Review

Scand J Work Environ Health 2010;36(1):3–24

Systematic evaluation of observational methods assessing biomechanical exposures at work

by Esa-Pekka Takala, PhD,[1] Irmeli Pehkonen, MSc,[1] Mikael Forsman, PhD,[2] Gert-Åke Hansson, PhD,[3] Svend Erik Mathiassen, PhD,[4] W Patrick Neumann, PhD,[5] Gisela Sjøgaard, PhD,[6] Kaj Bo Veiersted, PhD,[7] Rolf H Westgaard, PhD,[8] Jørgen Winkel, PhD [9]

Takala E-P, Pehkonen I, Forsman M, Hansson G-Å, Mathiassen SE, Neumann WP, Sjøgaard G, Veiersted KB, Westgaard RH, Winkel J. Systematic evaluation of observational methods assessing biomechanical exposures at work. *Scand J Work Environ Health*. 2010;36(1):3–24.

Objectives This systematic review aimed to identify published observational methods assessing biomechanical exposures in occupational settings and evaluate them with reference to the needs of different users.

Methods We searched scientific databases and the internet for material from 1965 to September 2008. Methods were included if they were primarily based on the systematic observation of work, the observation target was the human body, and the method was clearly described in the literature. A systematic evaluation procedure was developed to assess concurrent and predictive validity, repeatability, and aspects related to utility. At least two evaluators independently carried out this evaluation.

Results We identified 30 eligible observational methods. Of these, 19 had been compared with some other method(s), varying from expert evaluation to data obtained from video recordings or through the use of technical instruments. Generally, the observations showed moderate-to-good agreement with the corresponding assessments made from video recordings; agreement was the best for large-scale body postures and work actions. Postures of wrist and hand as well as trunk rotation seemed to be more difficult to observe correctly. Intra- and inter-observer repeatability were reported for 7 and 17 methods, respectively, and were judged mostly to be moderate or good.

Conclusions With training, observers can reach consistent results on clearly visible body postures and work activities. Many observational tools exist, but none evaluated in this study appeared to be generally superior. When selecting a method, users should define their needs and assess how results will influence decision-making.

Key terms posture; review; risk assessment; workload.

Observational methods are probably the most often used approach to evaluate physical workload in order to identify hazards at work, monitor the effects of ergonomic changes, and conduct research on these issues. The number of available methods is large, but no single one is suitable for all purposes – different approaches are needed for different goals. Due to differences in methods and diversity in user needs, the selection of an appropriate tool can be challenging.

The selection of a method should be based on (i) the objectives of its use, (ii) the characteristics of the work to be assessed, (iii) the individual(s) who will use the method, and (iv) the resources available for collecting and analyzing data. In epidemiological research, the

1 Finnish Institute of Occupational Health, Helsinki, Finland.
2 Karolinska Institutet, Stockholm, Sweden.
3 Occupational and Environmental Medicine, Lund University, Lund, Sweden.
4 Centre for Musculoskeletal Research, University of Gävle, Sweden.
5 Ryerson University, Toronto, Canada.
6 University of Southern Denmark, Odense, Denmark.
7 National Institute of Occupational Health, Oslo, Norway.
8 Norwegian University of Science and Technology, Trondheim, Norway.
9 University of Gothenburg, Gothenburg, Sweden and the National Research Centre for the Working Environment, Copenhagen, Denmark.

Correspondence to: Dr E-P Takala, Finnish Institute of Occupational Health, Topeliuksenkatu 41, FI-00250 Helsinki, Finland. [E-mail: esa-pekka.takala@ttl.fi]

appropriate methods for studying health hazards usually differ from those more suitable for the assessment of the needs for change at a single workplace, the evaluation of the effects of specific ergonomic improvements, or the study of the usability of hand tools. Official regulations may make explicit demands on the assessment, for instance in safety inspections, or when evaluating the workload in order to make compensability decisions for injured workers.

Potential users rarely know about more than a very limited selection of methods. There are several reasons for this knowledge gap. The reports describing methods may be in a language that is unfamiliar to the user, or they may have been published in unknown media. Methods may have been developed for specific occupations only and, consequently, not be suitable for the setting that the user needs to address.

In order for observational data to provide a sound basis for decision-making, the assessment should be valid for the targeted purpose and the result should be reproducible. An ideal way to assess validity is to compare results with a "gold standard". There is, however, no general "gold standard" for assessing biomechanical exposures, even though postures can be accurately measured with direct technical measurements. Another aspect of validity – the ability of the method to predict risks – can be studied by analyzing the associations between exposures obtained by the method and the outcomes of interest, such as musculoskeletal disorders (MSD). Assessment of reproducibility should cover both intra- and inter-observer repeatability (1, 2) and, in the case of off-line observation using video recordings, the possible errors and variances associated with the filming procedure itself.

Several aspects of observational methods for assessing physical workload have been reviewed earlier (3–8). However, we could not find an up-to-date systematic and critical comparison of methods devoted to guiding users in selecting appropriate tools for different purposes.

Thus, the aims of this project were systematically to identify published observational methods for assessing physical workload (biomechanical exposures) and to evaluate critically these methods from the perspective of different users, such as researchers, occupational health and safety personnel, safety inspectors, ergonomists, and work-system designers.

Methods

Search and selection of reference literature

Literature searches were conducted in the following electronic databases: PubMed, Embase, CISDOC, and ScienceDirect. We used Google and Google Scholar to identify sources via the internet. The searches were conducted covering material from 1965 to September 2008.

The searches were made using several combinations of the following search terms related to observational methods (using "OR"): observation, workload, lifting, manual material handling, risk assessment, task analysis, posture, ergonomic, and occupational exposure. These terms were combined (using "AND") with terms related to the musculoskeletal system (using "OR"): musculoskeletal, back, neck, extremities (eg, [observation OR workload] AND [back OR neck]). A search strategy using only these terms was ineffective in identifying several known methods and, therefore, additional searches were performed with the names of the identified methods and using the option of "related articles" of the key references. The reference lists of key papers were also scanned to identify additional references.

We screened the articles by title and abstract first. About 580 potential references were identified, including original scientific reports, reviews, and internet sources. Full texts of these references were collated in electronic format for further evaluation. The inclusion criteria were as follows: (i) systematic observation of work should be the principal exposure assessment tool; (ii) the method should be described in a manner allowing the procedure to be reproduced; and (iii) the observation target should be the human locomotor system (eg, back/trunk, neck, or extremities). Many original articles describing a method used only in one specific study were discarded because the tool was too inadequately described to be reproduced and evaluated. Only methods that were publically available in scientific or other reports or common textbooks were included; this excluded commercial products lacking a detailed public description. We also excluded methods that were not developed for visual observations in occupational field settings, such as those based entirely on the measurements of postural angles from video recordings.

Developing the framework for evaluation

There is no generally accepted procedure to evaluate methods for the assessment of workload even though several previous reviews have addressed this issue (3–9). In meetings to discuss our review, we developed the structure and contents of the evaluation procedure in an iterative manner. The procedure included items describing the basic features of the methods, as well as assessments of their validity and repeatability and practical issues for the users.

Validity assessment included concurrent validity (ie, how well does the method correspond with more valid methods?) and predictive validity (ie, how well have risk estimates generated by the method been shown to

be associated with MSD?) (1, 2). In addition, we evaluated the intra- and inter-observer repeatability. We also assessed face validity (ie, does the method appear to measure what it is intended to measure?) to help potential users, especially in cases where no formal studies on validity have been performed.

For each individual report describing the validity of a method, we evaluated the following questions related to *concurrent validity*: (i) "Is the reference method valid for comparison?" (ii) "What is the quality of the comparison?" and (iii) "Are results generated by the method valid?" The rating was done on a 4-step scale (ranging from "perfect/almost perfect" to "major error/mistake") with comments justifying our rating. We evaluated the *predictive validity* of a method in cross-sectional and longitudinal trials, if available, with four options ("yes"; "no"; "conflicting results"; "cannot be estimated from documents").

In assessing the *repeatability* of each method, we considered the results and their interpretation – as given by the authors of the report – and rated the methods on the following scale: "probably reliable"; "potential error"; "obvious error"; "cannot be estimated".

Finally, on the basis of all available information on concurrent validity and repeatability, we rated each method overall as "good", "moderate", or "low" relative to these performance aspects.

We evaluated *face validity* with the following questions: (i) "Is the content of the method such that a relevant assessment can be expected?" (ii) "Do the items to be observed have a sound basis?" (iii) "Is the operationalization of the items to be observed sound?" (iv) "Is the process of data collection and analysis sound?" (v) "Can the output help in decision-making?" In addition, we evaluated the tool's strengths and limitations, and the potential user groups thereof.

Evaluation

Two researchers from our group read the selected publications and independently completed the basic description and documentation of all the methods in the evaluation form. After that, they discussed any differences and reached a consensus on the written documentation.

Based on this documentation and the original articles, each method was evaluated in terms of validity, repeatability, and practical issues independently by at least two evaluators blinded to each other. The original two researchers evaluated all identified methods; a third member of our group evaluated 14 additional methods. Discrepancies were resolved by discussion amongst the evaluators in order to establish consensus. If no consensus was reached, an additional evaluator was prepared to participate in the discussion, in accordance with the predetermined protocol. This option was not needed in any case.

Results

A total of 30 eligible observational methods were identified. These 30 methods were, altogether, described and referenced in 285 documents, ranging from 1–50 references for a particular method. Additional methods referred to in 70 papers were excluded, mainly due to insufficient information on the particular method. In the following description, methods are classified in three groups according to whether the main focus was to assess (i) general workload, (ii) upper-limb activities, or (iii) manual material handling.

The methods are presented in chronological order of appearance in the literature, since newer methods generally share some features with older ones. Table 1 summarizes the basic characteristics of the methods, table 2 shows the evaluation of validity and repeatability, and table 3 describes our opinions related to practical issues for the user(s) of the method. A descriptive overview of findings for each tool is provided below. Detailed information of the evaluated methods with full references can be found in a website associated with the present project (10).

Methods to assess general workload

Ovako working posture assessment system (OWAS). OWAS was developed in a steel industry company to describe workloads during the overhauling of iron smelting ovens (11). Aspects to be observed include the weight of the load handled (three categories) and postures of the back (four postures), arms (three postures), and lower extremities (seven postures), resulting in 252 possible combinations, which have been classified to four action categories indicating a need for ergonomic change. The observations were made as "snapshots" and sampling has usually been carried out using fixed-time intervals. OWAS ratings of postures have been well associated with perceived loading and discomfort (12, 13). For time spent in bent postures, agreement between OWAS and direct technical measurements has been rather low (14), which may partly be explained by differences in sampling strategies between the methods. In an evaluation of lifting situations, OWAS results were clearly different from those obtained by the NIOSH lifting equation (see p13–14), probably due to the different basic approaches of these two methods (15). Associations between OWAS ratings and the occurrence of back disorders have been reported in cross-sectional studies (16). The method has shown good intra- (17, 18) and inter-observer repeatability (11, 17, 19, 20).

Arbeitswissenschaftliches erhebungsverfahren zur tätigkeitsanalyse [(AET) ergonomic job analysis procedure]. AET is a job and stress analysis procedure offering a broad-spectrum description of work characteristics.

Table 1. Description of observational methods. Exposures included in the method: posture (P), force (F), duration (D), frequency of actions (Fr), movements (M), and vibration (Vib). (RPE=rating of perceived exertion; NIOSH=National Institute of Occupational Safety and Health; VAS=visual analog scale; TLV=threshold limit value; MMH=manual material handling)

Method and year of first publication	Target exposures and dimensions	Metrics	Observation strategy	Mode of recording
General methods				
Ovako working posture assessment system (OWAS), 1973	P, F	Frequency of items	Time sampling	Pen & paper, computerized
Arbeitswissenschaftliches erhebungsverfahren zur tätigkeitsanalyse [(AET) ergonomic job analysis procedure], 1979	P, F, Fr, Vib	Profile of items	No detailed rules	Pen & paper
Posture targeting, 1979	P	Frequency of postures	No detailed rules	Pen & paper
Ergonomic analysis (ERGAN), 1982	-	Borg RPE scale	No detailed rules	Video, computerized
Task recording and analysis on computer (TRAC), 1992	P, F, D, Fr,	Distribution/duration of observed items	Time sampling/ continuous observations	Computerized
Portable ergonomic observation (PEO), 1994	P, F, D, Fr, M	Start/end of postures	Continuous observations	Computerized, video
Hands relative to the body (HARBO), 1995	P	Start/end of postures	Continuous observations	Computerized, video
Plan för identifiering av belastningsfaktorer [(PLIBEL) a method assigned for the identification of ergonomics hazards], 1995	P, F, Fr, M	Yes/no answers; profile of items	Selection by general knowledge of work and observations	Pen & paper
Posture, activity, tools and handling (PATH), 1996	P, F, work activity	Time spent in postures	Time sampling	Pen & paper, (video) computerized
Quick exposure check (QEC), 1999	P, F, D, Fr, M	Sum score of weighted items	"Worst case" of the task	Pen & paper
Rapid entire body assessment (REBA), 2000	P, F	Sum score of weighted items	Most common/ prolonged/loaded/postures	Pen & paper
Washington State ergonomic checklists, 2000,	P, F, D, Fr, M, Vib	Yes/no answers that are regular in work	Screening for tasks	Pen & paper
Video- och datorbaserad arbetsanalys [(VIDAR) a video- and computer-based method for ergonomic assessments], 2000	P, F, D, Fr, M	Borg RPE scale	By worker's needs	Computerized, video
Postural loading on the upper-body assessment (LUBA), 2001	P	Posture discomfort score	Most common/loaded postures	Pen & paper, video
Chung's postural workload evaluation, 2002	P	Posture discomfort score	No detailed rules	Computerized, video
Methods assessing workload on upper limbs				
Health and Safety Executive (HSE) upper-limb risk assessment method, 1990	P, F, D, Fr, Vib	Yes/no answers	Tasks involving high repetition/low variety	Pen & paper
Stetson's checklist, 1991	P, F, D, Fr	Frequency of items by their duration	No detailed rules	Pen & paper
Rapid upper-limb assessment (RULA), 1993	P, F, static action	Sum score of weighted items	No detailed rules	Pen & paper, video
Keyserling's cumulative trauma checklist, 1993	P, F, D, Fr, Vib	Sum score of positive findings	Screening of job with questions put to the worker	Pen & paper
Strain index, 1995	P, F, D, Fr	Multiplied score; risk index	No detailed rules	Pen & paper
Occupational Repetitive Actions (OCRA), 1996	P, F, D, Fr, Vib	Sum score of weighted items; risk index	Assessment of repetitive action incl. in profile of work	Pen & paper
American Conference of Governmental Industrial Hygienists hand activity level (ACGIH HAL), 1997	M, F,	Hand activity & force requirement on VAS	"Typical activity"	Pen & paper, (video)
Washington State ergonomic checklists, 2000	P, F, D, Fr, Vib	Yes/no to questions combining risk factors	Items selected by caution zone checklist	Pen & paper
Ketola's upper-limb expert tool, 2001	P, F, D, Fr, Vib	Yes/no answers; profile of items	No detailed rules	Pen & paper
Methods assessing mainly manual material handling				
NIOSH lifting equation, 1981 (revised 1991)	P, F, D, Fr	Multiplied score; risk index	No detailed rules	Pen & paper, computerized
Arbouw, 1997	P, F, D, Fr	3 levels of risk tables	No detailed rules	Pen & paper
New Zealand code for material handling, 2001	P, F, D, Fr	Sum score of weighted items indicating risk	Flowchart; tasks including hazardous MMH	Pen & paper
Manual handling assessment charts (MAC), 2002	P, F, Fr	Item profile; sum score indicating risk	Selection by general knowledge of work	Pen & paper, (video)
Washington State ergonomic checklists, 2000	P, F, D, Fr	Lifting limit computed as multiplied score	Worst & most common lifts	Pen & paper
Manual tasks risk assessment (ManTRA), 2004	P, F, D, Fr, Vib	Sum score of risk	Rules stated in Queensland manual tasks advisory standard	Pen & paper
ACGIH lifting TLV, 2004	P, F, D, Fr	Hazardous lifting TLV	No detailed rules	Pen & paper
Back-Exposure Sampling Tool (BackEst), 2008	P, F, Vib	Frequency of items	Time sampling	Pen & paper

Table 2. Validity and repeatability of observational methods [– =Insufficient information; NIOSH=National Institute of Occupational Safety and Health; MMH=manual material handling]

Method	Correspondence with 'valid' reference [a]	Association with musculoskeletal disorders (MSD) [b]	Intra-observer repeatability	Inter-observer repeatability
General methods				
Ovako working posture assessment system (OWAS)	Moderate (discomfort, technical measures)	X	Good	Good
Arbeitswissenschaftliches erhebungsverfahren zur tätigkeitsanalyse [(AET) ergonomic job analysis procedure]	–	–	–	–
Posture targeting	–	–	–	–
Ergonomic analysis (ERGAN)	–	–	–	–
Task recording and analysis on computer (TRAC)	Moderate (technical measures)	X	–	Moderate–good
Portable ergonomic observation (PEO)	Moderate (video, technical measures)	X	Good	Moderate–good
Hands relative to the body (HARBO)	Moderate (technical measures)	–	–	Good
Plan för identifiering av belastningsfaktorer [(PLIBEL) a method assigned for the identification of ergonomics hazards]	Moderate (AET)	–	–	Moderate
Posture, activity, tools and handling (PATH)	Moderate–good (video, technical measures)	–	Moderate–good	Moderate–good
Quick exposure check (QEC)	Good (video, technical measures)	X	Moderate	Moderate
Rapid entire body assessment (REBA)	Moderate (OWAS)	–	–	Low–moderate
Washington State ergonomic checklists	Moderate	X	–	Moderate
Video- och datorbaserad arbetsanalys [(VIDAR) a video- and computer-based method for ergonomic assessments]	–	–	–	–
Postural loading on the upper-body assessment (LUBA)	–	–	–	–
Chung's postural workload evaluation system	–	–	–	–
Methods to assess workload on upper limbs				
Health and Safety Executive (HSE) upper-limb risk assessment method	–	–	–	–
Stetson's checklist	–	–	–	Moderate
Rapid upper-limb assessment (RULA)	Low–moderate (technical measures, ACGIH HAL, OCRA, strain index)	X	–	Moderate–good
Keyserling's cumulative trauma checklist	Moderate (video, workplace data)	–	–	Low–moderate
Strain index (SI)	Moderate (RULA, ACGIH HAL)	L, X	Moderate–good	Moderate–good
Occupational Repetitive Actions (OCRA)	Moderate (SI, RULA, ACGIH HAL)	X	–	–
American Conference of Governmental Industrial Hygienists hand activity level (ACGIH HAL)	Moderate (video, SI)	L, X	Good	Moderate
Washington State ergonomic checklists	–	X	–	Moderate
Ketola's upper-limb expert tool	Low–moderate (technical measures)	–	–	Moderate
Methods to assess mainly manual material handling				
NIOSH lifting equation	.	X	–	–
Arbouw	Moderate (NIOSH lifting equation)	–	–	–
New Zealand code for material handling	–	–	–	–
Manual handling assessment charts (MAC)	–	–	Moderate–good	Moderate–good
Washington State ergonomic checklists	Moderate (NIOSH lifting equation)	X	–	Moderate
Manual tasks risk assessment (ManTRA)	–	–	–	–
ACGIH lifting threshold limit value	Moderate (NIOSH lifting equation)	–	–	–
Back-exposure sampling tool (BackEst)	Low–moderate (technical measures)	–	–	Moderate

[a] Correspondence with valid reference/repeatability: Good, Moderate, Low,
[b] Association with musculoskeletal disorders: X = association in cross-sectional studies; L = prediction in longitudinal studies,

Table 3. Practical issues relating to observational methods. (R=Researchers; O=Occupational safety/health practitioners/ergonomists; W=Workers/supervisors; – =Insufficient information; ?=not clear; NIOSH=National Institute of Occupational Safety and Health)

Method	Strengths	Limitations	Decision rules	Potential users
General methods				
Ovako working posture assessment system (OWAS)	Widely used and documented	Does not separate right and left upper extremities. Assessments of neck and elbows/wrists are missing. Posture coding crude for shoulders. Time-consuming. Does not consider repetition or duration of the sequential postures	Decision rules based on frequency distribution are arbitrary	R
Arbeitswissenschaftliches erhebungsverfahren zur tätigkeitsanalyse [(AET) ergonomic job analysis procedure]	Gives a broad description of work characteristics. Large experience and databank of results from various fields of occupations to be used as reference.	Only 17 items of 216 are targeted to assess musculoskeletal load	Tentative: "high" scores indicate potentially harmful jobs	R
Posture targeting	Illustrative output:The presentation of the postures in polar coordinates provides quantitative measures on ordinal scales, which are possible to validate using, for example, technical measures.	Suitable only for static postures. Does not consider duration and frequency. It is hard to observe many body segments simultaneously.	–	O, R
Ergonomic analysis (ERGAN)	Computerized registration; illustrative output	Time-consuming. Availability of the software unknown.	–	?
Task recording and analysis on computer (TRAC)	Computerized registration; the software counts distribution of frequency and duration of the events. Flexibility to select the items to be observed according to the purpose.	Mainly focussed on assessing exposure levels; frequencies can be retrieved only in the real-time set-up of the method.	–	O, R
Portable ergonomic observation (PEO)	The method enables registration of posture duration. The data allows further analysis for different purposes	Availability of software unknown. Time-consuming if detailed data is needed. If work pace is rapid, the assessment of several exposure categories is not possible	–	R
Hands relative to the body (HARBO)	Easy to learn, simple to use. Registers the duration of the postures on computer.	Registers only 5 postures to be used as proxy for body postures. Availability of software unknown	–	R
Plan för identifiering av belastningsfaktorer [(PLIBEL) a method assigned for the identification of ergonomics hazards]	General and simple screening tool.	Does not quantify the risk. Relative low repeatability due to the subjective decisions of "no"/"yes".	Tasks with a higher number of "Yes" ticks may require more immediate action	O
Posture, activity, tools and handling (PATH)	Thoroughly developed for easy use at worksite, including a procedure for making job-specific templates for observation. The sampling approach is systematic and well-designed. Data are processed in an automatized procedure on computer.	The method only addresses exposure levels, and only in relative durations. Requires considerable training	–	R
Quick exposure check (QEC)	Easy to use. Applicable for a wide range of tasks. Takes account interaction of risk factors.	Not suitable, when tasks are highly varied. Concentrates on work tasks; the user must decide, which tasks are most loaded	Tentative limits indicating level of risk	O, W, R(?)
Rapid entire body assessment (REBA)	Rapid to use. Computerized registration available in public domain.	Right and left hand have to be assessed separately and there is no method to combine this data; the user has to decide what to observe. Duration and frequency of items not included	Tentative limits indicating level of risk	O, R(?)
Washington State ergonomic rule checklists	Simple, quick, and takes into account most risk factors with duration and frequency.	Limited to screening of risks	Straight forward decision rules.	O, W(?)
Video- och datorbaserad arbetsanalys [(VIDAR) a video- and computer-based method for ergonomic assessments]	Easy to use. Encourages participation of workers. Illustrative output can help workers to understand ergonomic problems in their work.	Subjective evaluation of loading is based on discomfort, which may hamper the decision on the numerical value especially in group assessment. General limitations of video-recordings.	Based on QEC and Swedish regulations	O, W
Postural loading on the upper-body assessment (LUBA)	Simple, easy to use. Scoring based on physiological data. Numeric output can make the decisions easier than a qualitative description.	Does not consider force, duration and repetition	Four action categories proposed by the postural index	O(?), R(?)
Chung's postural workload evaluation system	Computerized registration; Illustrative output. Combines video and rating of postures.	Does not consider external forces.	–	R, O(?)

(continued)

Table 3. Continued.

Method	Strengths	Limitations	Decision rules	Potential users
Methods to assess workload on upper limbs				
Health and Safety Executive (HSE) upper-limb risk assessment method	Easy to use. Straight forward questions. Advice for potential solutions.	Does not consider interaction of the risk factors. Subjective rating: definition of observed items not always clear. No metric measure to quantify the risk.	Tasks with a "Yes" ticks require more detailed risk assessment.	O, W(?)
Stetson's checklist	Selection of most items based on research literature.	Too many items to be observed simultaneously. Does not consider the interaction of risk factors. Evaluation of duration of cycle lengths is probably impossible without a chronometer.	–	R
Rapid upper-limb assessment (RULA)	Easy to use. Computerized registration available in public domain.	Right and left hands have to be assessed separately but there is no method available to combine these scores. Does not consider duration of exposures.	Tentative limits indicating level of risk.	O, R
Keyserling's cumulative trauma checklist	Quick and easy to use.	Rating system is qualitative. Does not consider the interaction of risk factors.	Items with "yes" ticks should be considered potential for actions.	O, R
Strain index (SI)	The method includes main risk factors for distal upper-limb disorders. Takes into account the interaction of observed variables. One figure gives comparison of jobs.	Limited to distal upper limb exposure/risk assessment in monotask jobs. Multiplier values are hypothetical. Subjective assessment; definitions of the criteria are not very clear. Does not consider vibration and contact stress.	Sensitivity and specificity of index described in the literature.	O, R
Occupational Repetitive Actions (OCRA)	Takes into account recovery periods. Estimates the workers risk level by considering all the repetitive tasks in a complex job. The checklist is easy and quick to use.	The use is time consuming. Well trained observers needed.	Cut-off limits indicating needs for actions.	O, R
American Conference of Governmental Industrial Hygienists hand activity level (ACGIH HAL)	Rapid and simple to use. In the assessment individual capacity is considered.	Subjective assessment. Covers a limited number of risk factors.	Clear threshold values for actions for monotask work with duration >4 hours.	O, R
Washington State ergonomic checklists	Simple, quick, takes into account most risk factors with duration and frequency.	Limited to screening of risks.	Straight forward decision rules.	O, W(?)
Ketola's upper-limb expert tool	Quick, easy to use.	Rating system is qualitative. Does not combine the duration and other risk factors.	Items with "yes" ticks should be considered potential for actions.	O
Methods to assess mainly manual material handling				
NIOSH lifting equation	Well documented and tested in several laboratory studies. Sound background based on scientific studies. Outcome related to the risk of the health of back. Calculators available in internet	Plenty of practical limitations for use. Requirement of several technical measures and calculations means increased requirements for skills and time to make the estimation.	Clear threshold values indicating actions.	O, R
Arbouw	Covers lifting, carrying, pushing, and pulling.	Relative time-consuming, but does not give very detailed information.	Clear threshold values indicating actions.	O
New Zealand code for material handling	Includes information on risk factors and solution ideas; takes account many important factors such as size and shape of the load and slippery floor.	The user has to make many decisions with vague rules.	Clear threshold values indicating actions.	O
Manual handling assessment charts (MAC)	Relative simple and easy to use. Well described process for assessment.	Assesses only monotonous lift/carry tasks, not jobs or compound tasks. Includes frequency but not duration of the lifting.	Four level grading for action limits.	O, W(?)
Washington State ergonomic checklists	Simple, quick, takes into account most risk factors with duration and frequency.	Limited to screening of risks.	Straight forward decision rules.	O, W(?)
Manual tasks risk assessment (ManTRA)	Quick and easy to use. Takes into account for the general risk of manual material handling (also duration and repetition).	Definition of the criteria not very clear; subjective assessment. Not clear how to combine multiple tasks to get a job level exposure.	Proposed limits for action categories.	O, R(?), W(?)
ACGIH lifting threshold limit value	Quick, easy to use	Limited to two-handed monolifting tasks.	Clear threshold values for actions.	O
Back-exposure sampling tool (BackEst)	Quite simple. A thorough validation data against technical measures gives a possibility to transform the observed results accordingly	Time-consuming due to the sampling strategy.	–	R

The job itself is evaluated – not the individual doing the job. Of 216 items, 143 concern the person-at-work system, 31 deal with the analysis of tasks, and 42 relate to an analysis of job demands. Of these last 42 items, 17 are directly related to muscular work (ie, analysis of demands/activity). In coding the items, the user should combine observations at the worksite with interviews. The data can be used to characterize the job or task profile. A database with over 7000 jobs serves as a reference for comparisons (21–23). We found no studies comparing AET observations related to musculoskeletal load with more valid measures. Inter-observer repeatability has been studied, but the reports only give figures for the whole method and no separate results for the part related to musculoskeletal workload (21, 23–25).

Ergonomic analysis (ERGAN, formerly ARBAN). In the ERGAN method, the work situation is filmed and the workload on different body parts is assessed from single frames of the video using Borg's scale. Events are counted and registered in time from video recordings. These observations have been used as input for computer software that gives the workload profile in time sequences (26). No studies testing the validity or repeatability of the method were found.

Posture targeting. The posture targeting method involves the observation of static postures with respect to the "standard" anatomical position, which is selected as the centre of the "target" of each body part. A target comprises four concentric circles – similar to polar coordinates – representing angles of 45º, 90º, and 135º deviations from the neutral, center position. Deviations are marked by a cross on targets, which shows the frequency of postures during the observed period (27). Field observations have been compared with observations made from simultaneous photographs (27). However, the trial was so inadequately described that we could not evaluate validity or repeatability.

Task recording and analysis on computer (TRAC). TRAC is a generic method to record tasks, actions, or postures in real time or with computerized time sampling. The events to be observed (eg, postures in a particular category) must be defined a priori. Real-time observations allow users to analyze both the duration and sequence of the selected events and contextual factors of interest during these events (28). In the multi-moment application, the observer must monitor the situation repeatedly at previously selected time intervals, given as auditory signals from the computer. Posture categories in TRAC are user-configurable and a number of applications have adopted the OWAS categorization scheme (29). TRAC observations of postures have shown moderate correspondence with technical measurements (30). Inter-observer repeatability has been moderate to good (29–31).

Portable ergonomic observation (PEO). PEO is a method for the continuous computer-based observation of workers at the workplace. Every time the worker adopts a new predefined posture, performs an action, or moves from one posture to another, the observer hits the corresponding keys and the software records the start time of the event. When the posture changes, or when the activity is terminated, the observer hits the same key again. This triggers the software to calculate and store the duration of this particular event. From this information, anchored in real time, the software calculates the frequency and duration of each item for the observation period. If the frequency and/or total time of the observed task is known, the software calculates the cumulative frequency and duration of postures or actions for a longer period (day/week). The information on daily frequency of tasks needed for the calculation of cumulative exposures is obtained by interviews or production output (32). The method has shown moderate-to-good correspondence with data obtained by video (32) and direct technical measures (33, 34). An association between PEO observations and musculoskeletal discomfort has been seen in cross-sectional studies (35, 36). Intra-observer repeatability has been good (32, 37) and inter-observer repeatability has been moderate to good (32).

Hands relative to the body (HARBO). The HARBO method was developed to assess exposures in epidemiological studies or ergonomic prevention and intervention programs in all types of jobs. Five postures, defined through the position of the hands, can be measured for several hours: (i) standing/walking with hands above shoulder level; (ii) standing/walking with hands between shoulder and knuckle level, not fixed with load; (iii) standing/walking with hands between shoulder and knuckle level, fixed with load; (iv) standing/walking with hands fixed below knuckle level; and (v) sitting. The position of the hands is regarded as a proxy for postural demands on the neck, shoulders, and lower back. Observations are made and registered in real time with a hand-held computer using the software originally developed for the PEO method (38). Observations have been moderately correlated with direct technical measurements of arm and trunk postures. The inter-observer repeatability has been good (38). No reports on associations with MSD were found.

Plan för identifiering av belastningsfaktorer [(PLIBEL) a method assigned for the identification of ergonomics hazards]. PLIBEL is a simple checklist, intended as a rapid screening tool of major ergonomic risks which may have injurious effects on the musculoskeletal system. Time

aspects and environmental and organizational factors are also included as hazard modifiers (39, 40). Some of the items for classifying the workplace have shown moderate correspondence with similar items in AET, and the inter-observer repeatability has been moderate to good (39). No reports on associations with MSD were found.

Posture, activity, tools, and handling (PATH). The basis of the PATH method is work sampling (ie, observing "snapshots"), from which a frequency distribution of observed items is obtained. Before the data collection, a template must be customized for relevant work activities, tools, and handling. The classification of postures is based on OWAS (41, 42). PATH observations have shown good correspondence with observations made using another, less-well described tool for the posture assessment of video recordings (41) and technical measurements in simulated tasks (42). Intra-observer repeatability was good for arm and leg postures but less so for those of the neck and trunk. Inter-observer repeatability has been moderate to good (41). No reports on associations with MSD were found.

Rapid entire body assessment (REBA). REBA was designed as a quick and easy observational postural analysis tool for whole-body activities in healthcare and other service industries. The basic idea of REBA is similar to that of the rapid upper-limb assessment (RULA) method (see p12): positions of individual body segments are observed and postural scores increase when postures deviate from the neutral position. Group A includes trunk, neck, and legs, while group B includes upper and lower arms and wrists. These groups are combined into one of 144 possible posture combinations that are transformed to a general postural score ("grand score"). Additional items are observed and scored including: the load handled, couplings with the load, and physical activity. These scores are summed up to give one score for each observation, which can then be compared to tables stating risk at five levels, leading to the necessity of actions (ranging from "none" to "necessary now") (43–45).

REBA observations have corresponded moderately to those of the OWAS method, although the former classified more postures to have a higher level of risk (18). No reports on associations with MSD were found. Inter-observer repeatability was moderate to good for leg and trunk postures but low for upper limbs (46).

Quick exposure check (QEC). QEC is intended for the rapid assessment of tasks after minimal training of observers. The observer watches and rates postures of the back, shoulder–arm, wrist–hand, and neck on 2- or 3-step scales using "fuzzy logic" (ie, using non-technical language without exact borders between the classes). In addition, the observed worker is required to (self) rate the weights handled, the daily time spent doing the observed task, the level of hand force involved, visual demands, driving of vehicles, the use of vibrating tools, and difficulties to keep up with the work as well as the stressfulness of the work. The ratings are weighted into scores and added up to summary scores for different body parts and other items (driving, vibration, work pace, and stress). Based on these scores, priority levels for intervention are proposed to provide a basis for decision-making and communication within organizations (47–49). Back and shoulder posture results were found to be well correlated with technical measures in simulated tasks. (47, 49). QEC practioners' evaluations corresponded well to those of experts, and the intra- and inter-observer repeatability was moderate (48, 49). No studies on associations with MSD were found.

Washington State ergonomic checklists. These checklists were developed as part of a regulatory effort to control exposure to musculoskeletal hazards in workplaces in the state of Washington, USA. Epidemiological and other scientific studies were the basis for the selection of items to be observed. The evaluation of workplaces covering the main hazards for musculoskeletal disorders is done by two checklists: (i) the "caution zone" checklist is used as a screening tool; (ii) a more comprehensive "hazard zone" checklist is used for those jobs screened to represent potential hazards (50). Only the manual material handling aspect of the method has been compared with other methods (51). Jobs observed to have excessive exposures have had higher occurrence of MSD (52). The assessment has been shown to have good repeatability among observers (53). (See p13 and 14 for details on those parts of the checklist concerning of upper limbs and manual material handling.)

Video- och datorbaserad arbetsanalys [(VIDAR) a video- and computer-based method for ergonomic assessments]. VIDAR's approach differs from those of the previously mentioned methods, all of which are based on the observation of predefined postures and other items by an external observer. VIDAR is a participative method, mainly based on worker's assessment (54, 55). It was developed to support participative interventions and ergonomics training at workplaces. First, an employee is video recorded when performing his/her daily work. He/she then makes an assessment of physically and psychologically demanding situations. For physical situations, the worker marks affected body regions and rates perceived exertion using Borg's category ratio (CR-10) scale. Two checklist modules have been added to the program: one is based on the QEC and the other on official Swedish ergonomic

regulations. The checklists were implemented to clearly point to the needs for intervention and increase the ability of VIDAR to measure the effects of interventions. No studies testing the validity or repeatability of the method were found.

Postural loading on the upper-body assessment (LUBA). In LUBA analysis, postures are rated on scales developed from psychophysical experiments recording discomfort in comparable postures. All scores are summed up to one score describing the urgency of intervention actions. The summary score of postures is compared to experimental maximal holding times in different postures, and this analysis is used to formulate decision rules for the priority of actions needed (56). No studies testing the validity or repeatability of the method were found.

Chung's postural workload evaluation system. In this observational method, postures are rated according to a "discomfort score" associated with each joint posture or combination of postures. The more the joint position deviates from neutral, the higher the score. The scores have been determined in a series of laboratory experiments. Observations can be done in the workplace or from photographs or videos. Computer software has been developed to help code and analyze observations. The rating of postures is partly the same as for LUBA (57, 58). No studies testing the validity or repeatability of the method were found.

Methods to assess workload on upper limbs

Health and Safety Executive (HSE) upper-limb risk assessment method. The HSE risk filter and risk assessment worksheets provide a two-stage assessment process to help reduce work-related upper-limb disorders. As a first step, the risk filter (including questions on symptoms and general risk factors for MSD) is used to identify situations where a more detailed assessment is necessary. Risk assessment worksheets are then used to conduct a more detailed risk assessment for these selected tasks (59). No studies testing the validity or repeatability of the method were found.

Rapid upper-limb assessment (RULA). In the RULA method, positions of individual body segments are observed and scored, with scores increasing in line with growing deviation from the neutral posture. Summary scores are first calculated separately for both upper and lower arms and wrists (group A) and trunk, neck and legs (group B), and then transformed to a general postural "grand score". Additional weights are given to the postures according to forces/loads handled and the occurrence of static/repetitive muscular activity. These scores are then compared to tables stating risk on four levels and the actions needed (ranging from "acceptable" to "immediate investigation and change needed") (60, 61). RULA's posture scores have been shown to have low correspondence to the hand-use intensity scores of the strain index (see p13) (62). Still, in another study, it showed saw-filers work to be risky in line with the evaluations made using REBA, the American Conference of Governmental Industrial Hygienists threshold limit value for hand activity level (ACGIH HAL), the strain index, and occupational repetitive actions (OCRA) methods (see following pages) (63). Observations made using RULA have also been compared with those using the OWAS and REBA methods (18). The correspondence with all these methods has been moderate at best, but it remains unknown which method better reflects the underlying MSD risks for varying tasks (18). Higher RULA scores have shown an association with increased discomfort in laboratory studies (60, 64) and with perceptions of MSD in two cross-sectional field studies (65, 66). The inter-observer repeatability of RULA has been found to be good, although the methodological information on the repeatability studies is so scant that the quality thereof cannot be evaluated (60, 66).

Stetson's checklist for the analysis of hand and wrist. This method was developed as a quantitative measure of repetitive hand exertions for studies of cumulative trauma disorders. Observed objects include hand exertions while using power tools, pinch grip, high force, palm as a striking tool, and "involuntary" wrist deviation. The number of these exertions is recorded for "standard" work cycles and classified by the duration of exertions. This information, multiplied by the number of work cycles per shift, produces a quantifiable measure of repetitiveness (67). No studies testing the validity of the method were found. Inter-observer repeatability for the counts of observed cycles has been reported to be moderate for pinch and exertion (67, 68).

Keyserling's cumulative trauma checklist. The aim of the checklist is to determine the presence of ergonomic risk factors associated with the development of upper-extremity cumulative trauma disorders. Repetitiveness, local contact stresses, forceful manual exertions, awkward upper-extremity posture, and hand-tool usage are evaluated; each detected risk factor is recommended to be further evaluated (69). The method has shown moderate correspondence with observations from video recordings of postures, contact forces, and tool use, as well as with product output data describing repetitiveness (69). No studies on associations with MSD were found. The inter-observer repeatability has been fair to moderate for pinch grip and shoulder elevation above 45º, but poor for wrist deviations (68).

Ketola's upper-limb expert tool. This is a semi-quantitative time-based method for assessing the presence ("yes" or "no") of risk factors for upper-limb disorders. The limits for "yes" versus "no" are defined by the time proportion of the cycle, during which the exposure occurs as in the Keyserling's checklist. A higher total number of "yes" answers leads to a greater predicted risk of upper-limb disorders (70). When expert observation was used as a reference standard, validity has ranged from moderate to good for repetitive use of the hand, hand force, pinch grip, and non-neutral wrist posture. Correspondence was low when the observations were validated against wrist goniometric data and force estimations by electromyography. No studies on the association with MSD were found. The inter-observer repeatability has been moderate (70).

Strain index. The strain index is a semi-quantitative job analysis method yielding a numerical score, which is intended to correlate with the risk of developing distal upper-extremity disorders. According to the index six task variables describing hand exertions must be observed and scored on five levels. The six variables include: (i) intensity of exertion, (ii) duration of exertion, (iii) exertions per minute, (iv) hand–wrist posture, (v) speed of work, and (vi) duration of work per day. Each score is then weighted based on physiological (endurance, fatigue, recovery), biomechanical (internal forces, nonlinear relationship between strain and intensity of effort), and epidemiological principles. Multiplying the weighted scores gives a single figure: the strain index (71, 72). Comparison of the index with RULA has shown a limited correspondence with respect to the identification of risks (62, 73). The correspondence of the strain index with ACGIH HAL (see below) was moderate, but the former gave generally higher risk estimates than the latter (74, 75). It is not known which of these methods is more valid to assess risk. In a prospective study, the sensitivity was 0.91 and the specificity 0.83 to predict upper-limb disorders, when using a cut-off point of strain index=5.0 (76). Clear associations with upper-limb disorders have also been demonstrated in several retrospective studies (71, 75, 77, 78). Intra-observer and inter-observer repeatability have been reported to be moderate to good (79, 80).

The American Conference of Governmental Industrial Hygienists threshold limit value for hand activity level (ACGIH HAL). The ACGIH HAL evaluates the risk of developing disorders in the hand, wrist, or forearm on the basis of HAL and peak hand forces. It is aimed at the assessment of single-task jobs with at least four hours per day of repetitive handwork. HAL is rated on a visual-analogue scale (VAS) of 0–10 and addresses exertion frequency, recovery time, and the speed of motion. Peak force can be measured using a strain gauge or other instrumentation (normalized on a scale of 0–10) or estimated by a trained observer using a modified Borg CR-10 perceived effort scale. Peak force is judged relative to the population characteristics at the evaluated worksite, and so depends on factors like age and gender. The combination of HAL and peak hand force is evaluated against two limits: one indicating the necessity of the action and the other showing an absolute maximum allowance for hand-intensive work (81).The method was originally validated against detailed information from video recordings (number of exertions per second, recovery time per cycle, cycle time) with which it showed moderate correspondence (82). The action threshold levels have been compared to those generated by the strain index. The correspondence has been moderate, even though the strain index identified more hazardous jobs than the HAL method (74, 75). An association with upper–limb disorders has been seen in several cross-sectional studies (75, 78, 83, 84) as well as in prospective longitudinal studies (85–90). The intra- and inter-observer repeatability has been moderate to good (82, 91).

Occupational repetitive actions (OCRA). OCRA is a synthetic index describing risk factors related to repetitive actions at work. The index is the total number of technical actions performed during the shift divided by the total number of recommended technical actions. The latter is counted from observed actions multiplied by weights given for the following factors: muscle force, posture of the parts of the upper limb, lack of recovery periods, the daily duration of the repetitive work, and "additional factors". A simplified OCRA checklist is intended for use as a preliminary screening tool (92–95). OCRA has shown moderate correspondence with ACGIH HAL and the strain index (63, 96). Prevalence of upper-limb disorders has been greater in jobs with a higher, as opposed to lower OCRA index (97, 98). No studies on the repeatability of the method were found.

Washington State ergonomic checklists. The Washington State ergonomic checklists includes the following items to evaluate the risks for upper-limb disorders: (i) working with elevated arms, (ii) high hand force, (iii) highly repetitive motions, and (iv) high impact on the hand. The sensitivity and specificity of the tool to identify upper-limb disorders has been found to be low (52). The checklist has shown moderate inter-observer repeatability (53).

Methods to assess manual material handling

US National Institute of Occupational Safety and Health (NIOSH) lifting equation. The NIOSH lifting equation method was developed to assess the risk of low-back disorders in jobs with repeated lifting. Six factors related to the lifting conditions should be observed

and entered into an equation, yielding a recommended weight limit for the task. Multipliers are based on biomechanical, physiological, psycho-physiological, and epidemiological data. The lifting index (LI) is the ratio of the actual weight handled to the recommended weight limit. LI values below 1.0 are regarded to be safe for the average population, and the greater the index the greater the risk of low-back injury (99–101). The LI is restricted to jobs consisting of one or a few similar lifting tasks. For jobs with multiple tasks, procedures have been proposed to calculate a composite LI (100, 101) or a sequential LI (102) for the overall job. The results of using the NIOSH lifting equation have been compared with those of several other methods (15, 51, 103–106), but it is not possible to state whether any of these methods are more valid than the lifting equation. Due to its complex structure of multipliers in defining the recommended weight limit, there are no technical measures that can be used as a "gold standard" for the NIOSH evaluation.

A sensitivity analysis of laboratory simulations has shown that frequency and horizontal location are the most decisive parameters in the NIOSH equation, but that these parameters also tended to have the highest measurement errors (107). Several studies have found the occurrence of low-back pain to be higher in jobs with a higher LI (108–111). Inter-observer variability of measures had little influence on the total LI except in situations where large weights were handled (112).

Arbouw guidelines on physical workload. The Arbouw method was developed for the assessment of five areas of manual material handling: lifting, pushing and pulling, carrying, static load, and repetitive work. The guidelines are based on the NIOSH lifting equation and standards for manual material handling. A traffic light (green, yellow, red) approach is applied to guide recommended actions (113). The Arbouw lifting guidelines can be considered as a simplified version of the NIOSH lifting equation (15). The Arbouw method has shown a moderate correspondence with results from the NIOSH method (15). No studies were found on either its association with MSD or its repeatability.

New Zealand code of practice for manual handling. In the New Zealand code, manual material handling risks are first evaluated by means of a checklist. If this initial screening suggests a potential hazard, a detailed observational check is performed to calculate a risk score, which serves as a guide on the urgency and type of control measure required to reduce the risk. If the risk score is >10, a "factors assessment" can be used to determine the significance of contributing factors (ie, load, environment, people, task, and management) and thereby generate solutions for controlling risks associated with each factor. From this, the importance of the risk is assessed and an action plan is established for control thereof (114, 115). No studies testing the validity or repeatability of the method were found.

Manual tasks risk assessment (ManTRA). ManTRA was developed to assist safety inspectors auditing workplaces across all industries for compliance with the Queensland (Australia) manual tasks advisory standard (116). For each task in a job, the observer rates on a 5-step scale: (i) total time spent on the task during a typical day, (ii) repetition (combination of duration and cycle time), (iii) exertion (combination of force requirements and speed of movements), (iv) awkwardness (deviation from the mid range of movements), and (v) vibration. The scale relates to four body regions (ie, lower limbs, back, neck–shoulder, arm–wrist–hand). High scores of individual items or a high summary score are assumed to indicate an increased risk for MSD (117, 118). No studies testing the validity or repeatability of the method were found.

Manual handling assessment charts (MAC). MAC is a checklist designed to help health and safety inspectors to assess the most common risk factors in lifting, carrying, and team handling operations. The method sets out 11 items of manual handling to be evaluated according to a four-grade "traffic light"; a summary score is counted to prioritize those tasks that require urgent attention and check the effectiveness of those improvements (119–121). The properties of the method have been benchmarked against several other methods in a qualitative manner (120), but no formal comparison on validity was found. The assessment has shown moderate-to-good intra- and inter-observer repeatability on observations from video recordings (122, 123).

Washington State ergonomic checklist for manual handling. The lifting analysis included in this checklist is a simplified version of the NIOSH lifting equation, but the Washington checklist allows higher acceptable weights of the handled load (124). In a comparative trial, the method indicated a substantially lower risk level in a particular job than the NIOSH lifting equation, but the results were similar to those obtained using a biomechanical model (51). When used to detect jobs with an increased risk of back disorders, the checklist showed moderate sensitivity and specificity (52). The assessment has shown moderate inter-observer repeatability (53).

ACGIH lifting threshold limit value for low-back risk. The ACGIH lifting threshold limit value is estimated from the location of the handled material relative to the body, as well as the frequency and daily duration of lifting. ACGIH provides tables setting out the threshold limit

value for weights, below which no health risk is assumed to occur. The method was developed to ensure a lifting guideline that was accurate, based on the latest scientific information, and easy to use (125). The ACGIH lifting threshold limit value method has shown moderate-to-high correspondence with results obtained by the NIOSH lifting equation and Snook's psychophysical method for setting lifting limits (51). No studies were found on either associations with MSD or repeatability.

Back-exposure sampling tool (BackEST). In connection with a large-scale epidemiological study, BackEST aimed to evaluate physical back injury risk factors in demanding work conditions. In the development phase, a literature review suggested 53 relevant exposure variables; these were reduced to 20 items concerning posture, manual material handling, and whole-body vibration. The items were observed once every 60 seconds over a full-shift to produce data that can be analyzed according to the purposes of the individual research (126). The proportion of time in demanding back postures was compared with technical measures and the match was found to be moderate at best, although the finding may have been affected by the different sampling procedures of the compared tools (126). Inter-observer repeatability has been moderate (126).

Discussion

Observation-based assessments of biomechanical exposures (physical loads) on the musculoskeletal system have mostly been targeted at postures of the whole body or individual body regions, as well as exerted manual forces or weights handled manually. Still, there are no common metrics which enable a direct comparison of the different methods, although previous reviews have applied several approaches to overcome this problem. In his review of 12 methods, Genaidy (3) classified postures into macro- or micro-postural or postural-work activities. In other reviews, angles used for the classification of postures and scales of other items of methods have been tabulated (4, 5, 7). A general conclusion of these reviews is that the observational variables generated by different methods are not directly comparable, mainly due to the use of different body angles or angle sectors. Some reviews have mainly been devoted to describing existing methods published in the literature (6, 8). Validity and repeatability of the methods has been addressed in two previous reviews (4, 7), in which some statistical figures of intra- and inter-observer repeatability as well as descriptions of tests on internal and external validity have been tabulated. Several workload standards also contain items meant to be observed (127), but in the present evaluation we included only those methods with observations as the principal component.

The aim of our evaluation was to present information that may help potential users select the most appropriate method(s) for their particular purpose. In situations where practical ergonomic problems have to be solved, simplicity, utility, and face validity of the method are more important than expressing results in exact numeric figures (6, 8, 128). Validity and repeatability of data are particularly important in research and when comparing exposures to safety limits. Thus the selection of the most appropriate tool must involve consideration of the analysis' objectives and how the results will be used – a tool to help identify improvement opportunities may not have the same precision requirements as one being used to judge the safety of a job in a pass–fail determination process.

Validity and repeatability of observational methods

The concept of validity includes several aspects (1, 2). Since there is no "gold standard" to measure physical workload, criterion validity was, in this study, assessed in terms of concurrent validity (ie, the agreement of the observational method with some other measurement method considered to be more valid). Of the 30 methods included in our review, 19 had been compared with some other method(s), varying from expert evaluations to observations made from video recordings, and direct measurements with technical instruments. Generally, the worksite observation methods showed moderate-to-good agreement with measures based on visual recordings, and the correspondence was best for macro-postures and work actions. Micro-postures [like those of the wrist and hand (67, 70, 129), neck (34) and trunk rotation (126)] seem to be more difficult to observe with satisfactory accuracy.

When technical measurements have served as reference, correspondence has generally been lower than when using video-based observations as the reference (14, 30, 33, 34, 70, 126). In these comparisons, the variables of interest have been mainly frequency or duration of postures classified according to category limits set by the observational method. This kind of comparison may be sensitive to the "border lines" of the categories (32, 34, 67, 129–132). In other words, if the observer's perception is systematically biased in comparison to the limit used by the accurate technical method, the probability of having a high correspondence will decrease if the true posture is close to a category border line ("boundary zone problem"). This source of disagreement was seen in some reports, for instance when categorizing severe and moderate trunk flexion postures in PATH (41). Unfortunately, none of the studies had conducted a sensitivity analysis to see if the

shift of the technical measure limit might change the correspondence with observations. In the QEC method, exact limits between the observed categories have intentionally been omitted in order to improve tool usability (47). Another possible explanation for divergent results may be that the technical reference measures were collected on a continuous time-line, while the observations were done using limited sampling at fixed-time intervals (14, 30, 41, 126).

Different methods used simultaneously on the same object have given different results in several studies. If one of the methods is known or suspected to give more accurate information than another, it can be regarded as a validation reference, provided the two methods are measuring the same variable. Still, many of these studies have chosen to compare methods rather than validating one using the other as a reference since none of the methods has been shown to be systematically more accurate than the others (eg, comparisons between the strain index, ACGIH HAL and OCRA). If the output of the assessment method is a compound risk index (as in QEC, REBA, RULA, strain index, OCRA, NIOSH equation), its concurrent validity can, in theory, be estimated on the basis of technical measurements of each individual item. However, compound indices and sum scores are sensitive to the weights given to individual items in the calculation of the index or score. For example, a laboratory simulation of lifting tasks showed that the value of the NIOSH lifting index was highly sensitive to the frequency of lifting and the horizontal location of the load (107).

Historically, observational methods have developed from the common experience that some visually detectable postures and actions are related to discomfort or disorders in the musculoskeletal system; a notion which has later been demonstrated in numerous experimental and epidemiological studies. Theoretical constructs, combining the physiological, epidemiological, and biomechanical knowledge, have shown that mechanical forces acting on the tissues is probably the most important factor in explaining how MSD can develop. In addition to the magnitude of the exposures, time aspects related to physiological responses are of importance (133–137). Therefore an observation method with a good content validity (1, 2) for mechanical exposures should include the frequency and duration of items quantifying exposures – like external forces or postures – in addition to their magnitude.

In our evaluation, we addressed whether findings using a particular observation method had been associated with MSD (the so-called "predictive validity"). Even though assessments of several methods have correlated with the occurrence of MSD in cross-sectional settings, only a few cohort studies have analyzed possible associations using terms such as sensitivity and specificity, which can support conclusions on the predictive validity of the method (75, 85–90, 138, 139). Outcomes used in these risk studies have varied from discomfort and fatigue in laboratory settings to well-defined clinical diseases in prospective cohorts; this variety in outcome also hinders the comparisons of results from different studies, and thus obstructs the drawing of conclusions on the validity of risk limits given by any particular method.

Intra- and inter-observer repeatability were reported only for 7 and 17 methods, respectively. They were mostly reported to be good or moderate. Generally the repeatability is better within observers than between them, as suggested for those methods where both sources of variance had been studied. However, it should be noted that in most cases inter-observer reliability was assessed without consideration of the effects of intra-observer variability. If this effect is not acknowledged, estimates of inter-observer variability will be systematically inflated. The inter-observer reliability of many methods may, therefore, be better than what is reported in the studies.

Repeatability is highly related to the user skills, which can be enhanced with appropriate training (4). In some trials, the observers had to improve their skills until a preset agreement was reached (ie, TRAC, PATH) this resulted in a good reliability rating in our evaluation. For most methods, the literature did not mention the duration of training needed to reach a satisfying proficiency in using the tool.

Identification and selection of references

Guidelines on how to conduct systematic reviews (140–142) have stressed the importance of systematic search strategies in electronic databases. Using the various combinations of search terms set out earlier and a manageable number of references, we were unable to identify all of the relevant methods in our preliminary searches. Therefore, we continued searching using the names of the identified methods and the option of an automatic search for "related studies", followed by a visual screening of the resulting lists of publications. We were also aware that practitioners may use methods that have not necessarily been subjected to scientific testing and, therefore, will not be found in databases compiling scientific reports. Consequently, we supplemented our searches with internet searching once we had identified the names of the methods. In addition to papers in English, we accepted reports in German, French, Italian, and Scandinavian languages. Through this extensive search, we have probably identified most observational methods used for the assessment of biomechanical exposures. We acknowledge the existence of additional methods that were not accessible to us due to their limited availability

(eg, commercial methods or additional academic studies not listed in the internet databases). Previous reviews have also mentioned references in conference books that were not available to us.

We did not make a formal quality assessment of studies to be included or excluded for evaluation. Our selection of methods was more inclusive than restrictive. This decision was deliberate – even the methods that have not been tested in a formal way may have features that make them suitable for some users, or they may provide ideas for further development for specific purposes. Our results showed that most of the included methods, although appearing to a large extent in the scientific literature, had not been tested in a systematic, scientific way.

Evaluation process

There are no standards for the evaluation of methods assessing biomechanical exposures. Therefore, our group had to develop the evaluation procedure. The selection of items to be evaluated was based on a consensus of what are perceived to be important matters associated with the assessment of physical workload in the practice and research of occupational health and safety (4, 6, 8, 134, 143–145). The evaluation model was developed in an iterative manner and tested with a subset of methods. We aimed to base the conclusions of concurrent validity and repeatability on original studies of good quality. We had to relinquish this requirement because peer-reviewed reports were very few, or even absent, for a majority of the methods. However, the evaluations of the methods, done by experts blinded to the each others' results, were mostly identical. This probably reflects the fact that a common understanding of items and evaluation criteria was reached during the development of the evaluation process.

We do not consider direct numerical comparisons between studies and methods to be warranted. This is a result of considerable differences in study design and settings, the number and definition of items to be observed, sample sizes (including the amount of work observed and the number of observers used), and the statistical methods used, for instance to quantify reliability. Therefore, we decided to evaluate validity and repeatability in qualitative terms ("good"," moderate", "low") without strictly defined borders between these categories.

Information on the repeatability of most methods was limited in the original reports, excluding the details of the repeatability study itself. In our evaluation, this information was taken for granted if the main part of the report seemed to be otherwise credible. Studies comparing observations with other methods generally gave a more detailed description of their design, and so it was possible to identify potential biases in some reports. We did not make a formal assessment on the quality of the assessments of agreement, but considered the quality as sufficient if no obvious source of error was detected. We believe that this evaluation will be more valuable than simply labeling most of the studies as having "insufficient information", as the readers can obtain the original reports if they wish to scrutinize the data further.

Likewise, we did not make a formal assessment of content validity with a predefined list of items that the observational methods should address based on theoretical constructs. In the assessment of face validity and practical issues for the use of methods, the contents of each method was evaluated keeping the purpose of the method in mind. During the evaluation, we noticed that the ranking scale we planned to use for face validity was not used consistently by all the evaluators. Therefore, in table 3 we have reported only the expert's verbal comments on practical aspects.

Future development and evaluation of methods would probably benefit from a set of guidelines on "minimal requirements for reporting", similar to those adopted for clinical trials (146) and epidemiological studies (147).

Selection of methods for different purposes

The aim of this project was to analyze observational methods in order to help users select the most appropriate method(s) for their specific purposes. None of the evaluated methods will be universally ideal for all purposes, due to the variety of user needs and the diversity of settings in working life. Besides validity issues, the user will have to consider the trade-off between accuracy, complexity, costs, and ease-of-use when identifying an appropriate method in a particular setting (134). Sometimes rough and qualitative information may be sufficient, while in other situations detailed precise information is required as a sound basis for decision-making. When selecting a method, users should first define their needs and constraints, after which our evaluation of, and further details in, the original reports may help them to select the optimal method to be used or modified for their specific purpose. Since the observation methods described in this report are often only partly correlated, and all have at least face validity, it may be useful for practitioners to apply several methods while in the field. In addition, more qualitative assessments, including interviews are recommended, so as to improve the "coverage" of the assessment and minimize the risk of missing potentially hazardous elements in the job.

Most jobs and tasks show daily or seasonal variability in biomechanical exposures, and, therefore, the sampling strategy – in terms of the number of subjects and measurements per subject – is decisive for the precision and credibility of the obtained result (148–156). Some of

the methods included in the present selection are based on sampling at fixed-time intervals (eg, OWAS, PATH), while others apply continuous observation for longer periods (eg, PEO, TRAC). In both cases, variability within and between days, as well as between subjects, has to be considered when designing the measurement strategy; even a random sampling may be efficient (157, 158). Examples of random interval sampling exist that may overcome problems with regular observations on routine cycle jobs (157). A particular problem with instantaneous work sampling ("snapshots") is that the frequency and duration of actions cannot readily be assessed. Very few of the reviewed reports give any advice on appropriate sampling strategies when using the method, and in most cases reliability is assessed using indices – typically the kappa coefficient – that cannot be used for designing new measurement strategies with an appropriate performance. Identifying an appropriate sampling strategy can be at least as important for the credibility and interpretation of the obtained result as the chosen assessment instrument per se.

Based on the analysis of the methods and the items discussed before, we make the following recommendations: (i) Before commencing the observation, clearly define aims and needs as well as resources and a sampling strategy, acknowledging the variability of exposures over time and between subjects. It may be worth considering the use of multiple tools to capture different aspects of the job in question. Self-reports may be an option, but users should be aware of the problems related to their validity (159); (ii) For general screening purposes, use simple and rapid methods (eg, QEC, Washington State ergonomic checklists, ACGIH HAL, Arbouw, other checklists) to check if more accurate ones are needed; (iii) For problematic tasks identified, consider the use of methods giving numeric output (eg, OCRA, strain index, ACGIH lifting threshold limit value); (iv) If detailed data on frequency and duration of actions is needed, we recommend continuous observations (eg, TRAC, continuous analysis from video recordings, or technical measurements); (v) Observers should be sufficiently trained in the use of the tool to reduce methodological errors caused by intra- and inter-observer variability.

Concluding remarks

Numerous observational tools have been developed to assess biomechanical exposures and, in general, they all address the same known risk factors. Only some of these tools have been tested in a systematic manner for validity, repeatability, and aspects related to their practical use. Repeatability studies have shown that different observers will report reasonably similar results when observing large-scale body postures and work activities if they have adopted similar concepts and skills through sufficient training. Visual observation of smaller body regions and movements seems to be more challenging and, hence, less reliable, especially when the movements are fast. Observational results have in general agreed moderately with detailed information collected from video recordings, while correspondence with technical measurements has generally been lower.

No single tool appears to have a clear advantage over any other. When trying to select the most appropriate method in a specific setting, users should thoroughly define their needs and how the information will affect decision-making. In addition to choosing an appropriate method, the sampling strategy is essential if the results are to be generalized beyond the observed sample.

Acknowledgements

This study was financially supported by the Nordic Council of Ministers (Project 411040 – 70107)

References

1. Streiner DL, Norman GR. Health measurement scales: a practical guide to their development and use. 2nd ed. Oxford (NY): Oxford University Press; 1995.
2. Last JM, editor. A dictionary of epidemiology. 3rd ed. New York (NY): Oxford University Press; 1995.
3. Genaidy AM, Al-Shedi AA, Karwowski W. Postural stress analysis in industry. Appl Ergon. 1994;25(2):77–87.
4. Kilbom Å. Assessment of physical exposure in relation to work-related musculoskeletal disorders – what information can be obtained from systematic observations? Scand J Work Environ Health. 1994;20 special issue:30–45.
5. Juul-Kristensen B, Fallentin N, Ekdahl C. Criteria for classification of posture in repetitive work by observation methods: a review. Int J Ind Ergonomics. 1997;19(5):397–411.
6. Li G, Buckle P. Current techniques for assessing physical exposure to work-related musculoskeletal risks, with emphasis on posture-based methods. Ergonomics. 1999;42(5):674–95.
7. Denis D, Lortie M, Rossignol M. Observation procedures characterizing occupational physical activities: critical review. Int J Occup Saf Ergon. 2000;6(4):463–91.
8. David GC. Ergonomic methods for assessing exposure to risk factors for work-related musculoskeletal disorders. Occup Med. 2005;55(3):190–9.
9. Kilbom Å, Horst D, Kemmlert K, Richter A. Observationsmetoder för registering av belastningar på rörelseapparaten – en litteraturstudie [Observational methods to register load on the locomotor system – literature review]. Stockholm: Arbetarskyddsverket; 1986. Arbete och hälsa (21); p 1–90.

10. Observational methods assessing biomechanical exposures at work [Internet]. Helsinki: Finnish Institute of Occupational Health [cited 5 October 2009]. Available from: www.ttl.fi/workloadexposuremethods

11. Karhu O, Kansi P, Kuorinka I. Correcting working postures in industry: a practical method for analysis. Appl Ergon. 1977;8(4):199–201.

12. Kayis B, Kothiyal K. Multilevel approach to manual lifting in manufacturing industries. Int J Occup Saf Ergon. 1996;2(3):251–61.

13. Olendorf MR, Drury CG. Postural discomfort and perceived exertion in standardized box-holding postures. Ergonomics. 2001;44(15):1341–67.

14. Burdorf A, Derksen J, Naaktgeboren B, Van Riel M. Measurement of trunk bending during work by direct observation and continuous measurement. Appl Ergon. 1992;23(4):263–7.

15. van der Beek AJ, Mathiassen SE, Windhorst J, Burdorf A. An evaluation of methods assessing the physical demands of manual lifting in scaffolding. Appl Ergon. 2005;36(2):213–22.

16. Burdorf A, Govaert G, Elders L. Postural load and back pain of workers in the manufacturing of prefabricated concrete elements. Ergonomics. 1991;34(7):909–18.

17. de Bruijn I, Engels JA, van der Gulden JWJ. A simple method to evaluate the reliability of OWAS observations. Appl Ergon. 1998;29(4):281–3.

18. Kee D, Karwowski W. A comparison of three observational techniques for assessing postural loads in industry. Int J Occup Saf Ergon. 2007;13(1):3–14.

19. Kivi P, Mattila M. Analysis and improvement of work postures in the building industry: application of the computerised OWAS method. Appl Ergon. 1991;22(1):43–8.

20. Mattila M, Karwowski W, Vilkki M. Analysis of working postures in hammering tasks on building construction sites using the computerized OWAS method. Appl Ergon. 1993;24(6):405–12.

21. Rohmert W, Haider E, Landau K. Entwicklung und Anwendung eines Arbeitswissenschaftlichen Erhebungsverfahrens zur Tätigkeitsanalyse zum Anforderungsbereich Handlung (H-AET) [Development and practicability of an ergonomic job description questionnaire for the area of requirement "activity" (A-EJDQ)]. Int Arch Occup Environ Health. 1979;43(1):17–35.

22. Rohmert W. AET – a new job-analysis method. Ergonomics. 1985;28(1):245–54.

23. Landau K, Brauchler R, Rohmert W. The AET method of job evaluation. In: Karwowski W, Marras WS, editors. The occupational ergonomics handbook. Boca Raton (FL): CRC Press; 1999. p 355–70.

24. North K, Rohmert W. Ermittlung der Arbeitsanforderungen für Leistungsgeminderte [Job analysis applied to the disabled]. Int Arch Occup Environ Health. 1980;47(2):143–54.

25. Haider E, Rohmert W. Anforderungsermittlung im Bereich der Daten- und Textverarbeitung mit einem arbeitswissenschaftlichen Erhebungsverfahren zur Tätigkeitsanalyse (DTV-AET) [Ergonomic strain measurement in text processing by an ergonomic job description questionnaire (DT-EJDQ)]. Int Arch Occup Environ Health. 1982;50(3):245–60.

26. Holzmann P. ARBAN-A new method for analysis of ergonomic effort. Appl Ergon. 1982;13(2):82–6.

27. Corlett EN, Madeley SJ, Manenica I. Posture targetting: a technique for recording working postures. Ergonomics. 1979;22(3):357–66.

28. Frings-Dresen MHW, Kuijer PPFM. The TRAC-system: an observation method for analysing work demands at the workplace. Saf Sci. 1995;21(2):163–5.

29. van der Beek AJ, van Gaalen LC, Frings-Dresen MH. Working postures and activities of lorry drivers: a reliability study of on-site observation and recording on a pocket computer. Appl Ergon. 1992;23(5):331–6.

30. DeLooze MP, Toussaint HM, Ensink J, Mangnus C, van der Beek AJ. The validity of visual observation to assess posture in a laboratory-simulated, manual material handling task. Ergonomics. 1994;37(8):1335–43.

31. van der Molen HF, Mol E, Kuijer PP, Frings-Dresen MH. The evaluation of smaller plasterboards on productivity, work demands and workload in construction workers. Appl Ergon. 2007;38(5):681–6.

32. Fransson-Hall C, Gloria R, Kilbom Å, Winkel J, Karlqvist L, Wiktorin C. A portable ergonomic observation method (PEO) for computerized on-line recording of postures and manual handling. Appl Ergon. 1995;26(2):93–100.

33. Karlqvist L, Winkel J, Wiktorin C. Direct measurements and systematic observations of physical workload among medical secretaries, furniture removers and male and female reference populations. Appl Ergon. 1994;25(5):319–26.

34. Leskinen T, Hall C, Rauas S, Ulin S, Tönnes M, Viikari-Juntura E, et al. Validation of Portable Ergonomic Observation (PEO) method using optoelectronic and video recordings. Appl Ergon. 1997;28(2):75–83.

35. Murphy S, Buckle P, Stubbs D. Classroom posture and self-reported back and neck pain in schoolchildren. Appl Ergon. 2004;35(2):113–20.

36. Rolander B, Karsznia A, Jonker D, Oberg T, Bellner AL. Perceived contra observed physical work load in Swedish dentists. Work. 2005;25(3):253–62.

37. Rehn B, Nilsson T, Olofsson B, Lundström R. Whole-body vibration exposure and non-neutral neck postures during occupational use of all-terrain vehicles. Ann Occup Hyg. 2005;49(3):267–75.

38. Wiktorin C, Mortimer M, Ekenvall L, Kilbom Å, Wigaeus Hjelm EW. HARBO, a simple computer-aided observation method for recording work postures. Scand J Work Environ Health. 1995;21(6):440–9.

39. Kemmlert K. A method assigned for the identification of ergonomic hazards – PLIBEL. Appl Ergon. 1995;26(3):199–211.

40. Kemmlert K. PLIBEL – a method assigned for identification of ergonomic hazards. In: Marras WS, Karwowski W, editors. Fundamentals and assessment tools for occupational ergonomics. Boca Raton (FL): CRC Press; 2006. p 40:1–14.

41. Buchholz B, Paquet V, Punnett L, Lee D, Moir S. PATH: a work sampling-based approach to ergonomic job analysis for construction and other non-repetitive work. Appl Ergon. 1996;27(3):177–87.

42. Paquet VL, Punnett L, Buchholz B. Validity of fixed-interval observations for postural assessment in construction work. Appl Ergon. 2001;32(3):215–24.

43. Hignett S, McAtamney L. Rapid entire body assessment (REBA). Appl Ergon. 2000;31(2):201–5.

44. McAtamney L, Hignett S. Rapid entire body assessment. In: Stanton N, Brookhuis K, Hedge A, Salas E, Hendrick HW, editors. Handbook of human factors and ergonomics methods. Boca Raton (FL): CRC Press; 2005. p 8:1–11.

45. Hignett S, McAtamney L. REBA and RULA: whole body and upper limb assessment tools. In: Marras WS, Karwowski W, editors. Fundamentals and assessment tools for occupational ergonomics. Boca Raton (FL): CRC Press; 2006. p 42:1–12.

46. Janowitz IL, Gillen M, Ryan G, Rempel D, Trupin L, Swig L, et al. Measuring the physical demands of work in hospital settings: design and implementation of an ergonomics assessment. Appl Ergon. 2006;37(5):641–58.

47. Li G, Buckle P. Evaluating change in exposure to risk for musculoskeletal disorders—a practical tool [Internet]. Norwich (United Kingdom): Her Magesty's Stationery Office; 1999 [cited 15 October 2009]. Contract Research Report 251. Available from: http://www.hse.gov.uk/research/crr_pdf/1999/crr99251.pdf

48. David G, Woods V, Buckle P. Further development of the usability and validity of the Quick Exposure Check (QEC) [Internet]. Norwich (United Kingdom): Her Magesty's Stationery Office; 2005 [cited 15 October 2009]. Research Report: RR211/2005]. Available from: http://www.hse.gov.uk/research/rrpdf/rr211.pdf

49. David G, Woods V, Li G, Buckle P. The development of the Quick Exposure Check (QEC) for assessing exposure to risk factors for work-related musculoskeletal disorders. Appl Ergon. 2008;39(1):57–69.

50. Washington-State-Department of Labor and Industries. Concise Explanatory Statement (RCW 34.05.325.6a) of WAC 296–62–051, Ergonomics. Olympia (WA): Washington-State-Department of Labor and Industries; 2003.

51. Russell SJ, Winnemuller L, Camp JE, Johnson PW. Comparing the results of five lifting analysis tools. Appl Ergon. 2007;38(1):91–7.

52. Eppes S. Washington State Ergonomics Tool: predictive validity in the waste industry [dissertation][Internet]. College Station (TX): Texas A&M University; 2004 [cited 15 October 2009]. Available from: http://repository.tamu.edu/bitstream/handle/1969.1/547/etd-tamu-2004A-SENG-Eppes-1.pdf?sequence=1

53. 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(6):414–22.

54. Kadefors R, Forsman M. Ergonomic evaluation of complex work: a participative approach employing video-computer interaction, exemplified in a study of order picking. Int J Ind Ergonomics. 2000;25(4):435–45.

55. Forsman M, Stridqvist J, Persson O, Grundell L, Nyström A-K, Östman C. A video-based method for ergonomic evaluation – now with a checklist based on the Swedish ergonomic regulations. In: Saarela KL, Nygård C-H, Lusa S, editors. 38th Annual Nordic Ergonomics Society Congress: Promoting wellbeing in modern society; 2006; Hämeenlinna, Finland: PK-Paino Oy, Tampere, Finland; 2006. p. 196–200.

56. Kee D, Karwowski W. LUBA: an assessment technique for postural loading on the upper body based on joint motion discomfort and maximum holding time. Appl Ergon. 2001;32(4):357–66.

57. Chung MK, Lee I, Kee D, Kim SH. A postural workload evaluation system based on a macro-postural classification. Hum Factors Ergon Manuf. 2002;12(3):267–77.

58. Chung MK, Lee I, Kee D. Quantitative postural load assessment for whole body manual tasks based on perceived discomfort. Ergonomics. 2005;48(5):492–505.

59. Health and Safety Executives. Upper limb disorders in the workplace. 2 ed. Sudbury (United Kingdom): HSE Books; 2002.

60. McAtamney L, Corlett N. RULA: a survey method for the investigation of work-related upper limb disorders. Appl Ergon. 1993;24(2):91–9.

61. McAtamney L, Corlett N. Rapid upper limb assessment. In: Stanton N, Brookhuis K, Hedge A, Salas E, Hendrick HW, editors. Handbook of human factors and ergonomics methods. Boca Raton (FL): CRC Press; 2005. p 7:1–11.

62. Drinkaus P, Sesek R, Bloswick D, Bernard T, Walton B, Joseph B, et al. Comparison of ergonomic risk assessment outputs from Rapid Upper Limb Assessment and the Strain Index for tasks in automotive assembly plants. Work. 2003;21(2):165–72.

63. Jones T, Kumar S. Comparison of ergonomic risk assessments in a repetitive high-risk sawmill occupation: Saw-filer. Int J Ind Ergonomics. 2007;37(9–10):744–53.

64. Fountain LJ. Examining RULA's postural scoring system with selected physiological and psychophysiological measures. Int J Occup Saf Ergon. 2003;9(4):383–92.

65. Shuval K, Donchin M. Prevalence of upper extremity musculoskeletal symptoms and ergonomic risk factors at a Hi-Tech company in Israel. Int J Ind Ergon. 2005;35(6):569–81.

66. Breen R, Pyper S, Rusk Y, Dockrell S. An investigation of children's posture and discomfort during computer use. Ergonomics. 2007;50(10):1582–92.

67. Stetson DS, Keyserling WM, Silverstein BA, Leonard JA. Observational analysis of the hand and wrist: A pilot study. Appl Occup Environ Hyg. 1991;6(11):927–37.

68. Burt S, Punnett L. Evaluation of interrater reliability for posture observations in a field study. Appl Ergon. 1999;30(2):121–35.

69. Keyserling WM, Stetson DS, Silverstein BA, Brouwer ML. A checklist for evaluating ergonomic risk factors associated with upper extremity cumulative trauma disorders. Ergonomics. 1993;36(7):807–31.

70. Ketola R, Toivonen R, Viikari-Juntura E. Interobserver repeatability and validity of an observation method to assess physical loads imposed on the upper extremities. Ergonomics. 2001;44(2):119–31.

71. Moore JS, Garg A. The Strain Index: a proposed method to analyze jobs for risk of distal upper extremity disorders. Am Ind Hyg Assoc J. 1995;56(5):443–58.

72. Moore JS, Vos GA. The Strain Index. In: Stanton N, Brookhuis K, Hedge A, Salas E, Hendrick HW, editors. Handbook of human factors and ergonomics methods. Boca Raton (FL): CRC Press; 2005. p 9:1–5.

73. Lee EC, Rafiq A, Merrell R, Ackerman R, Dennerlein JT. Ergonomics and human factors in endoscopic surgery: a comparison of manual vs telerobotic simulation systems. Surg Endosc. 2005;19(8):1064–70.

74. 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(4):381–92.

75. Spielholz P, Bao S, Howard N, Silverstein B, Fan J, Smith C, et al. Reliability and validity assessment of the hand activity level threshold limit value and strain index using expert ratings of mono-task jobs. J Occup Environ Hyg. 2008;5(4):250–7.

76. Knox K, Moore JS. Predictive validity of the Strain Index in turkey processing. J Occup Environ Med. 2001;43(5):451–62.

77. Rucker N, Moore JS. Predictive validity of the strain index in manufacturing facilities. Appl Occup Environ Hyg. 2002;17(1):63–73.

78. Drinkaus P, Bloswick DS, Sesek R, Mann C, Bernard T. Job level risk assessment using task level strain index scores: a pilot study. Int J Occup Saf Ergon. 2005;11(2):141–52.

79. Stephens JP, Vos GA, Stevens EM Jr, Moore JS. Test-retest repeatability of the Strain Index. Appl Ergon. 2006;37(3):275–81.

80. Stevens EM Jr, Vos GA, Stephens JP, Moore JS. Inter-rater reliability of the strain index. J Occup Environ Hyg. 2004;1(11):745–51.

81. Armstrong T. The ACGIH TLV for hand activity level. In: Marras WS, Karwowski W, editors. Fundamentals and assessment tools for occupational ergonomics. Boca Raton (FL): CRC Press; 2006. p 41:1–14.

82. Latko WA, Armstrong TJ, Foulke JA, Herrin GD, Rabourn RA, Ulin SS. Development and evaluation of an observational method for assessing repetition in hand tasks. Am Ind Hyg Assoc J. 1997;58(4):278–85.

83. Latko WA, Armstrong TJ, Franzblau A, Ulin SS, Werner RA, Albers JW. Cross-sectional study of the relationship between repetitive work and the prevalence of upper limb musculoskeletal disorders. Am J Ind Med. 1999;36(2):248–59.

84. Franzblau A, Armstrong TJ, Werner RA, Ulin SS. A cross-sectional assessment of the ACGIH TLV for hand activity level. J Occup Rehabil. 2005;15(1):57–67.

85. Gell N, Werner RA, Franzblau A, Ulin SS, Armstrong TJ. A longitudinal study of industrial and clerical workers: Incidence of carpal tunnel syndrome and assessment of risk factors. J Occup Rehabil. 2005;15(1):47–55.

86. Werner RA, Franzblau A, Gell N, Hartigan A, Ebersole M, Armstrong TJ. Predictors of persistent elbow tendonitis among auto assembly workers. J Occup Rehabil. 2005;15(3):393–400.

87. Werner RA, Franzblau A, Gell N, Hartigan AG, Ebersole M, Armstrong TJ. Incidence of carpal tunnel syndrome among automobile assembly workers and assessment of risk factors. J Occup Environ Med. 2005;47(10):1044–50.

88. Werner RA, Franzblau A, Gell N, Hartigan AG, Ebersole M, Armstrong TJ. Risk factors for visiting a medical department because of upper-extremity musculoskeletal disorders. Scand J Work Environ Health. 2005;31(2):132–7.

89. Werner RA, Franzblau A, Gell N, Ulin SS, Armstrong TJ. Predictors of upper extremity discomfort: a longitudinal study of industrial and clerical workers. J Occup Rehabil. 2005;15(1):27–35.

90. Violante FS, Armstrong TJ, Fiorentini C, Graziosi F, Risi A, Venturi S, et al. Carpal tunnel syndrome and manual work: a longitudinal study. J Occup Environ Med. 2007;49(11):1189–96.

91. Ebersole ML, Armstrong TJ. Analysis of an observational rating scale for repetition, posture, and force in selected manufacturing settings. Hum Factors. 2006;48(3):487–98.

92. Occhipinti E, Colombini D. Proposta di un indice sintetico per la valutazione dell'esposizione a movimenti ripetitive degli arti superiori (OCRA index) [Proposal of a concise index for the evaluation of the exposure to repetitive movements of the upper extremity (OCRA index)]. Med Lav. 1996;87(6):526–48.

93. Colombini D. An observational method for classifying exposure to repetitive movements of the upper limbs. Ergonomics. 1998;41(9):1261–89.

94. Occhipinti E. OCRA: a concise index for the assessment of exposure to repetitive movements of the upper limbs. Ergonomics. 1998;41(9):1290–311.

95. Occhipinti E, Colombini D. The occupational repetitive action (OCRA) methods: OCRA index and OCRA checklis. In: Stanton N, Brookhuis K, Hedge A, Salas E, Hendrick HW, editors. Handbook of human factors and ergonomics methods. Boca Raton (FL): CRC Press; 2005. p 15:1–14.

96. Apostoli P, Sala E, Gullino A, Romano C. Analisi comparata dell'applicazione di quattro metodi per la valutazione del rischio biomeccanico per l'arto superiore [Comparative analysis of the use of 4 methods in the evaluation of the biomechanical risk to the upper limb]. G Ital Med Lav Ergon. 2004;26(3):223–41.

97. Occhipinti E, Colombini D. Updating reference values and predictive models of the OCRA method in the risk assessment of work-related musculoskeletal disorders of the upper limbs. Ergonomics. 2007;50(11):1727–39.

98. Occhipinti E, Colombini D. Metodo OCRA: aggiornarnento dei valori di riferimento e dei modelli di previsione della frequenza di patologie muscolo-scheletriche correlate al lavoro degli arti superiori (UL-WMSDs) in popolazioni lavorative esposte a movimenti e sforzi ripetuti degli arti superiori [The OCRA method: updating of reference values and prediction models

of occurrence of work-related musculo-skeletal diseases of the upper limbs (UL-WMSDs) in working populations exposed to repetitive movements and exertions of the upper limbs]. Med Lav. 2004;95(4):305–19.

99. Waters TR, Putz-Anderson V, Garg A, Fine LJ. Revised NIOSH equation for the design and evaluation of manual lifting tasks. Ergonomics. 1993;36(7):749–76.

100. Waters TR, Putz-Anderson V, Garg A. Applications manual for the revised NIOSH lifting equation [Internet]. Cincinnati (OH): National Institute for Occupational Safety and Health; 1994 [cited 15 October 2009]. Available from: http://www.cdc.gov/niosh/docs/94-110/

101. Waters T. Revised NIOSH lifting equation. In: Marras WS, Karwowski W, editors. Fundamentals and assessment tools for occupational ergonomics. Boca Raton (FL): CRC Press; 2006. p 46:1–28.

102. Waters TR, Lu ML, Occhipinti E. New procedure for assessing sequential manual lifting jobs using the revised NIOSH lifting equation. Ergonomics. 2007;50(11):1761–70.

103. Lavender SA, Oleske DM, Nicholson L, Andersson GB, Hahn J. Comparison of five methods used to determine low back disorder risk in a manufacturing environment. Spine. 1999;24(14):1441–8.

104. Marklin RW, Wilzbacher JR. Four assessment tools of ergonomics interventions: case study at an electric utility's warehouse system. Am Ind Hyg Assoc J. 1999;60(6):777–84.

105. Lee TH. Psychophysically determined asymmetrical lifting capabilities for different frequencies and containers. Ind Health. 2005;43(2):337–40.

106. Jones T, Strickfaden M, Kumar S. Physical demands analysis of occupational tasks in neighborhood pubs. Appl Ergon. 2005;36(5):535–45.

107. Dempsey PG, Burdorf A, Fathallah FA, Sorock GS, Hashemi L. Influence of measurement accuracy on the application of the 1991 NIOSH equation. Appl Ergon. 2001;32(1):91–9.

108. Wang MJ, Garg A, Chang YC, Shih YC, Yeh WY, Lee CL. The relationship between low back discomfort ratings and the NIOSH lifting index. Hum Factors. 1998;40(3):509–15.

109. Waters TR, Baron SL, Piacitelli LA, Anderson VP, Skov T, Haring-Sweeney M, et al. Evaluation of the revised NIOSH lifting equation: a cross-sectional epidemiologic study. Spine. 1999;24(4):386–95.

110. Marras WS, Fine LJ, Ferguson SA, Waters TR. The effectiveness of commonly used lifting assessment methods to identify industrial jobs associated with elevated risk of low-back disorders. Ergonomics. 1999;42(1):229–45.

111. Sesek R, Gilkey D, Drinkaus P, Bloswick DS, Herron R. Evaluation and quantification of manual material handling risk factors. Int J Occup Safety Ergon. 2003;9(3):271–87.

112. Waters TR, Baron SL, Kemmlert K. Accuracy of measurements for the revised NIOSH lifting equation. Appl Ergon. 1998;29(6):433–8.

113. Arbouw Foundation. Guidelines on physical workload for the construction industry [Internet]. Amsterdam: Arbouw Foundation; 1997 [cited 15 October 2009]. Available from: http://www.lhsfna.org/files/ARBOUW_Guidelines.pdf

114. Department of Labour Te Tar Mahi. Code of practice for manual handling [Internet]. Wellington (New Zealand): Occupational Safety and Health Service of the Department of Labour, the Accident Compensation Corporation; 2001 [cited 15 October 2009]. Available from: http://www.osh.dol.govt.nz/order/catalogue/pdf/manualcode.pdf

115. Coyle A. Comparison of the Rapid Entire Body Assessment and the New Zealand Manual Handling 'Hazard Control Record', for assessment of manual handling hazards in the supermarket industry. Work. 2005;24(2):111–6.

116. Gueensland Government. Manual Tasks Code of Practice 2000 [Internet]. Brisbane (Australia): Workplace Health and Safety, Department of Justice and Attorney-General [cited 15 October 2009]. Available from: http://www.deir.qld.gov.au/workplace/law/codes/manualtasks/index.htm

117. Burgess-Limerick R, Straker L, Pollock C, Egeskov R. Manual Tasks Risk Assessment Tool (ManTRA) V 2.0 [Internet]. Burgess-Limerick, Egeskov, Straker, and Pollock; 2000–2004 [cited 15 October 2009]. Available from: www.burgess-limerick.com/download/mantra2.pdf

118. Burgess-Limerick R. Ron Cumming Memorial Lecture 2004: a tale of two acronyms: PErforM and ManTRA. Ergonomics. 2004;18(4):10–3.

119. Monnington SC, Pinder AD, Quarrie C. Development of an inspection tool for manual handling risk assessment [Internet]. Sheffield (United Kingdom): Health and Safety Laboratory; 2002 [cited 15 October 2009]. Report 30. Available from: http://www.hse.gov.uk/research/hsl_pdf/2002/hsl02-30.pdf

120. Pinder AD. Benchmarking of the Manual Handling assessment Charts (MAC) [Internet]. Sheffield (United Kingdom): Health & Safety Laboratory; 2002 [cited 15 October 2009]. Report 31. Available from: http://www.hse.gov.uk/research/hsl_pdf/2002/hsl02-31.pdf

121. Care B, Monnington CQSC. Testing and improving the usability of the Manual handling Assessment Chart (MAC) [Internet]. Health and Safety Executive; 2002 [cited 15 October 2009].Available from: http://www.hse.gov.uk/msd/mac/pdf/usability.pdf

122. Tapley SE. Reliability of Manual handling Assessment Charts (MAC) developed for health and safety inspectors in the UK: a field study [Internet]. Sheffield (United Kingdom): Health and Safety Executive; 2002 [cited 15 October 2009]. Available from: http://www.hse.gov.uk/msd/mac/pdf/reliability.pdf

123. Lee D, Ferreira J. Reliability and usability evaluation of the Manual handling Assessment Charts (MAC) for use by non-regulatory professionals [Internet]. Sheffield (United Kingdom): Health & safety laboratory; 2003. Report 19 [cited 15 October 2009]. Available from: http://www.hse.gov.uk/research/hsl_pdf/2003/hsl03-19.pdf

124. Keyserling WM. Analysis of manual lifting tasks: a qualitative alternative to the NIOSH work practices guide. Am Ind Hyg Assoc J. 1989;50(3):165–73.

125. Marras WS, Hamrick C. The ACGIH TLV for low back risk. In: Marras WS, Karwowski W, editors. Fundamentals and

assessment tools for occupational ergonomics. Boca Raton (FL): CRC Press; 2006. p 50:1–15.

126. Village J, Trask C, Luong N, Chow Y, Johnson P, Koehoorn M, et al. Development and evaluation of an observational Back-Exposure Sampling Tool (Back-EST) for work-related back injury risk factors. Appl Ergon. 2008;40(3):538–44.

127. Fallentin N, Viikari-Juntura E, Wærsted M, Kilbom Å. Evaluation of physical workload standards and guidelines from a Nordic perspective. Scand J Work Environ Health. 2001;27 suppl 2:1–52.

128. Dempsey PG, McGorry RW, Maynard WS. A survey of tools and methods used by certified professional ergonomists. Appl Ergon. 2005;36(4):489–503.

129. Lowe BD. Accuracy and validity of observational estimates of wrist and forearm posture. Ergonomics. 2004;47(5):527–54.

130. Keyserling WM. Postural analysis of the trunk and shoulders in simulated real time. Ergonomics. 1986;29(4):569–83.

131. Denis D, Lortie M, Bruxelles M. Impact of observers' experience and training on reliability of observations for a manual handling task. Ergonomics. 2002;45(6):441–54.

132. Andrews DM, Arnold TA, Weir PL, vanWyk PM, Callaghan JP. Errors associated with bin boundaries in observation-based posture assessment methods. Occup Ergon. 2008;8(1):11–25.

133. Armstrong TJ, Buckle P, Fine LJ, Hagberg M, Jonsson B, Kilbom Å, et al. A conceptual model for work-related neck and upper-limb musculoskeletal disorders [review]. Scand J Work Environ Health. 1993;19:73–84.

134. Winkel J, Mathiassen SE. Assessment of physical work load in epidemiologic studies: concepts, issues and operational considerations. Ergonomics. 1994;37(6):979–88.

135. Wells R, Van Eerd D, Hägg G. Mechanical exposure concepts using force as the agent [review]. Scand J Work Environ Health. 2004;30(3):179–90.

136. Mathiassen SE. Diversity and variation in biomechanical exposure: what is it, and why would we like to know?. Appl Ergon. 2006;37(4 special issue):419–27.

137. Wells R, Mathiassen SE, Medbo L, Winkel J. Time – a key issue for musculoskeletal health and manufacturing. Appl Ergon. 2007;38(6):733–44.

138. Moore JS, Rucker NP, Knox K. Validity of generic risk factors and the strain index for predicting nontraumatic distal upper extremity morbidity. AIHA J. 2001;62(2):229–35.

139. Werner RA, Franzblau A, Gell N, Ulin SS, Armstrong TJ. A longitudinal study of industrial and clerical workers: predictors of upper extremity tendonitis. J Occup Rehabil. 2005;15(1):37–46.

140. Moher D, Cook DJ, Eastwood S, Olkin I, Rennie D, Stroup DF. Improving the quality of reports of meta-analyses of randomised controlled trials: the QUOROM statement. Lancet. 1999;354(9193):1896–900.

141. Stroup DF, Berlin JA, Morton SC, Olkin I, Williamson GD, Rennie D, et al. Meta-analysis of observational studies in epidemiology: a proposal for reporting. JAMA. 2000;283(15):2008–12.

142. Higgins J, Green S, editors. Cochrane Handbook for Systematic Reviews of Interventions Version 5.0.2 [updated September 2009][Internet]. The Cochrane Collaboration [cited 15 October 2009]. Available from: http://www.cochrane.org/resources/handbook/

143. Westgaard RH, Winkel J. Guidelines for occupational musculoskeletal load as a basis for intervention: a critical review. Appl Ergon. 1996;27(2):79–88.

144. Wells R, Norman R, Neumann P, Andrews D, Frank J, Shannon H, et al. Assessment of physical work load in epidemiologic studies: common measurement metrics for exposure assessment. Ergonomics. 1997;40(1):51–61.

145. van der Beek AJ, Frings-Dresen MH. Assessment of mechanical exposure in ergonomic epidemiology. Occup Environ Med. 1998;55(5):291–9.

146. Moher D, Schulz KF, Altman DG. The CONSORT statement: revised recommendations for improving the quality of reports of parallel group randomized trials. BMC Med Res Methodol. 2001;1:2.

147. Vandenbroucke JP, von Elm E, Altman DG, Gøtzsche PC, Mulrow CD, Pocock SJ, et al. Strengthening the Reporting of Observational Studies in Epidemiology (STROBE): explanation and elaboration. PLoS Med. 2007;4(10):e297.

148. Burdorf A. Sources of variance in exposure to postural load on the back in occupational groups. Scand J Work Environ Health. 1992;18(6):361–7.

149. van der Beek AJ, Kuiper JI, Dawson M, Burdorf A, Bongers PM, Frings-Dresen MH. Sources of variance in exposure to nonneutral trunk postures in varying work situations. Scand J Work Environ Health. 1995;21(3):215–22.

150. Burdorf A, van Riel M. Design of strategies to assess lumbar posture during work. Int J Ind Ergonomics. 1996;18(4):239–49.

151. Hoozemans MJM, Burdorf A, Van Der Beek AJ, Frings-Dresen MHW, Mathiassen SE. Group-based measurement strategies in exposure assessment explored by bootstrapping. Scand J Work Environ Health. 2001;27(2):125–32.

152. Mathiassen SE, Möller T, Forsman M. Variability in mechanical exposure within and between individuals performing a highly constrained industrial work task. Ergonomics. 2003;46(8):800–24.

153. Paquet V, Punnett L, Woskie S, Buchholz B. Reliable exposure assessment strategies for physical ergonomics stressors in construction and other non-routinized work. Ergonomics. 2005;48(9):1200–19.

154. Svendsen SW, Mathiassen SE, Bonde JP. Task based exposure assessment in ergonomic epidemiology: a study of upper arm elevation in the jobs of machinists, car mechanics, and house painters. Occup Environ Med. 2005;62(1):18–27.

155. Hansson GÅ, Arvidsson I, Ohlsson K, Nordander C, Mathiassen SE, Skerfving S, et al. Precision of measurements of physical workload during standardised manual handling. Part II: Inclinometry of head, upper back, neck and upper arms. J Electromyogr Kines. 2006;16(2):125–36.

156. Bao S, Howard N, Spielholz P, Silverstein B. Two posture analysis approaches and their application in a modified Rapid

Upper Limb Assessment evaluation. Ergonomics. 2007;50 (12):2118–36.

157. Neumann WP, Wells RP, Norman RW, Frank J, Shannon H, Kerr MS. A posture and load sampling approach to determining low-back pain risk in occupational settings. Int J Ind Ergonomics. 2001;27(2):65–77.

158. Mathiassen SE, Burdorf A, Van der Beek AJ, Hansson G-Å. Efficient one-day sampling of mechanical job exposure data – a study based on upper trapezius activity in cleaners and office workers. Am Ind Hyg Assoc J. 2003;64(2):196–211.

159. Barrero LH, Katz JN, Dennerlein JT. Validity of self-reported mechanical demands for occupational epidemiologic research of musculoskeletal disorders [review]. Scand J Work Environ Health. 2009;35(4):245–60.

Received for publication: 2 July 2009