

Assessment of physical exposure in relation to work-related musculoskeletal disorders – what information can be obtained from systematic observations?

by Åsa Kilbom, MD¹

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. Establishing an association between physical work load and work-related musculoskeletal disorders requires a quantitative assessment of exposure to the main risk factors. Commonly used methods are questionnaires, diaries, interviews, systematic observations, and direct measurements. While questionnaires provide subjective information and have low reliability, measurements are technically sophisticated, expensive, and unable to identify some important risk factors. Systematic observations offer a compromise. The observation methods devised during the past 15 years are described, together with their reliability and validity. Most of them use duration or frequency of certain postures or events as the main measure of exposure. Their ability to quantify the main risk factors posture, manual handling, and repetitive work differs. All available methods have drawbacks and have been used to a limited extent in epidemiologic studies. There have been improvements however, and it is concluded that observation methods will have a place in epidemiologic studies on musculoskeletal disorders.

KEY TERMS — epidemiology, exposure, manual handling, musculoskeletal, observation methods, posture, repetitive work, review.

In spite of the large number of experimental and epidemiologic studies that have appeared during the past 10 years, the etiology and risk factors of work-related musculoskeletal disorders remain insufficiently known. Hypotheses range from purely physical work load factors, to social and organizational factors at work, to individual characteristics like strength, personality, and psychological traits. Although physical work load has been considered important for many years, a critical review of the scientific support for this statement demonstrates many weaknesses. It is true that certain occupations, known to be physically demanding, have a very high prevalence and incidence of musculoskeletal disorders. However, many employees in these jobs also suffer from inadvertent psychosocial exposures. Moreover, a job title does not allow conclusions concerning the quality and quantity of physical work load and therefore does not allow the exposure-response relationship to be established and gives little guidance for prevention. Clearly there is a need for a more-detailed and accurate description of physical work loads, for use in epidemiologic studies of work-related musculoskeletal disorders, if the causal and aggravating factors are to be identified quantitatively (1, 2). Moreover it must be possible to quantitate

physical work loads when the effects of controlled interventions at the workplace are assessed.

The methods available for measuring physical work load can be categorized into a spectrum ranging from direct measurements, to observations, interviews and diaries, to questionnaires. The choice of a specific method depends on the resources available and the accuracy required of the data.

Direct measurements include electromyographic recordings, posture and movement recordings by goniometers, inclinometers, accelerometers, and optoelectronic devices (eg, sel-spot, CODA) (3—10). These methods are quantitative and highly accurate, but expensive due to the demands for sophisticated equipment and time-consuming calibration, registration, and analysis. Because of this limitation only a certain number of body regions and subjects can be assessed. Moreover, monitoring devices carried by the worker can hinder the worker and thereby influence work methods. Optoelectronic recordings can only be used for the limb and trunk movements of an otherwise stationary worker; furthermore the recording equipment requires even lighting and a relatively large space.

At the other end of the spectrum, questionnaires have been frequently used for exposure assessment due to their low cost. The use of questionnaires, diaries, and interview techniques has the potential of studying cumulative exposure over time, which is an important parameter usually not available with direct measurements. However, the relatively low reliabil-

¹ National Institute of Occupational Health, Solna, Sweden.

Reprint requests to: Dr Å Kilbom, National Institute of Occupational Health, S-171 84 Solna, Sweden.

ity and validity of the questionnaires developed so far make their use debatable, especially for retrospective assessments. Thus it is usually considered that only dichotomous exposure assessments can be obtained (11).

Recently, diaries and interviews have been demonstrated to have considerably higher validity than questionnaires. However, the experience so far is limited, and the subjective nature of the information cannot be disregarded (12, Wiktorin, personal communication). Thus observation methods may offer a compromise between the high cost of direct measurements and the low validity and the subjectivity of questionnaires, diaries, and interviews. The objective of this review is to describe and evaluate the potential usefulness of systematic observation techniques in the quantitative assessment of physical work load for epidemiologic studies of work-related musculoskeletal disorders, as well as for intervention.

What are the risk factors?

Selecting a method for assessing physical work load depends on the ability of the method to quantitate risk factors believed to cause or aggravate musculoskeletal disorders. Although the risk factors identified for disorders of different body regions differ slightly, it appears that the broad categories of static posture, manual handling, repetitive work, and vibration are frequently identified in reviews (2, 13–16). The pathophysiological mechanisms of exposure to these risk factors are biologically plausible, although insufficiently known (17), and they have been identified in several epidemiologic studies. However, there still remains considerable doubt concerning the quantitative importance of physical risk factors.

Several other risk factors, for example, high peak loads, high speed or the acceleration of movements, and high demands for precision, have been suggested (8, 18–20). They all require more scientific support and need to be quantified in epidemiologic studies before the risk factors can be satisfactorily coupled with musculoskeletal disorders.

Review of available methods

In table 1 nineteen observation methods for quantitatively assessing physical exposures have been summarized. The main quantitative measure of exposure is the duration or frequency of a certain posture or activity. Therefore all of the observations are performed over time. In the portable ergonomic observation (PEO) method, the observation is supplemented with an interview in which all major work tasks during one week are recorded with their proportional duration (Fransson-Hall et al, unpublished manuscript). This procedure enables the computer program to calculate the total exposure profile. All of the

methods selected for inclusion in table 1 assess posture, and some of them also assess manual handling and repetitive work. It is generally acknowledged that exposure to vibration is monitored better via direct measurements; hence no observation methods have been designed specifically for vibration assessment.

The methods selected for table 1 have all appeared since 1979. Thus methodological development has been very fast. However, already before 1979, some methods defining postures and movements were described (22, 23). The Benesh notation system (22) was developed for dancing and is highly specific for the flow of movement in ballet. The system is difficult to transform into digital data and was not developed with the aim of identifying musculoskeletal risk factors. The "posturegram" method developed by Priehl (23) gives accurate information on the location and movement of body parts in three coordinate planes. However, the system is time-consuming and therefore has not been used for studying changes in posture over time. A similar method is posture targeting (table 1), which is faster and can therefore be applied for a sequence of postures (24).

Postures have also been described for specific work situations and linked to workstation layout (eg, 24–29). The aim of these methods are ergonomic (ie, studying the fit between anthropometry, posture, and workstation and work task can help identify poor design for subsequent improvements). In general, the postural analysis of these methods concentrates on one or a few "critical" postures or makes a summary evaluation of the range of postures during a certain task. Thus risk factors are not assessed in real time or by time sampling, and therefore a quantitative evaluation of risk factors is impossible. Ergonomic methods often have an in-built "acceptability" limit, deduced from previous research or consensus agreements. In this respect too they differ from methods used in musculoskeletal epidemiology, for which one of the aims is to identify, with higher precision, the limit(s) of risk increase. It is true, however, that the borderline between observation methods used for ergonomic purposes and those used for musculoskeletal epidemiology is relatively vague and that the methods overlap. For this reason, some of the methods also used for ergonomic purposes — the Ovako work posture analyzing system (OWAS) (30), posture targeting (24), ARBAN (31) — have been included in table 1. Of these, the OWAS method is the most widely used, and many case studies of ergonomic applications have been reported (32–36).

Generality of the method

The methods presented in table 1 have been developed for three types of epidemiologic applications, characterized by the range of job titles and exposures that can be assessed (ie, the generality of the method). First, there are general methods for assessing exposure, applicable to a large range of occupational

Table 1. Summary of methods used for the systematic observation of physical exposures during the last 15 years. The methods included in the table all concentrate on postures.

Method	Principle	Scale for posture classification	Posture			
			Arms	Back	Other body regions	Whole body
Armstrong et al, 1979 (65)	Pencil and paper, real time via film, three frames per second	Interval	Shoulder: 4—6 categories of flexion-extension, abduction-adduction and rotation; elbow: 3—4 categories of flexion-extension and rotation; wrist: 3—5 categories of flexion-extension and deviation		Fingers: six categories of grips	
Baty et al, 1986 (46)	Pencil and paper, time sampling, 15-s intervals	Dichotomous	Reach	Twist, bend > 15 degrees		Stand, walk, squat, sit, kneel, support, brace
van der Beek et al, 1992 (48)	Computerized, time sampling and real time via video (based on ROTA principle), 15-s interval	Interval and dichotomous	Both <60 degrees, one >60 degrees, both >60 degrees "pump"	Flexion or rotation: 0—15, 15—45, 45—75 and >75 degrees	Legs both 0—15 degrees, one >15 degrees, both >15 degrees, one loose	Stand, sit, walk, kneel, squat
Chang et al, 1987 (43)	Video, time sampling, two frames per 15 s every 2 min; continuous, 60 frames per second for 1 min; side and oblique lateral view	Continuous	Shoulder flexion, rotation, abduction, elbow flexion	Flexion	Neck flexion	
Corlett et al, 1979 (24) Posture targetting	Pencil and paper, time sampling, variable intervals	Continuous in two planes, horizontal and vertical	Right and left upper arm, forearm	Flexion, side-bending	Also head, thighs, calves	Strike, crank, squeeze, twist, wipe, walk
Feuerstein & Fitzgerald, 1992 (19)	Video (frontal view), pencil and paper, real time; frequency of movements and events	Dichotomous	Wrist-hand deviations			
Foreman & Troup, 1987 (60); Foreman et al, 1988 (40)	Computerized, direct or video-based, real time	Dichotomous		Flexion > 15 degrees, stoop		Stand, sit, squat, walk, kneel
Fransson-Hall et al (unpublished manuscript) Leskinen et al (unpublished manuscript) PEO	Computerized, direct or via video, real time; frequency, duration of postures and activities	Interval and dichotomous	Right and left, hands above or below shoulder	Flexion 20—60 and >60 degrees, twist >45 degrees	Neck flexion, >20 degrees; neck rotation, >45 degrees	Kneel, squat
Fransson-Hall et al (unpublished manuscript) Hand-PEO	Computerized, real-time observation via video; frequency and duration of hand-grip and activities	Interval and dichotomous	Precision, power grip			

Manual materials handling		Other categories of exposure	Reliability	Internal validity	Developed and applied for	Comments
Lift, carry	Push, pull					
.	.	Supplemented with electro-myographic and MTM analysis	.	.	Poultry processing	Used for ergonomic intervention
Lift, carry	Push, pull	.	Interobserver: >80% agreement in frequent postures (bend, stand, walk, sit); poor agreement some infrequent postures (lift, brace, carry, reach, twist)	Comparison with accelerometer yielded poor agreement — due to accelerometer artefacts in fast movements?	Nursing	Poor agreement with questionnaire
<2, 2—10, >10 kg	<2, 2—175, >175 kg	.	Interobserver: kappa ~0.80 except trunk, arm, legs posture; improved by training; when MMH excluded, all posture kappas >0.50	Compared with optoelectronic measurements: error <5 degrees in trunk flexion at static postures	Delivery truck drivers	Modified version using two observers suggested, one observing postures, one observing activities and MMH in real time
.	.	Angular velocities calculated from continuous recordings	.	.	Stationary, seated work (artists)	Larger variation in back and neck postures for artists doing large than for those doing small drawings
Hold	Push, pull	Always supplemented with discomfort rating of 15 body regions	Intraobserver coefficient of correlation 0.67—0.88 for body posture	Compared with frame-by-frame film: good for head and trunk, unsatisfactory for legs and lower arms	Industrial work	Very accurate for postures but does not record frequency of postures, less suitable for dynamic work
.	.	Analysis of hand and arm activities: number of rest breaks, hand excursions, high-impact hand contacts, pace and smoothness of hand and finger movements	Interobserver: intraclass correlation >0.89 for all except frequency of wrist deviation (0.76)	.	Sign language interpreters	Method concentrates on pacing, extreme postures and dynamics of hand and wrist rather than on static postures
Lift, carry	Push, pull	.	Interobserver agreement: high for frequency, except squatting, high for duration except stoop and lift; Intraobserver: high (>90% agreement) for duration and frequency for trained observer; stooping, lifting <80% for untrained observer	.	Nursing	Manual handling and posture codes combined in two-letter mnemonic
Lift 1—5, 6—15, 16—45, >45 kg; lifts above shoulder; below knee, twisted back noted	1—5, 6—15, 16—45, >45 kg (See lift, carry)	Repetitive work: pro-supination, wrist flexion, wrist extension, finger flexion, finger extension; "extra events" recorded; interview to calculate weekly exposure	Interobserver: no systematic difference in duration except neck and manual handling; low correlation for frequency of neck, lift and MMH (ANOVA); intraobserver: high agreement (ANOVA)	Sel-spot and video recordings: high agreement for postures (except neck) and long-lasting repetitive activities	All jobs	Interviews to calculate weekly exposure to all categories; observation of all categories simultaneously, or one at a time
Hold	.	Hand tools x 4, machine tools x 2, hand as hitting tool, manual assembly, hand as support	.	.	Developed for assembly work, applicable for many jobs	Separate observations right—left hand

(continued)

Table 1. Continued

Method	Principle	Scale for posture classification	Posture			
			Arms	Back	Other body regions	Whole body
Harber et al, 1987 (47)	Pencil and paper, real time; duration and frequency	Ordinal and dichotomous	·	·	·	Static posture: above shoulder, below waist, sagittal or symmetric, upright, squat
Holzmann, 1982 (31) ARBAN	Computerized, film-based, time sampling, 1- to 2-s intervals	Interval (Borg's 10-level scale with ratio properties)	Right and left 10 levels Borg scale	Ten levels Borg scale	Also head, legs, 10-level Borg scale	·
Karhu et al, 1977 (30), Kivi & Mattila, 1991 (33) OWAS	Pencil and paper, computerized, time sampling (usually 30 s) or direct	Ordinal	Both hands on or below shoulders, one on or above, both on or above, modified versions of four arm postures	Straight, bent, straight and twisted, bent and twisted	Legs in seven categories, modified version; five head postures	Optional additional posture: sitting
Keyserling, 1986 (39)	Computerized, real time observation via video; frequency and duration of postures, number of postural changes	Interval and dichotomous	Right and left shoulder flexion-abduction ≤ 45, 45–90, > 90 degrees	Standing: extension > 20 degrees; flexion-bending: neutral, 20–45 and > 45 degrees, twist or bend > 20 degrees	·	Lie, sit neutral, sit mild flexion, sit twisted or bent
Kilbom et al, 1986 (44) Jonsson et al, 1988 (21) VIRA	Computerized via two video recordings (rear and side), real time; frequency and duration of postures and movements	Interval and dichotomous	Shoulder elevation; upper arm abduction 0–30, 30–60, 60–90, > 90 degrees, upper arm flexion < 0, 0–30, 30–60, 60–90, > 90 degrees	·	Neck flexion > 20 degrees	·
Ridd et al, 1989 (37) ROTA	Computerized, direct or video-based, real time or time sampling	Optional	·	Optional definition of postures	·	Optional definition of postures
Rohmert & Landau, 1979 (38) AET	Pencil and paper, classification done at end of observation period	Ordinal	·	Stand straight, stand slightly bent, stand considerably bent	·	Sit straight, sit forward bent
Ryan, 1989 (42)	Pencil and paper, time sampling (10 s during 30 min)	Dichotomous	·	Flexion > 40 degrees, twisting > 20 degrees	·	Stand, walk, sit, kneel, squat
Stetson et al, 1991 (45)	Pencil and paper, direct or video, real time; frequency and duration of hand exertions	Dichotomous and interval	Wrist flexion, extension, ulnar deviation, pinch grip	·	·	·
Wickström et al, 1985 (18)	Pencil and paper, time sampling; observer 1: task and manual handling, observer 2: postures	Interval and dichotomous	·	Straight, bent 15–90 degrees, bent > 90 degrees, backwards, rotation	·	·

titles (30, 37, 38, Fransson-Hall, unpublished manuscript). The OWAS and AET (arbeitswissenschaftliche Erhebungsverfahren zur Tätigkeitsanalyse),

methods have been thoroughly documented, and reference data are available. Second, there are methods for a narrower group of occupations, for example,

Manual materials handling		Other categories of exposure	Reliability	Internal validity	Developed and applied for	Comments
Lift, carry	Push, pull					
Number of lifts >30 kg	•	Static—dynamic, patient—nonpatient contact	•	•	Nursing	•
"Dynamic stress" for six body regions, 10-level Borg scale	•	Total "ergonomic stress" = mean for each body region of posture, dynamic stress, isometric muscle stress and vibration or shock	•	Good agreement between expert panel and workers' ratings at high loads; observers overestimated low loads	Building industry	Analysis time-consuming, done in laboratory; "typical" work analyzed, not individual variation
<10, 10—20, >20 kg	•	Static—dynamic posture, combinations classified into four categories of acceptability; tasks can be recorded	Interobserver: coefficient of correlation 0.74—0.99, low for forward bend + twisting	•	Widely used in many branches of industry	Less suitable for fast dynamic activities
•	•	•	Interobserver agreement (inexperienced): moderate for left shoulder, less for trunk and right shoulder; intraobserver: frequency and duration: high for trunk, less consistency for nonneutral shoulder	Compared with frame-by-frame: good agreement except for twisted/bent trunk	Automobile assembly	Separate replay for each joint: 1. trunk; 2 left shoulder; 3 right shoulder
•	•	•	Interobserver agreement (inexperienced): all observations except percentage of work cycle time at rest and 0—30 degrees of flexion good; intraobserver: good except percentage of work cycle time 0—30 degrees flexion	•	Stationary, preferably seated work	No difference in work cycle time between morning—afternoon; large interindividual difference between workers with same task
Optional definition of manual handling activities	•	Task, workstation or tool recorded (nine categories)	•	•	Developed for all jobs; applied by van der Beek (48)	Versatile system where exposure categories can be selected by observer
Dynamic large, dynamic small muscle groups	•	Static postures, (three body regions) heavy dynamic (two regions), repetitive (three regions)	Interobserver: coefficient of correlation ~0.8	•	Widely used in Germany and Finland, for comparison between all jobs	Few posture classes, therefore crude; method also registers hygienic and work organizational factors; "typical" work analyzed, not individual variation
Hold, reach, lift	Push, pull	•	•	•	Supermarket workers	Strong association: symptoms feet and legs and duration of standing
•	•	Hand exertions as power tool use, use of high force or hand as striking tool	Interobserver, no significant differences (multiple regression)	•	Assembly, manufacturing	Direct analysis more accurate than video
5—20, >20 kg (weighing); horizontal distance <30, >30 cm; level of load: toe, hip, shoulder	•	Static—dynamic (back), work surface: good, intermediate, poor	•	•	Construction work	No difference in posture between workers in same occupation; minor accidents recorded by diary and telephone interview

manufacturing (24, 39), nursing (40, 46, 47), and construction work (18, 31). Some of these methods have features similar to the general methods and can

probably be applied to a wider range of occupations (39—41). The method presented by Keyserling (39) and posture targeting (24) have also been thorough-

ly documented. Finally, there are methods for one specific occupation, for example, sign language interpreters (19), supermarket workers (42), and artists (43), or one body region, generally the hand or shoulder (19, 44, 45, Fransson-Hall et al, unpublished manuscript). These methods have usually been documented only in conjunction with the epidemiologic study for which they were developed.

The advantage of the first group is that they permit comparisons between occupations. On the other hand, the necessity to record many risk factors simultaneously makes them relatively crude. One exception is the PEO method, which attempts to quantify a large number of risk factors, however at the cost of a somewhat lower reliability and validity of some variables (Leskinen et al, unpublished manuscript). The third group, and to some extent the second, concentrates on only a few risk factors, for example, wrist movements and handgrips (19, 45, Fransson-Hall et al, unpublished manuscript) or upper arm postures and movements in seated work (43, 44).

Risk factors assessed in observation methods

All of the methods in table 1 concentrate on postures, since it is one of the most commonly identified risk factors. Back and arm postures are registered by most methods, because of the common occurrence of back and shoulder musculoskeletal disorders. For categories of postural analysis, see the section on the design of scales.

In addition, nine methods also incorporate manual handling, either by observation only (24, 40, 42, 46, 47) or with supplementary measurement of the weights handled or forces exerted, usually categorized into two to four classes (30, 41, 48, Fransson-Hall, unpublished manuscript). Six of these methods distinguish between lifting-carrying, usually recorded as the number of actions per unit of time, and pushing-pulling. Only two methods record pushing-pulling in real time, which allows the time spent in these activities to be calculated (40, Fransson-Hall, unpublished manuscript). This may be an important parameter, since a case-referent study of reported occupational musculoskeletal disorders showed that the cases spent about three times longer pushing and pulling than the referents (49).

For a more-detailed and quantitative evaluation of lifting, the procedure recommended in the revised equation of the National Institute for Occupational Safety and Health (NIOSH) in the United States can be used (50). This method does not register the sequence of lifting over time, neither does it record pushing, pulling, carrying, or body postures. The revised equation includes the calculation of a lifting index (ie, the ratio between the load lifted and the recommended weight limit). This index may become an interesting measure of exposure to lifting, and it should be useful in epidemiologic studies when man-

ual lifting is identified as the major exposure hazard. In a study by Huang et al (51) a similar approach was used for a comparison between musculoskeletal disorders in two workplaces.

Only the PEO method (Fransson-Hall et al, unpublished manuscript) and the observation method used by Feuerstein & Fitzgerald (19) attempt to quantify the duration or frequency of repetitive movements of the forearm, wrist, and hand. Systematic observations of repetitive work have been performed successfully by Silverstein et al in several studies (52—54). A work task was defined as repetitive when work cycles were shorter than 30 s or when fundamental work cycles constituted more than 50% of the total work cycle. Each job was characterized with regard to its force demands into high or low force jobs, but no postures were systematically observed. The studies were performed in an industrial setting, and tasks with a high or low amount of repetition and a high or low amount of force could be observed reliably.

Two other methods record "dynamic stress," which may partly incorporate manual materials handling but is likely also to include dynamic activities without the exertion of force (31, 38). The OWAS method also permits the observer to record the static or dynamic conditions of the observation (30). Dynamic exercise is usually not included among risk factors for musculoskeletal disorders. This factor has been introduced into some methods because the methods were designed for jobs in which the worker was highly mobile, and it was important to distinguish between the fatiguing effect of static versus dynamic components.

The computerized methods presented by Ridd et al (37) and Fransson-Hall et al (PEO) (unpublished manuscript), as well as OWAS (30), also allow the observer to record the tasks (or special events). This is an important parameter if the observation aims at quantifying risk factors by task, for example, for subsequent intervention. Many other methods can be modified to record risk factors by tasks, but they were not primarily designed for this purpose.

In conclusion, among the major risk factors identified for musculoskeletal disorders, postures and manual handling are commonly assessed with observation methods. Few attempts have been made to quantify repetitive work, and a quantitative assessment of vibration requires measurements.

Reliability

The reliability of an observation method is the coherency of results when the observations are duplicated. Hence the intraobserver reliability is the extent of identical results obtained in repeated observations of the same work situation by the same observer, and the interobserver reliability is the extent of identical results obtained by several observers when observing the same work situation. The out-

come of reliability testing is highly dependent upon the methodology used. Several factors influence the reliability assessment of an observation method.

Basis for observations. Video recordings are the most frequently used in determinations of reliability, but it is also possible to use observations in three dimensions by utilizing a subject performing predefined tasks according to recorded instructions transmitted by a portable tape recorder with earphones (Leskinen et al, unpublished manuscript). Furthermore, the observations can be made from simple, stereotype movements or realistic work situations, and the temporal base varies from single operations to several work tasks. Naturally, the calculated reliability can be expected to decrease as the complexity of the study material increases.

Number of observers. Most studies use only one or two observers. However more information is obtained if several observers are used.

Potential learning effects. Intraobserver reliability can be influenced by learning effects (ie, the sequence of postures and activities can be memorized from one observation session to the next). This problem can be overcome if enough time is allowed between the observation sessions and, if possible, the studied sequences are alternated.

Number of simultaneously observed categories. The calculated reliability can be expected to decrease as the number of simultaneously observed categories increases.

Statistical methods. The choice of statistical method for reliability assessment is crucial and should be considered when investigations of reliability are planned, since it influences the number of observers, the number of observation sessions, and the number of situations required.

An analysis of variance (ANOVA) can be used for both intra- and interobserver reliability assessments for two or several observers to reveal systematic associations and also covariations between two or several factors. All available data (main- and subcategories, first and second observing session, observers, studied situations, etc) are included and considered in the model.

The kappa coefficient (55) can be calculated for both intra- and interobserver reliability assessments for two or several observers. Kappa is an improvement of the simpler calculation of percentage of agreement, correcting for observations by chance. However, kappa is not applicable to continuous data, not even when grouped into ordinal categories (56).

Hence, for evaluating observation methods, calculations of the kappa coefficient are restricted to time-sampled data; real-time observations should not be assessed with this method.

The proportion of agreement can be used for both intraobserver (two sessions) and interobserver (two observers) reliability assessments. The overall proportion of agreement (57) is the simplest and most frequently used index of agreement. If the category under study is relatively rare, the proportion of specific agreement is preferable (58, 59).

A multivariate data analysis of observer agreement can be used for interobserver reliability assessments for several observers to reflect the observers' congruity for each category.

Reliability can be expected to depend on the accuracy with which a posture or activity is assessed, the number of simultaneous exposures to be recorded, the training of the observer, and his or her familiarity with the work to be performed. In general, reliability testing confirms these expectations. However, few of the methods included in table 1 have been thoroughly tested for reliability. Interobserver reliability has been assessed by van der Beek using calculations of kappa (48). High reliability was obtained for observations of posture (four levels of arm and back postures on an interval scale), provided that manual materials handling was not simultaneously observed. The authors therefore concluded that, if information on manual materials handling is needed, it should be obtained by a second observer. Interobserver reliability has also been assessed by Fransson-Hall et al using an analysis of variance of observer correlations (unpublished manuscript). The reliability was higher for registrations of duration than of frequency, probably reflecting a discrepancy between the observers according to the accuracy with which the work was registered (ie, a specific posture or activity recorded as one continuous event or separated into several shorter events). In general, the interobserver reliability was high for trunk and arm posture, but low for registrations of neck posture and manual handling activities. Stetson et al (45), using multiple regression analysis, found no significant differences between four trained analysts who observed video recordings of hand and wrist postures and grip types. They also noted that the identification of extreme wrist postures yielded a high interobserver agreement, but that different interpretations sometimes occurred close to the decision criteria.

Most other authors have measured the proportion of agreement between two observers, but this method may give a falsely high estimate of reliability for exposure factors which occur infrequently. As expected, the percentage of agreement is high for simple, dichotomous assessments of frequently occurring exposures (sit-stand, back bent-back straight), when both duration and frequency of exposure are observed (40, 46). However, when the complexity of