If there was any doubt about eligibility, the article was retained. Those articles that passed this first stage of review were obtained in full manuscript form and read in-depth to check their eligibility. At both stages, articles were retained if all the answers to the three following questions were yes: (i) Did the study evaluate a mechanical demand (including assuming awkward postures, applying force or executing forceful movements, doing repetitive movements, and measures of physical exertion)? (ii) Did the study compare the self-report of demands with another measure of demands? (iii) Did the study test the validity of a single-question?

### Data extraction

We captured information that characterizes individual validity assessments of questions, including: type of self-report; type of criterion method (ie, direct measurement, observation, or other self-report based instrument); employed question and response scale; type of population (ie, blue- versus white-collar workers); industry; period of exposure covered by the self-report and criterion method; measure of validity (eg, correlations, regression parameters, Kappa statistics, and mean differences between self-report-based and criterion method-based measures of exposure, etc.); sample size; and study population demographics (ie, age, gender, and education).

### Self-report classification by exposure and exposure dimension

The classification of self-reports according to the exposure they evaluated was based on previously proposed taxonomy schemes of physical exposures (10, 16). We added a new category for the exposure “activity/task” to include validity assessments testing the occurrence of activities or tasks in the workplace (table 2).

The classification of self-reports according to the exposure dimension they address (duration, frequency, and/or magnitude) was conducted via a 3-step process. First, both the part of the question that defined the focus of the question (15) and the question response were separately classified into the three different dimension categories. Second, the question predicates were examined for particular words suggesting the intention to further quantify the exposure in terms of duration, frequency, or magnitude (table 3). Third, each self-report was reviewed individually in order to understand its purpose and integrate information from the previous two stages.

**Table 2.** Specification and examples of classification of self-reports by exposure

Specification	Example
Posture	Questions about whole body postures (sitting, standing, squatting, kneeling) and particular body part postures (hands above shoulder, neck posture). Therefore, questions asking what proportion of the day are you kneeling would be classified as a postural exposure, whereas a question asking how many times you kneel would be classified as a movement exposure.
Movement	Questions about the action of squatting or moving a body part (eg, bend or rotate back). An important subcategory is all questions about manual materials handling (lifting, carrying, pushing, and pulling).
Repetition	Questions specifically asking about the duration of repetitive movements or qualifying a repetitive movement (eg, how many times per minute...). Questions about movement of the arms, hands or fingers. Questions about work in an assembly line. In one case, the use of a keyboard was not classified as activity exposure, but as repetition because of the qualifier “intense” keying.
Physical exertion	Questions looking to rate the physical exertion of jobs or tasks.
Force	Questions asking about the use of pinch grasp; questions asking about the weight of objects; and questions asking about external force direction or body posture with the purpose of estimating internal loads (eg, compression or shear forces in the L4/L5 joint). Questions about manual material handling of objects of a given weight were classified in the movement exposure.
Vibration	Questions asking about the duration of the use of vibrating tools or questions asking about driving any or particular types of vehicles.
Activity/task	Questions asking about the duration or frequency of occurrence of activities such as computer use, walking, cleaning, maintenance, and meetings.
Other	Questions about workstation design, adjustability of the workstation, presence of pressure on the tip of the thumb, and the use of hands as a hammer.

When a self-report was said to evaluate the magnitude dimension of an exposure (ie, posture, movement, repetition, and vibration), we created two subcategories. In the first subcategory, we classified simpler questions investigating the presence of an exposure (eg, using a dichotomous response). In the second subcategory, we classified more complex questions attempting to quantify the magnitude of the exposure in greater detail (eg, using ordinal scales, or qualifying the magnitude of the exposure in the focus or predicate of the question). An illustration of this process is presented in figure 1.

### Trends in validity testing of self-reports research

All identified validity-tested self-reports were organized by: type of self-report; type of criterion method; demographic information; whether the actual question was presented in the manuscript; exposure and exposure dimension; industry; and the type of population being tested.

To avoid the double counting of non-independent validity assessments presented in the same study and therefore to have a more accurate representation of the distribution of validity assessments by methodological characteristics and topics, the following sources of non-

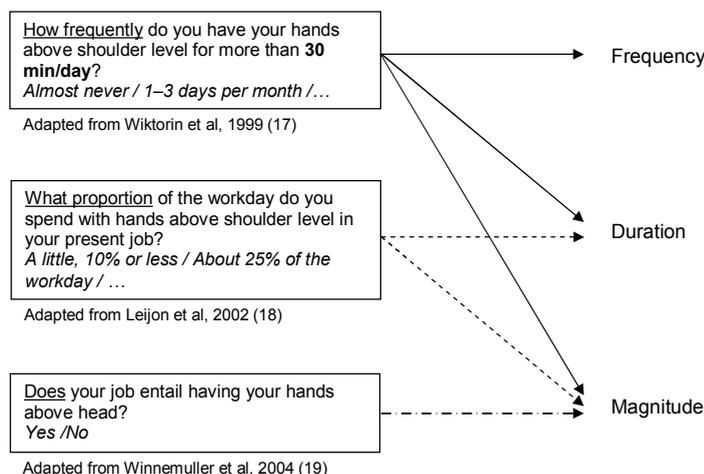
independence were considered: (i) validity assessments based on the same raw data, but transformed prior to analysis (eg, information originally collected with a larger number of categories and then collapsed in the analysis stage to fewer categories); (ii) validity assessments based on the same data but with more than one statistical measure of validity presented; and (iii) validity assessments of the same instrument in a subset (or subsets) of the original population. Assessments were considered independent when the same self-report was tested against the same criterion, but with a different time lapse between the application of the reference method and the administration of the self-report. We also considered assessments as independent when the same self-report was tested with a different reference method or with a different question-response type. Otherwise, we considered assessments coming from the same manuscript as independent.

### Relation of the comparability between self-reports and reference methods with reported validity

The subset of validity assessments reporting correlation coefficients (both Spearman and Pearson correlation coefficients) as statistical measures of validity were

**Table 3.** Classification of self-reports by exposure dimension based on question content.

Dimension	Question beginning	Question response	Question predicate
Duration	What proportion (of the day)? For how long?	Time, percentage/proportion (of time), start and end times, a little/somewhat/a lot (although it could also be frequency)	Continuously, without a pause
Frequency	How frequently? How often? How many?	Frequency, number of times (per unit of time)	Regularly, frequently, many times per day
Magnitude	How much? How exerting? How far? How heavy? Does your job require? What posture? How fast (hand movement)?	Distance, object weight, angle/posture, velocity, acceleration, amount (eg, a little, somewhat, a lot), level (of effort or activity), location/posture relative to workstation or workstation part. A special case for magnitude of posture was when the question aimed to identify whether the job was done mainly sitting, standing or a combination of both	Parts of the question qualifying an angle, qualifying the effort during manual materials handling, or qualifying the type of vibration source were flagged for the dimension, for example: back bent a lot, back bent between 20–60 degrees, carrying on one shoulder, use of hand-held polisher



**Figure 1.** An example of the process whereby questions are classified according to the exposure dimension being assessed. The underlined part of the question is the focus of the question. The predicate refers to the remaining part of the question. The response scale is presented in italics. In the upper box, there is an example of a question on the frequency, duration, and magnitude of the exposure. In the middle box, there is an example of a question on the duration and magnitude of the exposure. In the lower box, there is an example of a question on only the magnitude of the exposure.

included in this analysis. Initially, the comparability of self-report and reference methods was evaluated separately for the exposure time period and the exposure construct. Each factor was rated by two independent raters as “corresponds,” “does not correspond”, or “is arguable,” which included cases where the correspondence was difficult to discern given the information available in the manuscript. When disagreement occurred, consensus was reached after re-examining the available information. Then, using all potential combinations of the consensus evaluations of the time period and construct’s correspondence, we created an overall correspondence index (table 4) with three levels: high (+), medium (+/-), and low (-).

The time period was said to correspond when it was well defined in the self-report (eg, time spent sitting during the past work shift) and it matched the measurement period used by the criterion method. The time period was also said to correspond when the time period was not specified in the self-report (or was specified as an average or usual amount of time in the present job or in the past year) and the criterion method was applied (or referred) to more than two work shifts (thereby enabling a better reflection of the variability of the exposure).

The constructs were said to correspond when the units of the self-report corresponded with the units of the reference method (eg, both self-report and criterion method measured angles, time, or frequencies in the same units). Constructs were also said to correspond when self-reports asked for physical effort or physical exertion (eg, using Borg scales or other alpha-numeric scales) and the reference method used either heart rate-based or acceleration units-based (eg, from a portable accelerometer) measures of exposure. The correspondence between self-report-based measures of physical effort and acceleration units was considered appropriate because it has been shown that movement detected by acceleration-based instruments is related to muscle load and, therefore, to energy expenditure (20), a measure of physical effort.

Generalized estimating equations (GEE) (PROC GENMOD, SAS v9.1, SAS Institute, Cary, NC, USA) were used to assess the relation of the level of correspondence of self-reports and criterion methods (independent variable, rated as high or low) with their measured correlation (dependent variable, rated as high or low, depending on whether the correlation was above or below the median correlation of 0.49). GEE was used to account for “within study dependency” of validity assessments.

#### Relation of heterogeneity of study population with reported validity

For this analysis, we included reviewed validity assessments that: (i) used correlation coefficients as the statistical measure of validity, and (ii) included

information about the variability of the exposure of interest (ie, standard deviation of the exposure as measured by the criterion method or in its absence as measured by the self-report being tested), which was thought to be a measure of the heterogeneity of the population regarding the exposure being measured. Standard deviations were not used directly to measure the variability of the exposure in the validity studies. Instead, the coefficient of variation (CV), defined as the standard deviation divided by the mean of the exposure (or median when the mean was not reported) was used. The CV allows for the evaluation of the variability of a construct in relation to its mean; it is most useful in comparing the variability of several different samples, each with different means (21). Therefore, this measure was considered an appropriate surrogate for variability for this analysis where multiple studies (with multiple exposures measured in a variety of populations) were aggregated. When multiple sub-groups of workers were used as the study population and individual estimates of mean and variability were presented, we pooled them to get an estimate of the CV for the entire study population.

Again, we used GEE to conduct the association analysis while accounting for dependency of validity assessments. The dependent variable was the reported correlation (rated as high or low, depending on whether it was above or below the median-reported correlation of 0.49). The independent variable was the CV (rated as high or low, depending on whether it was above or below the median CV of 1.00).

## Results

The search identified a total of 9563 articles. A first screening based on the title and abstract resulted in 126 articles being selected for in-depth review. Of these 126 studies, 40 included at least one inter-method test of

**Table 4.** Comparability between self-reports and criterion methods index.<sup>a</sup>

Construct	Time period		
	Corresponds	Is arguable	Does not correspond
Corresponds	High (+)	High (+)	Medium (+/-)
Is arguable	High (+)	Medium (+/-)	Low (-)
Does not correspond	Medium (+/-)	Low (-)	Low (-)

<sup>a</sup> The comparability between self-reports and criterion methods is evaluated independently for the evaluated time period and the evaluated construct as “corresponds”, “arguable or hard to tell”, or “does not correspond” (see text for details on the evaluation criteria). The combination of the two factors were used to evaluate the overall comparability between self-reports and criterion methods as: high (+), medium (+/-) or low (-).

individual questions asking about mechanical demands. These 40 studies reported a total of 490 independent individual-question validity assessments (figure 2).

Three separate reports follow about (i) trends and characteristics of current validity assessments; (ii) the relation between the comparability of self-reports and criterion methods, and the reported validity of self-reports; and (iii) the relation between the variability of the exposure in the study population and the reported validity of self-reports.

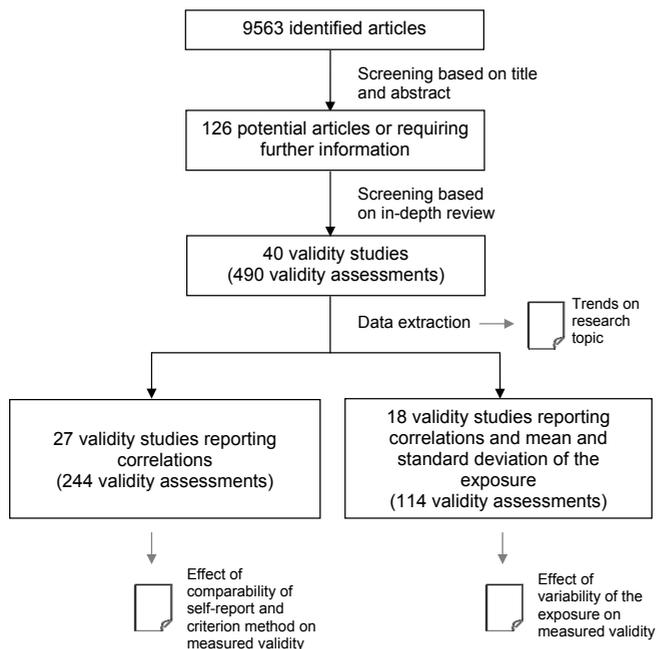
### Trends in validity testing of self-reports research

Self-administered questionnaires were the most commonly tested type of self-report (69.2% of validity assessments), followed by diaries/logs (16.8%) and interviews (13.3%). Of all evaluated self-reports, 74% were tested against observation-based methods. Direct measurement criterion methods were used only in 18.4% of all validity assessments. In 6.5% of the cases, self-reports were tested against other self-reports. Exposures in most categories were primarily tested against observation methods. Exposures in the “movement” category used observations as the criterion method in 96.5% of the assessments. In contrast, validity assessments of self-reported vibration exposures were tested mostly against direct measurement methods (figure 3). See the appendix for an inventory of validity assessments of single question for mechanical demands evaluation.

Postures and movements are the most frequently evaluated mechanical demands, accounting for 55% of all validity assessments. Overall, 35% of all the tested self-reports evaluated the magnitude of the exposure alone. Another 42% of the self-reports evaluated simultaneously the magnitude and duration of the demands. However, most self-reports (62.5%), evaluating simultaneously the magnitude and duration dimensions of the exposures, tested only the duration in which an exposure was present (figure 4).

Important demographic information such as the education level, age, and gender of the study population was frequently not included in the validity assessment. For example, only one out of the 40 studies reported the education level of the study participants (14). Twelve studies presented at least one validity assessment with missing information on the mean or the distribution of the age (eg, standard deviation or range of age) of the study population (17, 19, 22–31). Lastly, five studies did not report the gender distribution of the study population (19, 22, 23, 32, 33).

The exact wording or a condensed version of the evaluated self-report (14, 17–19, 22, 24, 25, 29–31, 34–45) was reported in 22 studies. Another three studies provided some validity assessments, for which there was full presentation of the self-report, but also provided validity assessments, for which the wording of the self-report was not fully presented (26, 46, 47). Five other studies reported assessments in which either the question



Three separate analyses which used subsets of all 490 validity assessment identified in this review are reported.

Figure 2. Manuscript search, identification, and selection process.

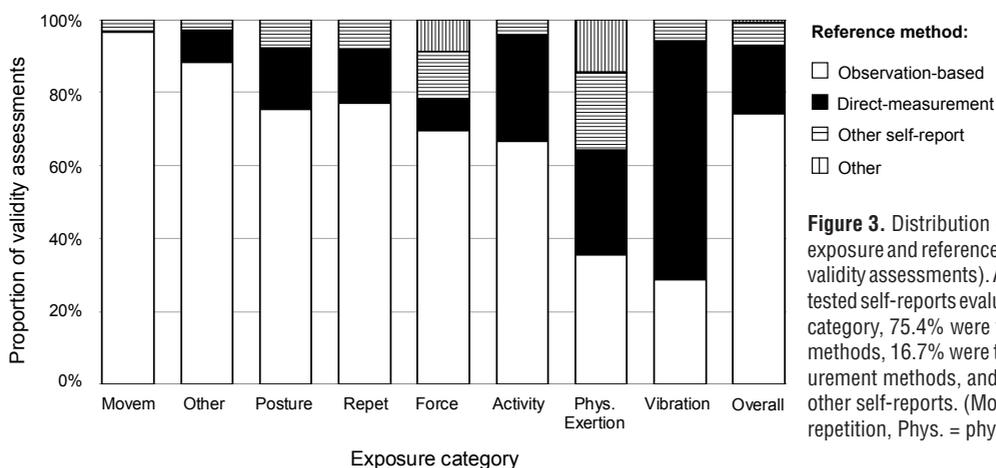
or the response was reported, but not both (4, 27, 33, 48, 49). For the remaining nine studies, there was little or no information at all about the wording of the evaluated self-report (23, 28, 32, 50–55).

**Relation of the comparability between self-reports and reference methods with reported validity**

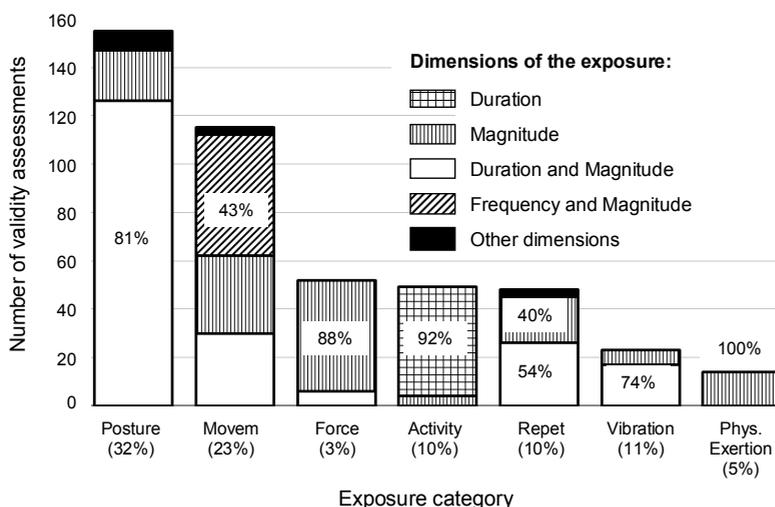
Of the 40 studies, 27 presented correlations as a measure of validity of self-reports, which resulted in a total of 244 validity assessments used in this analysis (figure 2). There was a wide range of correlations between self-reports and reference methods across exposures, exposure dimensions, types of self-report, and the level of correspondence. When these 244 assessments were classified depending on the reported correlation level (above or below the median correlation), we found that validity assessments of exposures in the force, repetition, and

movement categories tended to result more frequently in correlations below the median (100%, 85.7%, and 53.1%, respectively). Validity assessments of self-reports evaluating other exposure categories such as physical activity, posture, activity, and vibration resulted more frequently with correlations above the median (60%, 67.7%, 71.4%, and 72.7%, respectively)

Of all validity assessments reporting correlations, 62% (152 assessments) tested self-reports against reference methods measuring the same period and construct as the self-report. However, another 30% (72 assessments) of these validity assessments tested self-reports against reference methods for which either one or both of the evaluated period of time or construct did not match the self-report; 97% of this lack of correspondence was due to low correspondence in the time period being evaluated. Therefore, we could evaluate only the relation between time period correspondence and reported correlation.



**Figure 3.** Distribution of validity assessments by exposure and reference method employed (N=490 validity assessments). As an example: of all validity tested self-reports evaluating the posture exposure category, 75.4% were tested against observation methods, 16.7% were tested against direct measurement methods, and 7.7% were tested against other self-reports. (Movem = movement, Repet = repetition, Phys. = physical)



**Figure 4.** Distribution of validity assessments by exposure and exposure dimension assessed (N=490 validity assessments). As an example: 32% of all validity tested self-reports evaluated the posture exposure category; 81% of these self-reports evaluated simultaneously the duration and magnitude dimensions of the posture. (Movem = movement, Repet = repetition, Phys. = physical)

Validity assessments of exposure categories (such as physical activity, force, vibration, posture, and repetition) were conducted frequently under conditions where the time period evaluated by the self-report was said not to correspond with the time period evaluated by the criterion method (50%, 46.7%, 36.4%, 34.4%, and 32.1%, respectively).

There was a positive association between the level of comparability of self-reports and reference methods, and the reported validity of self-reports (OR 2.03, 95% CI 0.89–4.65) (table 5). There were not enough subsets of validity assessments using the same type of self-report and criterion method to be able to adjust for these variables.

#### Relation of heterogeneity of study population with reported validity

Of the 40 studies, 18 presented both correlations as a measure of validity and information to estimate the CV of the exposure of interest in the study population. These 18 studies resulted in 114 validity assessments that were used in this analysis (figure 2). Validity assessments of the repetition, activity, force, and posture exposure categories were conducted frequently in populations with exposure variability below the median variability of the studies included in this analysis (95%, 83.3%, 61.5%, and 45.2%, respectively).

**Table 5.** Unadjusted association analysis study characteristics and reported validity of self-reports. (r = correlation between the self-report and the reference method; CV = coefficient of variation)

Study characteristics	Validity		OR	95% CI <sup>a</sup>
	High ( $r > 0.49^b$ )	Low ( $r \leq 0.49^b$ )		
Comparability of time construct between self-reports and reference methods (N=226 <sup>c</sup> )			2.03	0.89–4.65
High	80	72		
Low	35	39		
Heterogeneity of population (N=114)			1.60	0.96–2.61
High (CV>1.00 <sup>d</sup> )	37	20		
Low (CV≤1.00 <sup>d</sup> )	19	38		

<sup>a</sup> Estimates using Generalized Estimating Equations.

<sup>b</sup> The cut-off corresponds to the median of all validity assessments reporting correlations (N=244); association analyses with subsets of those validity assessments are presented in this table.

<sup>c</sup> The comparability of the time period between self-reports and reference methods was judged “arguable or hard to tell” for 18 of the 244 validity assessments reporting correlations; only those validity assessments that were judged to have high or low correspondence were included in this analysis (226 assessments).

<sup>d</sup> The cut-off corresponds to the median coefficient of variation of all validity assessments for which this information could be estimated (N=114).

There was a positive association between the heterogeneity of the study population with regard to exposure of interest (as measured by the CV) and the reported validity of self-reports (OR 1.60, 95% CI 0.96–2.61) (table 5). There were not enough subsets of validity assessments, using the same type of self-report and criterion method, to be able to adjust for these variables.

## Discussion

### Main findings

This study aimed to examine whether methodological aspects of validity assessments are likely to explain findings of validity testing and to document current trends in this topic of research. Overall, our study supports the findings of Stock et al (10) that it is difficult to evaluate the validity of self-reported exposures based on currently available validity assessment studies. We provided quantitative evidence to support the claim that methods and characteristics of the study population used in current validity assessments may explain, at least partially, the low-to-moderate reported validity of self-reported mechanical demands. This evidence suggests that commonly accepted views about the poor-to-moderate validity assessment should be reconsidered. Furthermore, reported results cannot be easily generalized to other populations as study populations have not generally been well described.

We found that validity assessments have primarily been conducted for self-reports measuring posture duration and movement (including manual materials handling) frequency, together representing 55% of all tested exposure dimensions. Although this may simply reflect their commonality in occupational epidemiologic research, it seems reasonable that the development and testing of self-reports assessing other exposure categories and dimension combinations should be conducted before conclusions are made about the validity of self-reported mechanical demands as a whole. Also, the fact that mainly the duration and frequency dimensions of the exposure were evaluated for postures and movements respectively was not surprising. These dimensions capture what is traditionally more important regarding posture and movement. However, this marked distribution highlights the fact that more elaborated patterns of exposure (eg, whether the total duration of a posture during a day is achieved during a single moment of exposure or by adding multiple moments of a exposure throughout the day) were not typically assessed by the self-reports included in this review.

Self-reports were validated mainly against observation-based methods of exposure assessment. The use of

this type of reference method was particularly frequent for exposure categories such as movement, repetition, and posture. This may be of concern because the validity of observational methods themselves has also argued for the assessment of detailed postures and the assessment of rapid movements (56, 57). Therefore, in agreement with previous reviews (10), it is recommended to observe results about the validity of self-reported information cautiously in relation to the validity of the employed reference method.

Frequently, these validity studies did not report important demographic information on the study population such as age, gender and, most importantly, education, which is one likely determinant of an individual's cognitive capability to understand and respond to questions (15). Similarly, we found that validity studies often do not report the evaluated questions in full, which is essential to corroborate and evaluate previous research. Overall, the lack of this essential information makes it difficult to establish the degree to which these studies can be generalized to other situations and working populations.

This study's findings highlight that methodological and population differences across studies are related to the measured validity of self-reports and should, therefore, be considered when interpreting the aggregate results from multiple validity studies. We found a high frequency of mismatch between self-reports and criterion-based methods. Most often, the discrepancy corresponded to a self-report asking for average or usual exposures while the criterion method evaluated exposures for a specific time period (eg, a few hours or a work shift). We found that the mismatch was related to lower reported correlations. We also found that the higher the heterogeneity of the study population with regard to the exposure of interest, the greater the number of reported correlations.

Our analyses of the effect of design and study population characteristics on the measured validity of self-reports have important limitations. First, our analyses resulted in only moderate associations and were conducted for aggregated validity reports for multiple exposures and using various types of self-reports and reference methods. Therefore, we cannot rule out alternative explanations to observed correlation levels in the validity tests, other than the heterogeneity of the study population and the correspondence between the self-report and reference methods. Second, only studies reporting correlations could be used in these analyses. Although correlations were by far the most-used measure of validity, it remains unknown whether methods and study characteristics of studies reporting correlations are representative of all validity studies. Furthermore, the frequent use of correlations as a measure of validity raises another issue. Correlations only measure one aspect of the validity of self-reported information, namely, what proportion of the

variability of self-report estimates is due to the exposure as measured by the reference method. In this way, high correlations can be achieved as long as large values of self-report based estimates are associated with large values of the reference methods. This may occur even in the presence of systematic bias in the self-report based estimates. Therefore, our results should be observed within the limitations of correlations as a measure of exposure. Third, we used a qualitative method to assess the comparability between self-reports and reference methods. Therefore our assessment of the level of comparability may be subject to misclassification; however, efforts were made to have independent evaluations of the comparability study, which is expected to reduce error in this assessment. Lastly, the level of comparability between self-reports and reference methods, and the variability of the exposure in the study population were related to correlations regardless of whether they were Pearson or Spearman correlation coefficients. This was done to maintain a larger dataset. We expect little effect due to this aggregation of coefficients as they were distributed similarly across validity assessments that were said to have good and low comparability between self-reports and reference methods, and assessments with lower and higher exposure variability.

#### Implications for epidemiologic research on musculoskeletal disorders

Substantial research effort has been invested in assessing the validity of self-reports as a source of information of occupational mechanical exposures. Burdorf (1) noted that very few self-reports used in occupational epidemiologic research have been tested for validity. Since then, at least 40 studies have tested the validity of 490 individual questions dealing with a variety of mechanical exposures. These efforts have not been as effective as desired in firmly establishing the circumstances under which self-reports are more likely to be valid.

Our review suggests that the lack of strong conclusions about the validity of self-reported mechanical demands is largely due to two factors: (i) aggregating information from multiple observational studies is difficult, and (ii) currently employed field validity assessment methods are not optimal to learn about this type of self-reported information. Although a more radical solution could be to avoid the use of self-reported information of mechanical demands, this seems impractical both due to the cost of not using it and the difficulty of accomplishing this under certain conditions (58). Researchers may agree that, instead, self-reported information should be confined to certain pieces of information required for occupational exposure assessment (9). However, the question remains: to what parts of the exposure assessment process should self-reports be confined (59)? To answer this question,

we propose a two-pronged strategy: (i) the use of agreed-upon minimum standards for both testing and results reporting in field validity testing, which would allow for better use of this type of research; and (ii) the use of more controlled conditions that would allow for a more systematic understanding of self-reported information.

The use of field validity testing is to some extent unavoidable, as researchers are compelled to validate their tools and methods in a representative sample of the population in which they expect to apply these instruments. Therefore, the validation study is used to explore the magnitude and direction of bias that would be introduced to the association between the exposure of interest and musculoskeletal disorders; however, this information is typically not used to calibrate or formally correct the measurement error in the full population of interest. Although the results from observational validity studies are most applicable to the specific study population in which they are conducted, these studies can help stimulate important improvements by reporting complete information about factors that may influence the measured validity of self-reports.

Enabling a proper meta-analysis to be conducted on aggregate results, a summary of the proposed minimum information that should be reported from validity assessments of self-reports includes:

- exact wording of the body and response scale of the questions used;
- context of the question (eg, how many and what other questions were asked, time allowed, whether additional aid was available);
- distribution of the exposure of interest in the population, as measured with both the criterion method and the self-report measure;
- appropriate measures of validity other than (or in addition to) correlations such as mean differences, regression parameters, misclassification matrices;
- confidence intervals or other appropriate measures of variability of the estimates of validity;
- demographic characteristics of the evaluated population (age, gender and education, health status);
- distribution of the measurement errors as well as statistical measures of validity in relevant subgroups of the study population (eg, occupation, age, gender, musculoskeletal disorder status groups);
- key words in the title or abstract such as “inter-method reliability” or validity; and
- standard additional information for scientific publications.

We further propose to complement observational field-based validity information with experimental

research in order to investigate explanatory mechanisms for validity findings. Experimental research would enable a more appropriate adjustment for exposure distribution and workers' differences; it would also allow the use of more accurate reference methods. Finally, it would enable investigators to systematically test the effect of question content, format, and administration. This in turn would facilitate an understanding of the relation between the criterion and self-reported mechanical exposures under different work conditions and working populations, and refine methods and instruments for data collection.

For a long time, experimental testing has advanced our knowledge about the relationship between work load and perceived physical exertion (60, 61). This has permitted researchers to gain better insight into what reported working demands really mean (62, 63). Recently, two studies have assessed the potential determinants of accuracy of self-reported repetitive work and self-reported task durations (58, 64). Petersson et al (64) showed that the duration and repetition of tasks can affect individuals' accuracy; more specifically, they reported that while very short tasks (2.5 minutes) can be overestimated by 100%, longer tasks (37.5 minutes) can be reasonably estimated. In a more recent study, and using a working population, Barrero et al (58) demonstrated that task-related factors such as the work pattern (whether individuals execute tasks continuously or are intermittently interrupted by other tasks) and the physical and cognitive demands of the tasks can affect the accuracy of self-reported tasks durations, for example up to 30% in tasks lasting 40 minutes. In contrast to validity studies for specific populations, the results from systematic testing can be potentially used across multiple populations to correct bias and interpret the self-reported physical exposure information gathered (65).

In conclusion, the use of self-reported mechanical demands for occupational epidemiologic research requires further, better validity testing research. We believe the full potential of self-reports in occupational epidemiologic research is still to be discovered.

## References

1. Burdorf A. Exposure assessment of risk factors for disorders of the back in occupational epidemiology. *Scand J Work Environ Health*. 1992;18(1):1-9.
2. Wigaeus-Hjelm E, Winkel J, Nygård C-H, Wiktorin C, Karlqvist L; Stockholm MUSIC 1 Study Group. Can cardiovascular load in ergonomic epidemiology be estimated by self-report? *J Occup Env Med*. 1995;37:1210-7.
3. Wells R, Norman R, Neumann P, Andrews D, Frank J, Shannon H, et al. Assessment of physical work in epidemiological studies: common measurements metrics for exposure assessment. *Ergonomics*. 1997;40:51-61.

4. Torgén M, Winkel J, Alfredsson L, Kilbom Å, Stockholm MUSIC 1 study group. Evaluation of questionnaire-based information on previous physical work loads. *Scand J Work Environ Health*. 1999;25(3):246–54.
5. Knibbe JJ, Friele RD. The use of logs to assess exposure to manual handling of patients, illustrated in an intervention study in home care nursing. *Int J Ind Ergon*. 1999;24:445–54.
6. Van der Beek AJ, Fring-Dresen MHW. Assessment of mechanical exposure in ergonomic epidemiology. *Occup Environ Med*. 1998;55:291–9.
7. David GC. Ergonomic methods for assessing exposure to risk factors for work-related musculoskeletal disorders. *Occup Med (Lond)*. 2005;55:190–9.
8. Li G, Buckle P. Current techniques for assessing physical exposure to work-related musculoskeletal risks, with emphasis on posture-based methods. *Ergonomics*. 1999;42:674–95.
9. 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.
10. Stock SR, Fernandes R, Delisle A, Vézina N. Reproducibility and validity of workers' self-reports of physical work demands [review]. *Scand J Work Environ Health*. 2005;31(6):409–37.
11. Barriera-Viruet H, Sobeih TM. Questionnaires vs observational and direct measurements: a systematic review. *Theor Issues Ergon Sci*. 2006;7:261–84.
12. Winkel J, Mathiassen SE. Assessment of physical work load in epidemiological studies: concepts, issues and operational considerations. *Ergonomics*. 1994;37:979–88.
13. Balogh I, Orbaek P, Ohlsson K, Nordander C, Unge J, Winkel J, ET AL. Self-assessed and directly measured occupational physical activities – influence of musculoskeletal complaints, age and gender. *Appl Ergon*. 2004;35:49–56
14. Van der Beek AJ, Braam ITJ, Douwes M, Bongers PM, Fringsdresen MHW, Verbeek J, et al. Validity of a diary estimating exposure to tasks, activities, and postures of the trunk. *Int Arch Occup Environ Health*. 1994;66:173–8.
15. Tourangeau R, Rips LJ, Rasinski K. The psychology of survey response. New York (NY): Cambridge University Press; 2000.
16. Genaidy A, Karwowski W, Succop P, Kwon Y-G, Alhemedou A, Goyal D. A classification system for characterization of physical and non-physical work factors. *Int J Occup Saf Ergon*. 2000;6:535–55.
17. Wiktorin C, Vingard E, Mortimer M, Pernold G, Wigaeus-Hjelm E, Kilbom A, et al. Interview versus questionnaire for assessing physical loads in the population-based MUSIC-Norrtälje study. *Am J Ind Med*. 1999;35:441–55.
18. Leijon O, Wiktorin C, Harenstam A, Karlqvist L; MOA Research Group. Validity of a self-administered questionnaire for assessing physical work loads in a general population. *J Occup Environ Med*. 2002;44:724–35.
19. Winnemuller LL, Spielholz PO, Daniell WE, Kaufman JD. Comparison of ergonomist, supervisor, and worker assessments of work-related musculoskeletal risk factors. *J Occup Environ Hyg*. 2004;1:414–22.
20. Freedson PS, Miller K. Objective monitoring of physical activity using motion sensors and heart rate. *Res Q Exerc Sport*. 2000;71:21–9.
21. Rosner B. Fundamentals of biostatistics. Fifth edition. Pacific Grove (CA): Duxbury Thomson Learning; 2000. p 24–5.
22. Cai X, Ross S, Dimberg L. Can we trust the answers?: reliability and validity of ergonomic self-assessment surveys. *J Occup Environ Med*. 2007;49:1055–9.
23. Neumann WP, Wells RP, Norman RW, Andrews DM, Frank J, Shannon HS, et al. Comparison of four peak spinal loading exposure measurement methods and their association with low-back pain. *Scand J Work Environ Health*. 1999;25(5):404–9.
24. Wiktorin C, Karlqvist L, Winkel J. Validity of self-reported exposures to work postures and manual materials handling. *Scand J Work Environ Health*. 1993;19(3):208–14.
25. Hays M, Saurel-Cubizolles M-J, Bourguin M, Touranchet A, Verge C, Kaminski M. Conformity of workers' and occupational health physicians' descriptions of working conditions. *Int J Occup Environ Health*. 1996;2:10–7.
26. Heinrich J, Blatter BM, Bongers PM. A comparison of methods for the assessment of postural load and duration of computer use. *Occup Environ Med*. 2004;61:1027–31.
27. Homan MM, Armstrong TJ. Evaluation of three methodologies for assessing work activity during computer use. *AIHA J (Fairfax, VA)*. 2003;64:48–55.
28. Jensen OC, Sorensen JF, Kaerlev L, Canals ML, Nikolic N, Saarni H. Self-reported injuries among seafarers: questionnaire validity and results from an international study. *Accid Anal Prev*. 2004;36:405–13.
29. Nordstrom DL, Vierkant RA, Layde PM, Smith MJ. Comparisons of self-reported and expert-observed physical activities at work in a general population. *Am J Ind Med*. 1998;34:29–35.
30. Pope DP, Silman AJ, Cherry NM, Pritchard C, Macfarlane GJ. Validity of a self-completed questionnaire measuring the physical demands of work. *Scand J Work Environ Health*. 1998;24(5):376–85.
31. Somville PR, Van Nieuwenhuysse A, Seidel L, Masschelein R, Moens G, Mairiaux P. Validation of a self-administered questionnaire for assessing exposure to back pain mechanical risk factors. *Int Arch Occup Environ Health*. 2006;79:499–508.
32. Akesson I, Balogh I, Skerfving S. Self-reported and measured time of vibration exposure at ultrasonic scaling in dental hygienists. *Appl Ergon*. 2001;32:47–51.
33. Bao S, Howard N, Spielholz P, Silverstein B. Quantifying repetitive hand activity for epidemiological research on musculoskeletal disorders – part II: comparison of different methods of measuring force level and repetitiveness. *Ergonomics*. 2006;49:381–92.
34. Tiemessen IJH, Hulshof CTJ, Frings-Dresen MHW. Two way assessment of other physical work demands while measuring the whole body vibration magnitude. *J Sound Vib*. 2008;310:1080–92.
35. Lin CJ, Wang SJ, Chen HJ. A field evaluation method for assessing whole body biomechanical joint stress in manual lifting tasks. *Ind Health*. 2006;44:604–12.

36. 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;19:445–55.
37. Hansson GA, Balogh I, Bystrom 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.
38. Kallio M, Viikari-Juntura E, Hakkanen M, Takala EP. Assessment of duration and frequency of work tasks by telephone interview: reproducibility and validity. *Ergonomics.* 2000;43:610–21.
39. Laperriere E, Messing K, Couture V, Stock S. Validation of questions on working posture among those who stand during most of the work day. *Int J Ind Ergon.* 2005;35:371–8.
40. Unge J, Hansson GA, Ohlsson K, Nordander C, Axmon A, Winkel J, et al. Validity of self-assessed reports of occurrence and duration of occupational tasks. *Ergonomics.* 2005;48:12–24.
41. Karlqvist LK, Hagberg M, Koster M, Wenemark M, Nell 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.
42. Viikari-Juntura E, Rauas S, Martikainen R, Kuosma E, Riihimaki H, Takala EP, et al. Validity of self-reported physical work load in epidemiologic studies on musculoskeletal disorders. *Scand J Work Environ Health.* 1996;22(4):251–9.
43. Wiktorin C, Selin K, Ekenvall L, Alfredsson L. An interview technique for recording work postures in epidemiological studies. *Int J Epidemiol.* 1996;25:171–80.
44. Andrews DM. Comparison of self-report and observer methods for repetitive posture and load assessment. *Occup Ergon.* 1998;1:211–22.
45. Faucett J, Rempel D. Musculoskeletal symptoms related to display terminal use – an analysis of objective and subjective exposure estimates. *AAOHN J.* 1996;44:33–9.
46. Douwes M, Kraker Hd, Blatter BM. Validity of two methods to assess computer use: self-report by questionnaire and computer use software. *Int J Ind Ergon.* 2007;37:425–31.
47. Wigaeus Hjelm E, Winkel J, Nygard C-H, Wiktorin C, Karlqvist L; Stockholm MUSIC 1 Study Group. Can cardiovascular load in ergonomic epidemiology be estimated by self-report? *J Occup Environ Med.* 1995;37:1210–7.
48. Burdorf A, Laan J. Comparison of methods for the assessment of postural load on the back. *Scand J Work Environ Health.* 1991;17(6):425–9.
49. Spielholz P, Silverstein B, Morgan M, Checkoway H, Kaufman J. Comparison of self-report, video observation and direct measurement methods for upper extremity musculoskeletal disorder physical risk factors. *Ergonomics.* 2001;44:588–613.
50. Lindegård A, Karlberg C, Wigaeus Tornqvist E, Toomingas A, Hagberg M. Concordance between VDU-users' ratings of comfort and perceived exertion with experts' observations of workplace layout and working postures. *Appl Ergon.* 2005;36:319–25.
51. Martin F, Matthias P. Factors associated with the subject's ability to quantify their lumbar flexion demands at work. *Int J Environ Health Res.* 2006;16:69–79.
52. Sandmark H, Wiktorin C, Hogstedt C, Klenell-Hatschek EK, Vingard E. Physical work load in physical education teachers. *Appl Ergon.* 1999;30:435–42.
53. Mortimer M, Hjelm EW, Wiktorin C, Pernold G, Kilbom A, Vingard E. Validity of self-reported duration of work postures obtained by interview. *Appl Ergon.* 1999;30:477–86.
54. Palmer KT, Haward B, Griffin MJ, Bendall H, Coggon D. Validity of self reported occupational exposures to hand transmitted and whole body vibration. *Occup Environ Med.* 2000;57:237–41.
55. Pernold G, Tornqvist EW, Wiktorin C, Mortimer M, Karlsson E, Kilbom A, et al. Validity of occupational energy expenditure assessed by interview. *AIHA J (Fairfax, VA).* 2002;63:29–33.
56. Kilbom A. Assessment of physical exposure in relation to work-related musculoskeletal disorders – what Information can be obtained from systematic observations [review]. *Scand J Work Environ Health.* 1994;20, special issue:30–45.
57. Van der Beek AJ, van Gaalen LC, Fring-Dresen MHW. Working postures and activities of lorry drivers: a reliability study of on-site observations and recording on a pocket computer. *Appl Ergon.* 1992;23:331–6.
58. Barrero LH, Katz JN, Perry MJ, Krishnan R, Ware JH, Dennerlein JT. Work pattern causes bias in self-reported activity duration: randomized study of mechanisms and implications for exposure assessment and epidemiology. *Occup Environ Med.* 2009;66:38–44.
59. Stewart P. Challenges to retrospective exposure assessment. *Scand J Work Environ Health.* 1999;25(6):505–10.
60. Borg G. Perceived exertion as an indicator of somatic stress. *Scand J Rehab Med.* 1970;2:92–8.
61. Gamberale F. Perceived exertion, heart rate, oxygen uptake, and blood lactate in different work operations. *Ergonomics.* 1972;15:545–54.
62. Kumar S, Simmonds M. The accuracy of magnitude production of submaximal precision and power grips and gross motor efforts. *Ergonomics.* 1994;37:1345–53.
63. Gamberale F. Perception of effort in manual materials handling [review]. *Scand J Work Environ Health.* 1990;16 suppl 1:59–66.
64. Petersson NF, Mathiassen SE, Björing G, Winkel J. The accuracy of self-rating of exposure to repetitive work. *Int J Ind Ergon.* 2000;25:239–46.
65. Chatterjee N, Wacholder S. Validation studies: bias, efficiency and exposure assessment. *Epidemiology.* 2002;13:503–6.

Received for publication: 22 December 2008

**Appendix.** Inventory of validity assessments of single question for ergonomic demands evaluation. [Dur = duration; Mag = magnitude; Freq = frequency; Dir Meas = direct measurement; Mixed population = blue- and white-collar workers; Dist Healthcare = distribution in healthcare; Forestry & agri = forestry & agriculture, Warehousing & distr = warehousing & distribution]

Exposure	Dimension	Self-report	Criterion	Author	Industry	Population	Sample			
Posture	Magnitude	Self-administered	Observations	Andrews et al, 1997 (1)	Auto assembly	Blue-collar	99			
				Cai et al, 2007 (2)	Banking	White-collar	111			
				Hays et al, 1996 (3)	Food processing	..	497			
				Heinrich et al, 2004 (4)	Administration	White-collar	87			
				Laperriere et al, 2005 (5)	Mixed	White-collar	77			
				Pope et al, 1998 (6)	Mixed	Mixed	123			
				Dur & Mag	Dir Meas	Dir Meas	Hansson et al, 2001 (7)	Mixed	Mixed	71–78
							Spielholz et al, 2001 (8)	Mixed	Blue-collar	18
							Torgen et al, 1999 (9)	Mixed	Mixed	75
							Wiktorin et al, 1993 (10)	Mixed	Mixed	97
	Observations	Andrews et al, 1998 (11)	Auto assembly				Blue-collar	163		
		Burdorf & Laan, 1991 (12)	Steel manufacturing				Blue-collar	35		
		Jensen et al, 2004 (13)	Construction				Blue-collar	71		
		Pope et al, 1998 (6)	Mixed				Mixed	17–123		
		Sandmark et al, 1999 (14)	Education				White-collar	30		
		Somville et al, 2006 (15)	Dist healthcare				Mixed	147		
		Spielholz et al, 2001 (8)	Mixed				Blue-collar	18		
		Tiemessen et al, 2008 (16)	Transportation				Blue-collar	10		
	Torgén et al, 1999 (9)	Mixed	Mixed	40–75						
	Viikari-Juntura et al, 1996 (17)	Forestry & agric	Blue-collar	36						
	Wiktorin et al, 1993 (10)	Mixed	Mixed	97						
	Winnemuller et al, 2004 (18)	Mixed	Mixed	55						
	Other self-report	Leijon et al, 2002 (19)	Other self-report	Mixed	Mixed	202				
				Wiktorin et al, 1999 (20)	Mixed	Mixed	358–1421			
				Interview	Dir Meas	Martin & Matthias, 2006 (21)	Forestry & agric	Multiple	13–14	
						Mortimer et al, 1999 (22)	Mixed	Mixed	16–19	
	Observations	Mortimer et al, 1999 (22)	Observations	Mixed	Mixed	20				
				Wiktorin et al, 1996 (23)	Construction	Blue-collar	57			
	Diaries/logs	Burdorf & Laan, 1991 (12)	Diaries/logs	Steel manufacturing	Blue-collar	35				
				van der Beek et al, 1994 (24)	Transportation	Multiple	16–116			
				Viikari-Juntura et al, 1996 (17)	Forestry & agric	Blue-collar	36			
				Tiemessen et al, 2008 (16)	Transportation	Blue-collar	10			
Freq & Mag Dur, Freq & Mag	Self-administered	Other self-report	Torgén et al, 1999 (9)	Mixed	Mixed	69				
			Wiktorin et al, 1999 (20)	Mixed	Mixed	358–1421				
Movement	Magnitude	Observations	Pope et al, 1998 (6)	Mixed	Mixed	38–123				
			Sandmark et al, 1999 (14)	Education	White-collar	30				
			Somville et al, 2006 (15)	Dist healthcare	Mixed	76–147				
			Interview	Nordstrom et al, 1998 (25)	Mixed	Mixed	Mixed	28–33		
						Self-administered	Laperriere et al, 2005 (5)	Mixed	White-collar	85
	Frequency	Self-administered	Self-administered	Somville et al, 2006 (15)	Dist healthcare	Mixed	147			
				Dur & Mag	Burdorf & Laan, 1991 (12)	Steel manufacturing	Blue-collar	35		
	Pope et al, 1998 (6)	Mixed	Mixed				3–64			
	Tiemessen et al, 2008 (16)	Transportation	Blue-collar				10			
	Wiktorin et al, 1993 (10)	Mixed	Mixed				97			
	Interview	Nordstrom et al, 1998 (25)	Mixed	Mixed	Mixed	28–33				
				Diaries/logs	van der Beek et al, 1994 (24)	Transportation	Multiple	16–116		
	Freq & Mag	Self-administered	Self-administered				Andrews et al, 1998 (11)	Auto assembly	Blue-collar	155–166
				Pope et al, 1998 (6)	Mixed	Mixed	35–73			
				Sandmark et al, 1999 (14)	Education	White-collar	30			
				Somville et al, 2006 (15)	Dist healthcare	Mixed	147			
				Viikari-Juntura et al, 1996 (17)	Forestry & agric	Blue-collar	36			

(continued)

## Appendix. Continued.

Exposure	Dimension	Self-report	Criterion	Author	Industry	Population	Sample			
Movement	<i>cont</i>			Wiktorin et al, 1993 (10)	Mixed	Mixed	97			
				Winnemuller et al, 2004 (18)	Mixed	Mixed	55			
				Other self-report	Burdorf & Laan, 1991 (12)	Steel manufacturing	Blue-collar	35		
				Leijon et al, 2002 (19)	Mixed	Mixed	202			
				Diaries/logs	Observations	van der Beek et al, 1994 (24)	Transportation	Multiple	16–63	
				Viikari-Juntura et al, 1996 (17)	Forestry & agric	Blue-collar	28			
	Dur, Freq & Mag	Self-administered	Winnemuller et al, 2004 (18)	Mixed	Mixed	55				
Physical exertion	Magnitude		Dir Meas	Torgén et al, 1999 (9)	Mixed	Mixed	78			
				Wigaeus Hjelm et al, 1995 (26)	Mixed	Mixed	91			
				Observations	Lindegård et al, 2005 (27)	Mixed	White-collar	746–787		
				Somville et al, 2006 (15)	Dist healthcare	Mixed	147			
				Other self-report	Leijon et al, 2002 (19)	Mixed	Mixed	202		
				Wigaeus Hjelm et al, 1995 (26)	Mixed	Mixed	35–52			
				Other	Lin et al, 2006 (28)	Warehousing & dist	Blue-collar	31		
				Other	Dir Meas	Pernold et al, 2002 (29)	Mixed	Mixed	27	
		Wigaeus Hjelm et al, 1995 (26)	Mixed	Mixed	85					
Repetitions		Self-administered	Observations	Spielholz et al, 2001 (8)	Mixed	Blue-collar	18			
				Hays et al, 1996 (3)	Food processing	..	496			
				Spielholz et al, 2001 (8)	Mixed	Blue-collar	18			
				Interview	Nordstrom et al, 1998 (25)	Mixed	Mixed	28–33		
				Dur & Mag	Self-administered	Dir Meas	Hansson et al, 2001 (7)	Mixed	Mixed	77–80
				Spielholz et al, 2001 (8)	Mixed	Blue-collar	18			
				Observations	Pope et al, 1998 (6)	Mixed	Mixed	83–123		
				Spielholz et al, 2001 (8)	Mixed	Blue-collar	18			
				Torgén et al, 1999 (9)	Mixed	Mixed	63			
				Viikari-Juntura et al, 1996 (17)	Forestry & agric	Blue-collar	36			
				Winnemuller et al, 2004 (18)	Mixed	Mixed	55			
				Other self-report	Leijon et al, 2002 (19)	Mixed	Mixed	202		
				Interview	Observations	Nordstrom et al, 1998 (25)	Mixed	Mixed	28–33	
				Freq & Mag	Self-administered	Torgén et al, 1999 (9)	Mixed	Mixed	63	
Dur, Freq & Mag		Hays et al, 1996 (3)	Food processing	..	499					
		Other self-report	Heinrich et al, 2004 (4)	Administration	White-collar	87				
Force	Magnitude		Dir Meas	Bao et al, 2006 (30)	Mixed	..	402			
				Pope et al, 1998 (6)	Department Store	Blue-collar	20			
				Observations	Andrews et al, 1997 (1)	Auto assembly	Blue-collar	99		
				Neumann et al, 1999 (31)	Auto assembly	Blue-collar	142			
				Spielholz et al, 2001 (8)	Mixed	Blue-collar	18			
				Other self-report	Bao et al, 2006 (30)	Mixed	..	402		
				Burdorf & Laan, 1991 (12)	Steel manufacturing	Blue-collar	35			
				Interview	Observations	Nordstrom et al, 1998 (25)	Mixed	Mixed	28–33	
				Other	Andrews et al, 1997 (1)	Auto assembly	Blue-collar	99		
				Dur & Mag	Self-administered	Dir Meas	Spielholz et al, 2001 (8)	Mixed	Blue-collar	18
				Observations	Winnemuller et al, 2004 (18)	Mixed	Mixed	55		
				Interview	Nordstrom et al, 1998 (25)	Mixed	Mixed	28–33		
				Vibration	Magnitude	Self-administered	Other	Palmer et al, 2000 (32)	Mixed	Blue-collar
Palmer et al, 2000 (32)	Mixed	Blue-collar	179							
Interview	Observations	Nordstrom et al, 1998 (25)	Mixed					Mixed	28–33	
Dur & Mag	Self-administered	Palmer et al, 2000 (32)	Mixed					Blue-collar	179	
Somville et al, 2006 (15)	Dist healthcare	Mixed	147							
Winnemuller et al, 2004 (18)	Mixed	Mixed	55							
Other self-report	Wiktorin et al, 1999 (20)	Mixed	Mixed					358–1421		
Interview	Dir Meas	Akesson et al, 2001 (33)	Health services					White-collar	10	
									(continued)	

## Appendix. Continued.

Exposure	Dimension	Self-report	Criterion	Author	Industry	Population	Sample			
Vibration	<i>cont</i>		Observations	Nordstrom et al, 1998 (25)	Mixed	Mixed	28–33			
		Diaries/logs	Dir. Meas.	Akesson et al, 2001 (33)	Health services	White-collar	10			
			Observations	van der Beek et al, 1994 (24)	Transportation	Blue-collar	35			
Activity/task	Duration	Self-administered	Dir Meas	Heinrich et al, 2004 (4)	Administration	White-collar	87			
				Homan et al, 2003 (34)	Mixed	Mixed	51			
				Observations	Burdorf & Laan, 1991 (12)	Steel manufacturing	Blue-collar	35		
				Douwes et al, 2007 (35)	Banking	White-collar	83			
				Faucett & Rempel, 1996 (36)	Newspaper	White-collar	13			
				Homan et al, 2003 (34)	Mixed	Mixed	51			
				Tiemessen et al, 2008 (16)	Transportation	Blue-collar	10			
				Other self-report	Wiktorin et al, 1999 (20)	Mixed	Mixed	358–1421		
				Interview	Observations	Pernold et al, 2002 (29)	Mixed	Mixed	110	
					Other self-report	Kallio et al, 2000 (37)	Forestry & agric	Blue-collar	13	
					Diaries/logs	Observations	Burdorf & Laan, 1991 (12)	Steel manufacturing	Blue-collar	35
				Unge et al, 2005 (38)		Mixed	Multiple	20–22		
				van der Beek et al, 1994 (24)		Transportation	Multiple	16–35		
				Magnitude	Self-administered	Dir Meas	Sandmark et al, 1999 (14)	Education	White-collar	30
							Viikari-Juntura et al, 1996 (17)	Forestry & agric	Blue-collar	36
Wiktorin et al, 1993 (10)	Mixed	Mixed	97							
Other				Karlqvist et al, 1996 (39)	Mixed	White-collar	100			
				Observations	Cai et al, 2007 (2)	Banking	White-collar	111		
				Hays et al, 1996 (3)	Food processing	..	450			
				Heinrich et al, 2004 (4)	Administration	White-collar	87			
				Other self-report	Heinrich et al, 2004 (4)	Administration	White-collar	87		
				Interview	Observations	Nordstrom et al, 1998 (25)	Mixed	Mixed	28–33	
					Diaries/logs	Unge et al, 2005 (38)	Mixed	Multiple	20–22	
				Frequency	Self-administered		Winnemuller et al, 2004 (18)	Mixed	Mixed	55
				Dur & Mag			Interview	Nordstrom et al, 1998 (25)	Mixed	Mixed

## References

1. 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;19:445–55.
2. Cai X, Ross S, Dimberg L. Can we trust the answers?: reliability and validity of ergonomic self-assessment surveys. *J Occup Environ Med.* 2007;49:1055–9.
3. Hays M, Saurel-Cubizolles M-J, Bourguine M, Touranchet A, Verge C, Kaminski M. Conformity of workers' and occupational health physicians' descriptions of working conditions. *Int J Occup Environ Health.* 1996;2:10–7.
4. Heinrich J, Blatter BM, Bongers PM. A comparison of methods for the assessment of postural load and duration of computer use. *Occup Environ Med.* 2004;61:1027–31.
5. Laperriere E, Messing K, Couture V, Stock S. Validation of questions on working posture among those who stand during most of the work day. *Int J Ind Ergon.* 2005;35:371–8.
6. Pope DP, Silman AJ, Cherry NM, Pritchard C, Macfarlane GJ. Validity of a self-completed questionnaire measuring the physical demands of work. *Scand J Work Environ Health.* 1998;24(5):376–85.
7. Hansson GA, Balogh I, Bystrom 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.
8. Spielholz P, Silverstein B, Morgan M, Checkoway H, Kaufman J. Comparison of self-report, video observation and direct measurement methods for upper extremity musculoskeletal disorder physical risk factors. *Ergonomics.* 2001;44:588–613.
9. Torgén M, Winkel J, Alfredsson L, Kilbom Å, Stockholm MUSIC 1 study group. Evaluation of questionnaire-based information on previous physical work loads. *Scand J Work Environ Health.* 1999;25(3):246–54.
10. Wiktorin C, Karlqvist L, Winkel J. Validity of self-reported exposures to work postures and manual materials handling. *Scand J Work Environ Health.* 1993;19(3):208–14.
11. Andrews DM. Comparison of self-report and observer methods for repetitive posture and load assessment. *Occup Ergon.* 1998;1:211–22.

12. Burdorf A, Laan J. Comparison of methods for the assessment of postural load on the back. *Scand J Work Environ Health*. 1991;17(6):425–9.
13. Jensen OC, Sorensen JF, Kaerlev L, Canals ML, Nikolic N, Saarni H. Self-reported injuries among seafarers: questionnaire validity and results from an international study. *Accid Anal Prev*. 2004;36:405–13
14. Sandmark H, Wiktorin C, Hogstedt C, Klenell-Hatschek EK, Vingard E. Physical work load in physical education teachers. *Appl Ergon*. 1999;30:435–42.
15. Somville PR, Van Nieuwenhuysse A, Seidel L, Masschelein R, Moens G, Mairiaux P. Validation of a self-administered questionnaire for assessing exposure to back pain mechanical risk factors. *Int Arch Occup Environ Health*. 2006;79:499–508.
16. Tiemessen IJH, Hulshof CTJ, Frings-Dresen MHW. Two way assessment of other physical work demands while measuring the whole body vibration magnitude. *J Sound Vib*. 2008;310:1080–92.
17. Viikari-Juntura E, Rauas S, Martikainen R, Kuosma E, Riihimaki H, Takala EP, et al Validity of self-reported physical work load in epidemiologic studies on musculoskeletal disorders. *Scand J Work Environ Health*. 1996;22(4):251–9.
18. Winnemuller LL, Spielholz PO, Daniell WE, Kaufman JD. Comparison of ergonomist, supervisor, and worker assessments of work-related musculoskeletal risk factors. *J Occup Environ Hyg*. 2004;1:414–22.
19. Leijon O, Wiktorin C, Harenstam A, Karlqvist L; MOA Research Group. Validity of a self-administered questionnaire for assessing physical work loads in a general population. *J Occup Environ Med*. 2002;44:724–35.
20. Wiktorin C, Vingard E, Mortimer M, Pernold G, Wigaeus-Hjelm E, Kilbom A, et al. Interview versus questionnaire for assessing physical loads in the population-based MUSIC-Norrtalje study. *Am J Ind Med*. 1999;35:441–55.
21. Martin F, Matthias P. Factors associated with the subject's ability to quantify their lumbar flexion demands at work. *Int J Environ Health Res*. 2006;16(1):69–79.
22. Mortimer M, Hjelm EW, Wiktorin C, Pernold G, Kilbom A, Vingard E. Validity of self-reported duration of work postures obtained by interview. *Appl Ergon*. 1999;30:477–86.
23. Wiktorin C, Selin K, Ekenvall L, Alfredsson L. An interview technique for recording work postures in epidemiological studies. *Int J Epidemiol*. 1996;25:171–80.
24. Van der Beek AJ, Braam ITJ, Douwes M, Bongers PM, Fringsdresen MHW, Verbeek J, et al. Validity of a diary estimating exposure to tasks, activities, and postures of the trunk. *Int Arch Occup Environ Health*. 1994;66:173–8.
25. Nordstrom DL, Vierkant RA, Layde PM, Smith MJ. Comparisons of self-reported and expert-observed physical activities at work in a general population. *Am J Ind Med*. 1998;34:29–35.
26. Wigaeus-Hjelm E, Winkel J, Nygård C-H, Wiktorin C, Karlqvist L; Stockholm MUSIC 1 Study Group. Can cardiovascular load in ergonomic epidemiology be estimated by self-report? *J Occup Env Med*. 1995;37:1210–7.
27. Lindegård A, Karlberg C, Wigaeus Tornqvist E, Toomingas A, Hagberg M. Concordance between VDU-users' ratings of comfort and perceived exertion with experts' observations of workplace layout and working postures. *Appl Ergon*. 2005;36:319–25.
28. Lin CJ, Wang SJ, Chen HJ. A field evaluation method for assessing whole body biomechanical joint stress in manual lifting tasks. *Ind Health*. 2006;44:604–12.
29. Pernold G, Tornqvist EW, Wiktorin C, Mortimer M, Karlsson E, Kilbom A, et al. Validity of occupational energy expenditure assessed by interview. *AIHA J (Fairfax, VA)*. 2002;63:29–33.
30. Bao S, Howard N, Spielholz P, Silverstein B. Quantifying repetitive hand activity for epidemiological research on musculoskeletal disorders – part II: comparison of different methods of measuring force level and repetitiveness. *Ergonomics*. 2006;49:381–92.
31. Neumann WP, Wells RP, Norman RW, Andrews DM, Frank J, Shannon HS, et al. Comparison of four peak spinal loading exposure measurement methods and their association with low-back pain. *Scand J Work Environ Health*. 1999;25(5):404–9.
32. Palmer KT, Haward B, Griffin MJ, Bendall H, Coggon D. Validity of self reported occupational exposures to hand transmitted and whole body vibration. *Occup Environ Med*. 2000;57:237–41.
33. Akesson I, Balogh I, Skerfving S. Self-reported and measured time of vibration exposure at ultrasonic scaling in dental hygienists. *Appl Ergon*. 2001;32:47–51.
34. Homan MM, Armstrong TJ. Evaluation of three methodologies for assessing work activity during computer use. *AIHA J (Fairfax, VA)*. 2003;64:48–55.
35. Douwes M, Kraker Hd, Blatter BM. Validity of two methods to assess computer use: self-report by questionnaire and computer use software. *Int J Ind Ergon*. 2007;37:425–31.
36. Faucett J, Rempel D. Musculoskeletal symptoms related to display terminal use – an analysis of objective and subjective exposure estimates. *AAOHN J*. 1996;44:33–9.
37. Kallio M, Viikari-Juntura E, Hakkanen M, Takala EP. Assessment of duration and frequency of work tasks by telephone interview: reproducibility and validity. *Ergonomics*. 2000;43:610–21.
38. Unge J, Hansson GA, Ohlsson K, Nordander C, Axmon A, Winkel J, et al Validity of self-assessed reports of occurrence and duration of occupational tasks. *Ergonomics*. 2005;48:12–24.
39. Karlqvist LK, Hagberg M, Koster M, Wenemark M, Nell 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.