

Accuracy and validity of observational estimates of wrist and forearm posture

BRIAN D. LOWE*

National Institute for Occupational Safety and Health, 4676 Columbia Parkway,
MS C-24, Cincinnati, Ohio, OH 45226, USA

Keywords: Wrist posture; Exposure assessment; Musculoskeletal disorders

Numerous observational methods for analysis of working posture of the wrist/forearm have been reported in the literature yet few of these methods have been validated for the accuracy of their posture classification. The present study evaluated the accuracy of estimates of working posture made by 28 experienced ergonomists using methods of scaling upper limb posture typical of those reported in the literature. Observational estimates of wrist/forearm posture of four jobs presented on video-recording were compared with posture levels measured directly with an electrogoniometer system. Ergonomists using a visual analogue scale tended to underestimate peak and average wrist extension with mean errors of -29.4% and -10.5% of the joint ROM, respectively ($p < 0.05$). While estimates of wrist flexion, pronation and supination resulted in less bias, variability in observer error was large for all wrist postures. The probability of an analyst misclassifying the most frequently occurring posture using a three- and a six-category scale was 54 and 70%, respectively. The probability of misclassifying peak posture was 22 and 61% using a three- and a six-category scale respectively. This suggests a trade-off between the degree of precision afforded by the categorical scale and the likelihood of posture misclassification. Estimates of the temporal distribution of posture among the categories appeared to be biased towards more neutral postures than were measured for the jobs. This indicated the possibility of a trend towards underestimation of posture duration severity by the ergonomists.

1. Introduction

The documentation of physical exposure in the workplace by way of job analysis is critical to surveillance and risk identification, evaluation of ergonomic interventions, and quantifying exposure dose for epidemiologic studies. Job analysis methods are used by ergonomists in the research community as well as by practitioners. The precision of measures of physical exposure has ranged from simple job titles to measurements with direct recording instrumentation (van der Beek and Frings-Dresen 1998). Lying between these extremes is perhaps the most prevalent method of documenting exposure to physical work factors—observational job analysis (Winkel and Westgaard 1992, Kilbom 1994). Observational methods of job analysis offer obvious advantages in precision over the use of simple job titles, and their deficiency in precision over more sophisticated instrumentation-based measures is compensated for by their efficiency, lower cost, and ease of use (Kilbom 1994, Winkel and

*Author for correspondence. e-mail: blowe@cdc.gov

Mathiassen 1994). This may explain why so many methods for observational job analysis have appeared in the literature (e.g. Armstrong *et al.* 1982, Holzmann 1982, Kilbom *et al.* 1986, Drury 1987, Stetson *et al.* 1991, McAtamney and Corlett 1993, de Looze *et al.* 1994, Kemmlert 1995, Moore and Garg 1995, Leskinen *et al.* 1997, Occhipinti 1998, Seth *et al.* 1999).

While numerous observational methods for documenting levels of posture, forceful exertions, and repetitiveness have appeared in the literature, comparisons between studies involving physical exposure continue to be limited by a lack of standardization among methods of assessing exposure. As an example, Stock (1991) eliminated 18 of a potential 49 studies from an epidemiological meta-analysis of workplace factors and musculoskeletal disorders on the basis of inadequate exposure measures. Stock noted what were referred to as '... serious flaws in the measurement of exposure' with some studies, yet there is no consensus in the ergonomics community regarding how the physical work factors of posture, repetitiveness, and forcefulness of exertion should be operationalized or scaled when observation is used to acquire physical exposure data.

Working posture is a physical factor commonly documented in job analyses for the purpose of surveillance and job redesign. The operationalization of working posture is generally straightforward because posture can be expressed unambiguously in terms of joint angles. However, a clear lack of standardization exists in how working posture of the upper limbs is scaled for the purpose of categorizing exposure levels. Since postural analyses of industrial work are often conducted by visual observation from a video-recording of the work activities, and the limitations of visual observation for posture analysis have not been well defined, the lack of standardization is understandable.

Kilbom (1994) emphasized two types of validity integral to methods for exposure documentation. *External validity*, sometimes referred to as *predictive value*, is defined as the 'ability of an observational method to identify physical exposures associated with an increased risk of musculoskeletal disorders' (p. 38) and requires a tenable underlying hypothesis regarding the cause-effect relationship between physical exposure and pathophysiology of work-related musculoskeletal disorders (WMSDs) and a method design to permit the assessment of exposure with sufficient accuracy. Inter-observer agreement is a commonly reported measure of reliability. However, inter-observer reliability is more meaningful if the validity of the observational method has been related to a well established standard measure (Baty *et al.* 1986), namely one that is instrumentation-based (Sorock and Courtney 1996). A method has high *internal validity* if it accurately classifies the exposure of interest (Kilbom 1994). Relatively few studies have examined the internal validity of observational job analysis methods. Those studies that have examined internal validity suggest that inter-observer agreement is higher than accuracy. de Looze *et al.* (1994) evaluated the performance of two observers using the TRAC (Task Recording and Analysis on Computer) system. The two observers documented torso flexion, 'arms position' (which appears to be shoulder flexion), and 'legs position' (deviation from a straight leg when standing), for which angular measures were obtained with an optoelectric system. The authors reported a high proportion agreement index (P_0) and Cohen's κ suggesting that inter-observer agreement was high, but reported that *accuracy* (observation-measurement agreement) was low. Recently, Ketola *et al.* (2001) examined the validity of observational estimates of 'non-neutral' wrist posture defined as a deviation exceeding 20° from neutral. Their investigation operationa-

lized non-neutral wrist posture as a 20° deviation occurring for more than 33% of the work cycle. Their results suggest that even a trained expert exhibited difficulty in estimating non-neutral wrist posture when operationalized in this dichotomous manner, exhibiting a best-case Kappa of 0.32.

Spielholz *et al.* (2001) used electrogoniometry to measure wrist motions in three jobs and compared this to observational estimates of the percentage of the work cycle in which the wrist was observed in a non-neutral posture (defined as flexion/extension exceeding 30°/30°, radial/ulnar deviation exceeding 10°/15°, and pronation/supination exceeding 45°/45°). They found Pearson product-moment correlations of 0.33, 0.21, and 0.07 between observational analysis from a video recording and direct electrogoniometer measurement of these posture durations. While the correlation values were low, measures of agreement, derived by accounting for both location (difference between means) and scale (differences in variability) shift were higher: 0.49, 0.66, 0.82, respectively. Juul-Kristensen *et al.* (2001) compared estimates of the temporal distribution of wrist posture against electrogoniometric measures of these distributions from 19 workers. Their observer overestimated the percentage of the work cycle in moderate wrist extension by 18% which happened to offset the 17% underestimation of the percentage of the work cycle in neutral wrist posture. These findings illustrate that observational exposure assessment methods have the potential for misclassification of exposure.

The present study combined aspects of the studies described above (de Looze *et al.* 1994, Juul-Kristensen *et al.* 2001, Ketola *et al.* 2001, Spielholz *et al.* 2001) and included a larger number of observers from whom estimates of posture variables were obtained. It is difficult to gain insight into biases, trends, or systematic error generalizable to ergonomic analysts as a group from the performance of one or two analysts. This paper presents findings of a study conducted in our laboratory to investigate the accuracy of observational estimates of physical risk factors associated with musculoskeletal disorders of the upper limbs. The objective was to evaluate the accuracy of observational estimates of upper limb posture using methods for scaling posture which are representative of those appearing in the literature. These methods included categorical scales and a continuous scale which were used to estimate the extreme, average, and/or most frequently occurring working posture as estimated in a video-based analysis of representative work cycles.

2. Method

2.1. Apparatus

A tri-axial electrogoniometer system (Biometrics Ltd., Ladysmith, VA, USA) was mounted to the dominant hand (right) of each worker. The system included two goniometer units: a biaxial unit (model XM65) and a single axis torsionmeter (model Z110). The biaxial XM65 registered movement in the flexion/extension axis and radial/ulnar deviation axis. The XM65 was mounted so that the end blocks spanned the wrist joint aligned over the third metacarpal. The Z110 torsionmeter recorded torsional movement of the forearm in pronation/supination as the radius rotated about the ulna.

A hand-held 'mini-DV' camera was used to acquire video of the work tasks simultaneously with the electrogoniometric recording. The video recordings were made by an ergonomist with over 25 years experience acquiring video recordings of industrial jobs for the purpose of ergonomic evaluation. The video recordings with

the mini-DV camera were time-synchronized with the electrogoniometric recording by a manually-triggered LED pulse which appeared as a voltage pulse in an auxiliary analogue channel recorded with the electrogoniometer and as an instantaneous illumination of an LED in the video recording. Video frames were indexed to the frame showing the first appearance of the LED illumination and mapped to the corresponding electrogoniometer samples which were low pass filtered and time averaged to a 30 Hz sampling rate to match the 30 frames/s video standard. The digital video from the mini-DV recording was then dubbed to VHS format to create first-generation VHS format recordings that contained multiple work cycles of the jobs to be observed by ergonomists. The video was edited so that the work cycles presented to the analysts corresponded to the exact portions of the electrogoniometer recording that were used in calculating the posture measures of interest. Video footage of each job contained between 5–12 complete work cycles (see table 1). Half of the work cycles presented were from frontal views and half were from side (sagittal) views.

2.2. Simulation of jobs

Three research associates acted as 'workers' performing four jobs simulated in a laboratory setting. The heights of these workers were 177.0, 184.4, and 170 cm; their weights were 86.4, 80.3, and 69.5 kg respectively. They were all right-handed males and wore identical black sleeveless t-shirts while performing the jobs.

The jobs were repetitive mono-task jobs designed to require a variety of working postures of the upper limbs and to span a range in cycle time. They are described briefly in table 1. Only the posture of the right (dominant side) arm was quantified and ergonomists were instructed to observe and estimate only the posture of the right arm. The research associates were provided with some practice performing the jobs; however, the variability in cycle time and upper limb kinematics within-worker may have been higher than that typical at a production facility where routinized mono-task work was prevalent.

2.3. Analysts

Twenty-eight ergonomists were recruited equivalently from academia and industry/consulting to participate in this study. These professionals were recruited world-wide from personal contacts of the author (approximately one-third) and referrals from personal contacts of the author (approximately two-thirds). The criterion for inclusion was experience in performing analyses of jobs for the purpose of ergonomic evaluation. The 14 ergonomists recruited from academia were either full-time faculty with appointments in ergonomics-related programmes, or graduate students working

Table 1. Descriptions of the jobs

Job	Average cycle time (s)	No. of cycles presented	Description
A	13	8	Folding t-shirts
B	8	12	Assembly of small cardboard shipping boxes
C	56	5	Manual insertion of four screws with a standard screwdriver
D	46	5	Power-assisted insertion of five screws with a cordless drill

under supervision of such faculty, with experience in the area of upper limb musculoskeletal disorders and job analysis. The 14 ergonomists recruited from industry/consulting were practitioners who had experience with job analysis for ergonomic evaluation and who possessed educational background and/or professional certification in ergonomics. Of the 28 total participants, nine held board certification in ergonomics (BCPE) and two were registered Occupational Therapists. In terms of education, 12 held doctoral degrees, 13 held a M.S. degree as their highest degree, and three held a B.S. as their highest degree. The number of years experience in the field of ergonomics for these analysts averaged 9.1 years (6.5 years SD) and ranged between 1–30 years. Analysts were compensated at a rate comparable to an hourly consulting fee for a professional ergonomist. Informed consent was obtained prior to the collection of data and all procedures had been approved by the Institutional Review Board.

2.4. *Observational methods*

Most methods of scaling wrist/forearm posture for the purpose of observational job analysis have been proposed with categorical scales for posture recording. (table 2 lists a variety of job analysis methods that have appeared in the literature and the number of categories used to scale wrist/forearm posture.) Two categorical methods of independently scaling wrist/forearm postures were chosen for investigation in this study—scales with three and six posture categories. The lower resolution scale with three categories partitions the joint range of motion (ROM) with two boundaries (figure 1a). The higher resolution scale with six categories requires five boundaries (figure 1b). The rationale for choosing three- and six-category scales was that the three-category scale was believed to provide the minimum useful information for documenting wrist/forearm posture. It identifies a neutral category and two non-neutral categories of flexion or extension, radial or ulnar deviation, and pronation or supination. At the other extreme a categorical scale with more than six-categories begins to approach a continuous scaling method, which was the third scaling method investigated, and the majority of methods appearing in the literature have incorporated fewer than six posture categories. The choice of three- and six-category scales was made to reflect a lower and higher precision categorical scale. The category boundaries for the categorical scales were chosen to facilitate a relatively equal width among the category intervals while attempting to use boundaries somewhat consistent with those of other studies (McAtamney and Corlett 1993, Juul-Kristensen *et al.* 2001, Ketola *et al.* 2001, Spielholz *et al.* 2001).

The third scale evaluated was a continuous posture scale (Latko 1997). This visual analogue scale (VAS) was numbered from 0 to 10 with pole anchors representing 'neutral' and 'extreme' postures (figure 1c). The visual analogue scale partitioned each joint's ROM continuously, with the zero pole representing the neutral (0°) posture for the joint axis, and the value 10 representing the limit to the ROM for the joint axis. Joint ROMs for the wrist and forearm were based on the mean plus one standard deviation ROMs published by Bonebrake *et al.* (1990), who reported ROMs measured with the forearm in multiple planes, making their data more generalizable to functional work activities in which joint displacements occur in multiple axes concurrently. These ROMs (listed in figure 2) were presented in the instruction materials and on the data forms supplied to those analysts using the continuous VAS method.

(a) three-category scale

category ^a	flexion/extension	supination/pronation	radial/ulnar deviation
1	> 20° flex	> 40° sup	> 10° radial
2	20° flex – 20° ext	40° sup – 40° pro	10° radial – 10° ulnar
3	> 20° ext	> 40° pro	> 10° ulnar

category ^b	flexion	extension	supination	pronation	radial	ulnar
1	0° - 20°	0° - 20°	0° - 40°	0° - 40°	0° - 10°	0° - 10°
2	> 20°	> 20°	> 40°	> 40°	> 10°	> 10°

^a scale for estimating mode^b scale for estimating peak

(b) six-category scale

category ^a	flexion/extension	supination/pronation	radial/ulnar deviation
1	> 45° flex	> 60° sup	> 20° radial
2	20° – 45° flex	30° – 60° sup	10° – 20° radial
3	0° – 20° flex	0° – 30° sup	0° – 10° radial
4	0° – 20° ext	0° – 30° pro	0° – 10° ulnar
5	20° – 45° ext	30° – 60° pro	10° – 20° ulnar
6	> 45° ext	> 60° pro	> 20° ulnar

category ^b	flexion	extension	supination	pronation	radial	ulnar
1	0° - 20°	0° - 20°	0° - 30°	0° - 30°	0° - 10°	0° - 10°
2	20° - 45°	20° - 45°	30° - 60°	30° - 60°	10° - 20°	10° - 20°
3	> 45°	> 45°	> 60°	> 60°	> 20°	> 20°

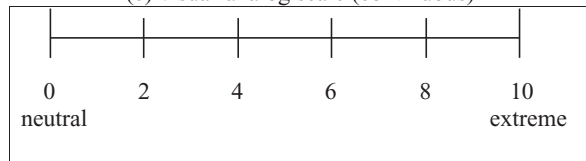
^a scale for estimating mode^b scale for estimating peak(c) visual-analog scale (continuous)^c^c “neutral” posture always corresponds to 0°, “extreme” posture corresponds to the limit of the joint ROM (shown in Figure 2).

Figure 1. Scales for observational estimates of working posture. (a) Three-category scale, (b) six-category scale, and (c) continuous visual analogue scale.

Analysts used the categorical scales to estimate the peak and most frequently occurring (mode) posture categories based on analysis of an average, or typical, work cycle. Analysts also used the categorical scales to provide an estimate of the temporal distribution of posture for the typical work cycle. They used the continuous scale to provide estimates of the peak and average posture for the typical work cycle.

Analysts were nested within the three posture scaling methods (three-category, six-category, continuous VAS), nested within three workers performing the four jobs, and crossed within the four jobs. Thus, each analyst used a single posture observation method to evaluate the wrist/forearm postures of a single worker to evaluate all four jobs. Analysts were assigned randomly to the posture method condition and were randomly assigned to the worker they observed. Of the 28

Table 2. Observational job analysis methods in the literature and their scaling of wrist/forearm posture

Posture categories	Flexion/extension	Radial/ulnar deviation	Pronation/supination
3	McAtamney and Corlett (1993) Occhipinti (1998)	Armstrong <i>et al.</i> (1982) Genaidy <i>et al.</i> (1993) McAtamney and Corlett (1993) Occhipinti (1998)	Armstrong <i>et al.</i> (1982) Genaidy <i>et al.</i> (1993) Occhipinti (1998) Seth <i>et al.</i> (1999)
5	Armstrong <i>et al.</i> (1982) Kilbom <i>et al.</i> (1986) Genaidy <i>et al.</i> (1993) Moore and Garg (1995) Seth <i>et al.</i> (1999)	Kilbom <i>et al.</i> (1986) Moore and Garg (1995) Seth <i>et al.</i> (1999)	Moore and Garg (1995)
8	Drury (1987)*	Drury (1987)*	Drury (1987)*

*Four zones used for each bi-directional posture.

Flexion/Extension	Radial/Ulnar Deviation	Pronation/Supination
ROM: 85° extension -95° flexion	ROM: 70° ulnar deviation -45° radial deviation	ROM: 130° pronation -145° supination

NOTE: Shaded regions do not reflect the ROM values listed. See text for description of the ROMs.

Figure 2. Conventions for the measurement of wrist posture. (Illustrations adapted from the American Academy of Orthopaedic Surgeons (1965)).

analysts, nine were assigned to use the three-category posture scales, 10 were assigned to use the six-category posture scales, and nine were assigned to use the visual analogue (continuous) scale.

All study materials, which included a VHS format video recording, data sheets, and detailed instruction materials, were mailed to ergonomists who performed the analyses at their home institution, on their own time. They were requested to read all of the instruction materials prior to observing video footage of the jobs. The instruction material described the layout of the data sheets, the presentation of job cycles on the video, and the posture scales to be used. Immediately prior to observing the job and immediately after completion of the analysis, analysts were asked to document the respective starting time and ending time so that the time to

completion of the posture analysis of each job was documented. Analysts reported the dimensions of the video monitor they used in observing the VHS video recording. The average diagonal dimension was 62 cm (standard deviation = 17 cm). The minimum reported monitor diagonal was 33 cm, the maximum was 94 cm.

2.5. *Direct measurement*

The electrogoniometers were calibrated with a fixture constructed to allow the worker's wrist to be fixed in three positions in the flexion/extension, radial/ulnar deviation, and pronation/supination axes in a manner similar to that described by Spielholz (1998). Readings were taken from the wrist while maintaining static posture in these angular positions combined factorially, to yield 27 total static recordings. Static posture recordings were taken as the average voltage output in each static position. Each electrogoniometer signal was sampled at 248 Hz and digitally low-pass filtered (6th order Butterworth) with a 10 Hz cut-off frequency. Calibrations were made by a multiple regression approach in which readings from all three axes were entered as regression variables and polynomial terms and all interactions were included in the regression model. The average coefficients of determination (R^2) predicting fixture-measured angular position were high for flexion/extension (0.99) and pronation/supination (0.94), but were lower for radial/ulnar deviation (0.80). These R^2 values agree fairly closely with those reported by Spielholz (1998).

Electrogoniometer calibration equations for wrist posture in radial/ulnar deviation resulted in unacceptably low predictions of radial/ulnar deviation as measured from the fixture. Discrepancies between the fixture-measured angle and goniometer-measured angle in ulnar deviation were as large as 10° at 30° deviation for one worker, contrasted with the much more acceptable 'worst case' discrepancies of 0.5° at 45° flexion and 2.6° at 45° pronation. For this reason the accuracy of observational estimates of radial/ulnar deviation were not based on the electrogoniometer measurements and no measures of accuracy were calculated for estimates of radial/ulnar deviation. Interrater agreement among analysts' estimates of radial/ulnar deviation posture are presented.

Job analysts are typically presented with a video recording showing multiple work cycles of a job in which posture levels are to be estimated and summarized for the job. The analyst is expected to observe multiple work cycles and to document estimates of posture variables that reflect those of the 'typical' or 'average' work cycle. The present study was conducted in the same manner. Analysts estimated the posture variable for each job based on an estimate of the average for the work cycles presented (the 'typical' work cycle). The accuracy of each analyst's estimate of each posture variable was calculated as the difference between the analyst's estimate of the variable and the measured variable as averaged over all of the individual work cycles presented. Implicit in the analysts' task is to ignore any variability that exists in the posture level between the observed work cycles when estimating the posture levels that reflect the typical or average work cycle. Since electrogoniometer measurements were summarized for the wrist/forearm posture variables by individual work cycle the study afforded the opportunity to examine the effect of postural variability between work cycles on the accuracy of analysts' posture estimates. Work cycle variability was hypothesized to be negatively correlated with the accuracy of analysts' posture estimates.

3. Results

3.1. Flexion/extension and pronation/supination—observational accuracy and bias

Analysts' estimates of wrist flexion/extension and forearm pronation/supination were compared directly to electrogoniometric measurements of the associated posture variables. For posture estimates obtained with the categorical scales, error was calculated as the number of categories deviation between the estimated and measured posture variable which ranged between $-(n-1)$ and $+(n-1)$, where n is the number of scale categories. Using the three-category scale the most frequently occurring (mode) posture was misclassified in 54.2% of the cases and the peak posture was misclassified in 22.2% of the cases. With the six-category scale these percentages were 70 and 61.2%, respectively. These probabilities represent 100% minus the probability of correct classification of the posture, which is the probability associated with a misclassification of zero (0) categories. These latter values are shown in boldface in table 3a, which lists the probability of misclassification as a function of the magnitude of misclassification in number of posture categories. Misclassifications of the most frequently occurring (mode) posture using the three-category scale were almost always between adjacent categories and errors in peak posture classification were constrained to adjacent categories because the three-category scale uses only two-categories for classifying peak posture. Posture misclassifications using the six-category scale occurred between non-adjacent categories with relatively low frequency, however, misclassification errors between -4 and $+2$ categories are evident in table 3.

The misclassification probabilities listed in table 3a are more informative when compared to those calculated from performance expected by chance. This calculation is illustrated graphically in figure 3. An $n \times n$ contingency table was generated for each scaling condition where n is the number of scale categories. Chance performance was calculated by retaining the marginal probabilities for the measured values and setting the marginal probabilities for the estimated values equal to $1/n$ (equal likelihood of classification among the n categories). Once the marginal probabilities for the measured and estimated values are determined completion of the chance contingency table is straightforward. The probabilities of categorical error between $-(n-1)$ and $+(n-1)$ are easily calculated from the chance contingency table. Figure 3 illustrates the calculation of the ergonomists' performance (upper contingency table) and chance performance (lower contingency table) for classification of most frequently occurring pronation/supination posture. The graph at the bottom of the figure plots the cumulative probability of absolute misclassification error (expressed in category units) for the ergonomists and for that of chance performance.

Weighted Kappa statistics (κ_w) expressing the agreement between the analysts and the electrogoniometric measurements of the posture variables are listed in table 3b. Spearman correlation coefficients (r_s) are also reported between the measured and estimated values. For several of the peak posture variables the low values of κ_w are somewhat misleading because the measured values for the peak postures were often, if not always, in the maximum category. Kappa is influenced by the prevalence of the characteristic being measured and is reduced by infrequently observed events which increase chance agreement (Burt and Punnett 1999). This is clearly evident when comparing the Kappas for peak extension and pronation (both 0) with those for wrist flexion (0.386) and forearm supination (0.453) in the three-category scale condition. Peak extension and pronation were always measured in the higher of the

Table 3a. Posture misclassification probabilities for categorical posture scales. Digits in each cell represent the probability of misclassifying the posture with an error of . . . , - 2, - 1, 0, + 1, + 2, . . . categories. Misclassification of error 0 (probability shown in bold) is the probability of correct classification of the posture. The number of error categories possible is shown below the designation of peak/mode and is between ± (n-1) for a categorical scale with n categories

<i>n</i> = 36 (9 analysts × 4 jobs)	Peak (two categories) - 1, 0 , + 1	Mode* (three categories) - 2, - 1, 0 , + 1, + 2
Wrist extension	25.0, 75.0 , 0	0, 38.9, 55.6 , 5.6, 0
Wrist flexion	22.2, 75.0 , 2.8	
Forearm pronation	13.9, 86.1 , 0	2.8, 52.8, 36.1 , 8.3, 0
Forearm supination	8.3, 75.0 , 16.7	
Average	17.4, 77.8 , 4.9	1.4, 45.8, 45.8 , 6.9, 0
<i>n</i> = 40 (10 analysts × 4 jobs)	Peak (three categories) - 2, - 1, 0 , + 1, + 2	Mode† (six categories) - 5, - 4, - 3, - 2, - 1, 0 , + 1, + 2, + 3, + 4, + 5
Wrist extension	27.5, 50.0, 22.5 , 0, 0	0, 2.5, 10.0, 15.0, 45.0, 22.5 , 5.0, 0, 0, 0, 0
Wrist flexion	20.0, 32.5, 37.5 , 10.0, 0	
Forearm pronation	17.5, 15.0, 52.5 , 15.0, 0	0, 7.5, 7.5, 17.5, 7.5, 37.5 , 17.5, 5.0, 0, 0, 0
Forearm supination	17.5, 10.0, 42.5 , 22.5, 7.5	
Average	20.6, 26.9, 38.8 , 11.9, 1.9	0, 5.0, 8.8, 16.3, 26.3, 30.0 , 11.3, 2.5, 0, 0, 0

*Most frequently occurring posture (mode) scales include flexion and extension; pronation and supination.

Table 3b. Weighted Kappa statistics and Spearman correlation coefficients for agreement between analyst estimates and electrogoniometric measures of posture variables

		Weighted Kappa		Spearman	
		Peak κ_w	Mode* κ_w	Peak r_s	Mode* r_s
Three category scale	Wrist extension	0	0.1628	na	0.209
	Wrist flexion	0.3864		0.440**	
	Forearm pronation	0	- 0.0286	na	- 0.007
Six category scale	Forearm supination	0.4527		0.460**	
	Wrist extension	0	- 0.0549	na	- 0.268
	Wrist flexion	0.1646		0.256	
	Forearm pronation	- 0.0656	0.0266	- 0.109	0.038
	Forearm supination	0.0265		0.070	

***p* < 0.01.

*Most frequently occurring posture (mode) scales include flexion and extension; pronation and supination.

na – no variability in measured category, correlation cannot be calculated.

two possible categories, while peak flexion was distributed as 17 and 83% between the low and high categories and peak supination was distributed as 69 and 31% between the low and high categories. A similar trend is evident in the six-category

ergonomists – most frequent pronation/supination

contingency table

		estimated						
		> 60° sup	30° - 60° sup	0° - 30° sup	0° - 30° pro	30° - 60° pro	> 60° pro	marginal
measured	> 60° sup	0.000	0.000	0.000	0.000	0.000	0.000	0.000
	30° - 60° sup	0.000	0.000	0.000	0.000	0.000	0.000	0.000
	0° - 30° sup	0.000	0.000	0.025	0.075	0.000	0.000	0.100
	0° - 30° pro	0.000	0.000	0.025	0.150	0.075	0.050	0.300
	30° - 60° pro	0.000	0.000	0.000	0.050	0.000	0.025	0.075
	> 60° pro	0.000	0.075	0.075	0.175	0.000	0.200	0.525
	marginal	0.000	0.075	0.125	0.450	0.075	0.275	1.000

probabilities of category error (# of categories)

	-5	-4	-3	-2	-1	0	+1	+2	+3	+4	+5
ergonomists	0.000	0.075	0.075	0.175	0.075	0.375	0.175	0.050	0.000	0.000	0.000

chance performance – most frequent pronation/supination

contingency table

		estimated						
		> 60° sup	30° - 60° sup	0° - 30° sup	0° - 30° pro	30° - 60° pro	> 60° pro	marginal
measured	> 60° sup	0.000	0.000	0.000	0.000	0.000	0.000	0.000
	30° - 60° sup	0.000	0.000	0.000	0.000	0.000	0.000	0.000
	0° - 30° sup	0.017	0.017	0.017	0.017	0.017	0.017	0.100
	0° - 30° pro	0.050	0.050	0.050	0.050	0.050	0.050	0.300
	30° - 60° pro	0.013	0.013	0.013	0.013	0.013	0.013	0.075
	> 60° pro	0.088	0.088	0.088	0.088	0.088	0.088	0.525
	marginal	0.167	0.167	0.167	0.167	0.167	0.167	1.000

probabilities of category error (# of categories)

	-5	-4	-3	-2	-1	0	+1	+2	+3	+4	+5
chance	0.088	0.100	0.150	0.167	0.167	0.167	0.079	0.067	0.017	0.000	0.000

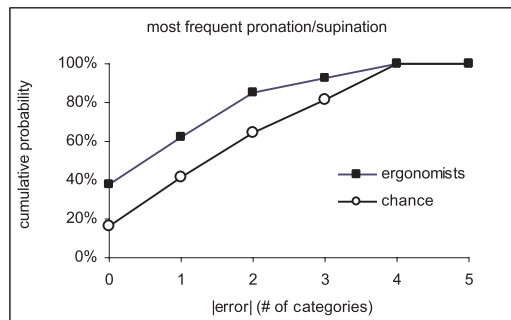
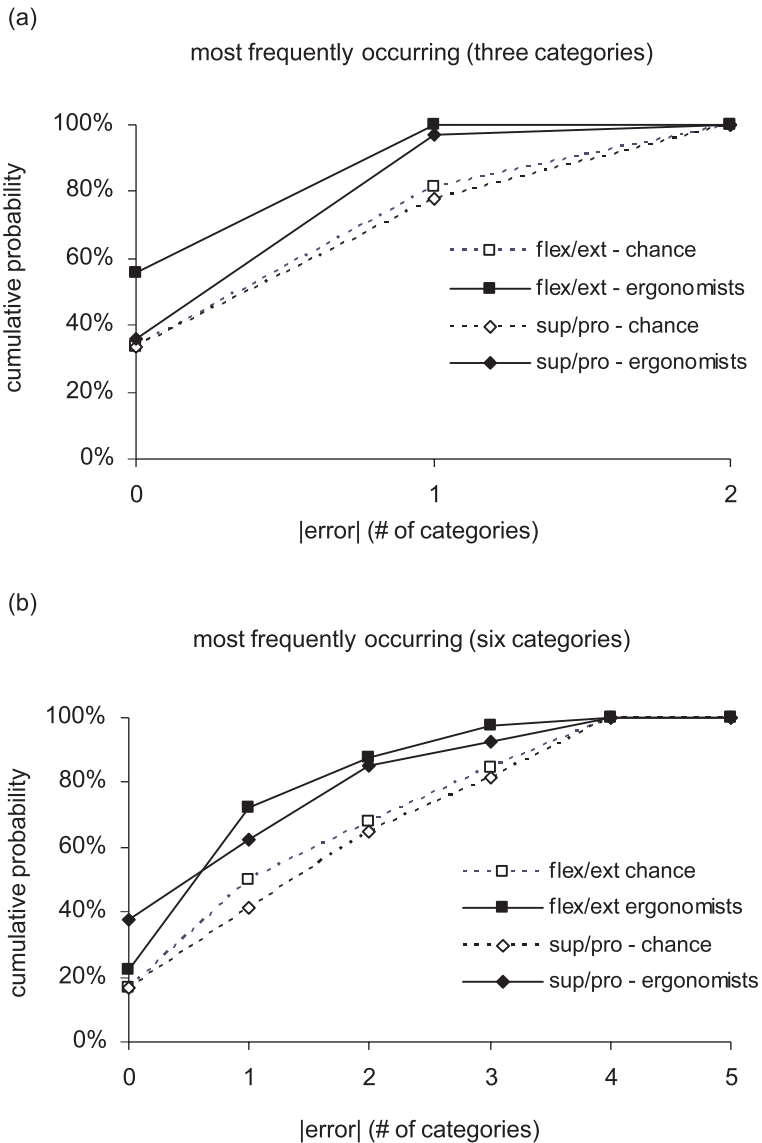


Figure 3. Example calculation of the probability of posture misclassification by ergonomists and by chance performance.

scale condition. Thus, the Kappa statistics, particularly in the case of the peak postures, may be more reflective of large differences in the prevalence of the postures being estimated than of the ability of analysts to estimate the different postures.

Figures 4 and 5 show the cumulative probability distributions for the remainder of the postures and scaling conditions. In the worst situations, such as that shown for peak wrist extension (figure 4c), the ergonomists' performance appears to be no better than that which would be expected by chance. Peak wrist extension, when



(Figure 4 continued overleaf)

rated as 0° – 20° , 20° – 45° , or $> 45^{\circ}$ resulted in a slightly lower probability of correct classification, 0.225, among ergonomists than that resulting by chance, 0.333. The probability of classifying wrist extension within ± 1 category of the measured category was slightly higher among the ergonomists', 0.725, than would be expected by chance, 0.667.

Figure 5 shows scatterplots of the estimates of the posture variables obtained with the continuous visual analogue scale vs. the measured posture variables. Means for the estimated and measured variables are shown graphically in figure 6, which correspond to the mean analyst errors (differences between the mean estimated and

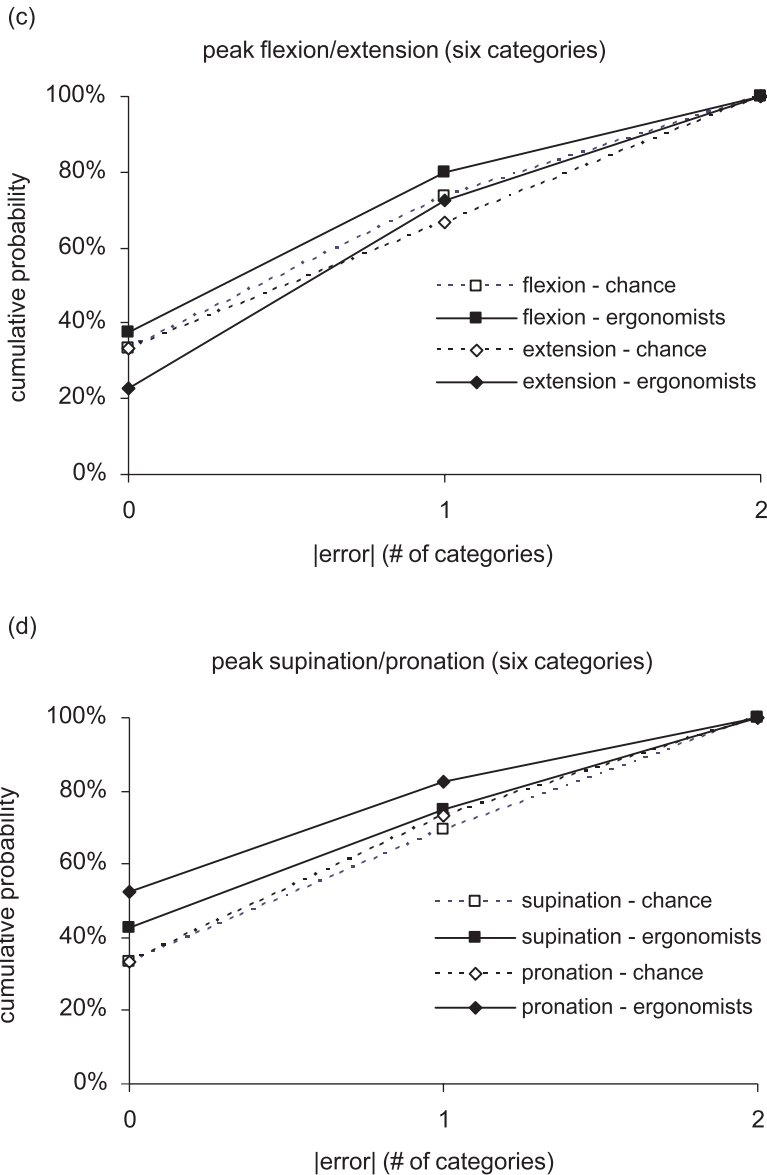
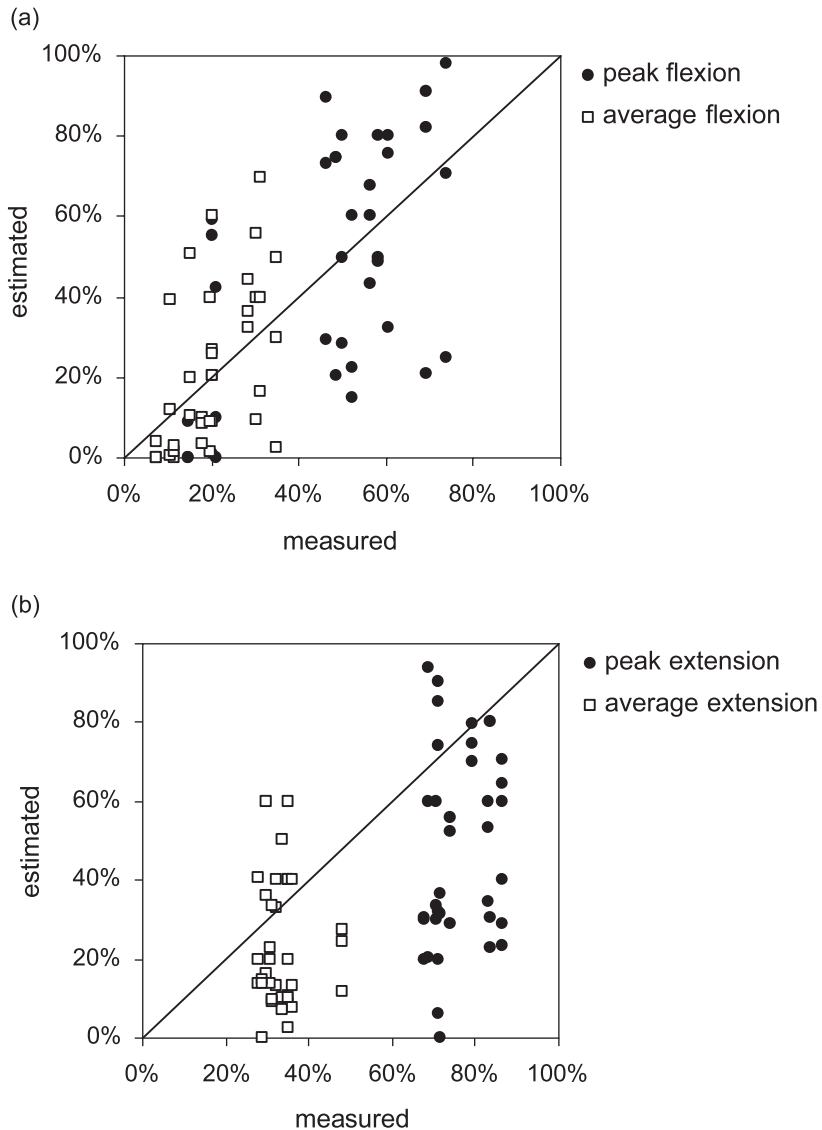


Figure 4. Cumulative probabilities of posture misclassification error for: (a) most frequently occurring flexion/extension and supination/pronation—three categories, (b) most frequently occurring flexion/extension and supination/pronation—six categories, (c) peak flexion/extension—six categories, (d) peak supination/pronation—six categories. Error is expressed as the absolute value of the number of categories deviation between the estimated and measured posture category. The probability of correct posture classification lies on the ordinate axis, where $|\text{error}| = 0$.

measured variables) listed in table 4. With the exception of wrist extension, differences between the means for estimated and measured postures were generally small. However, variability was high and linear regressions of estimated posture against the associated measured posture revealed poor coefficients of determination.



(Figure 5 continued overleaf)

Only the regression relationships involving peak ($R^2 = 0.311$, $\beta_1 = 0.89$) and average ($R^2 = 0.279$, $\beta_1 = 1.22$) wrist flexion were statistically significant ($p < 0.05$). In particularly troublesome cases ergonomists rated peak wrist extension as low as 0% of the range of motion when it was measured in excess of 70% of the extension ROM (see figure 5b). Conversely, peak forearm supination was rated as high as 100% of the ROM in a situation where no forearm supination was measured electrogoniometrically (see figure 5c).

Table 4 shows the mean error among analysts for the flexion/extension and pronation/supination posture variables. Table 4 expresses error in terms of a difference in percentage of the VAS distance for the particular joint posture, which

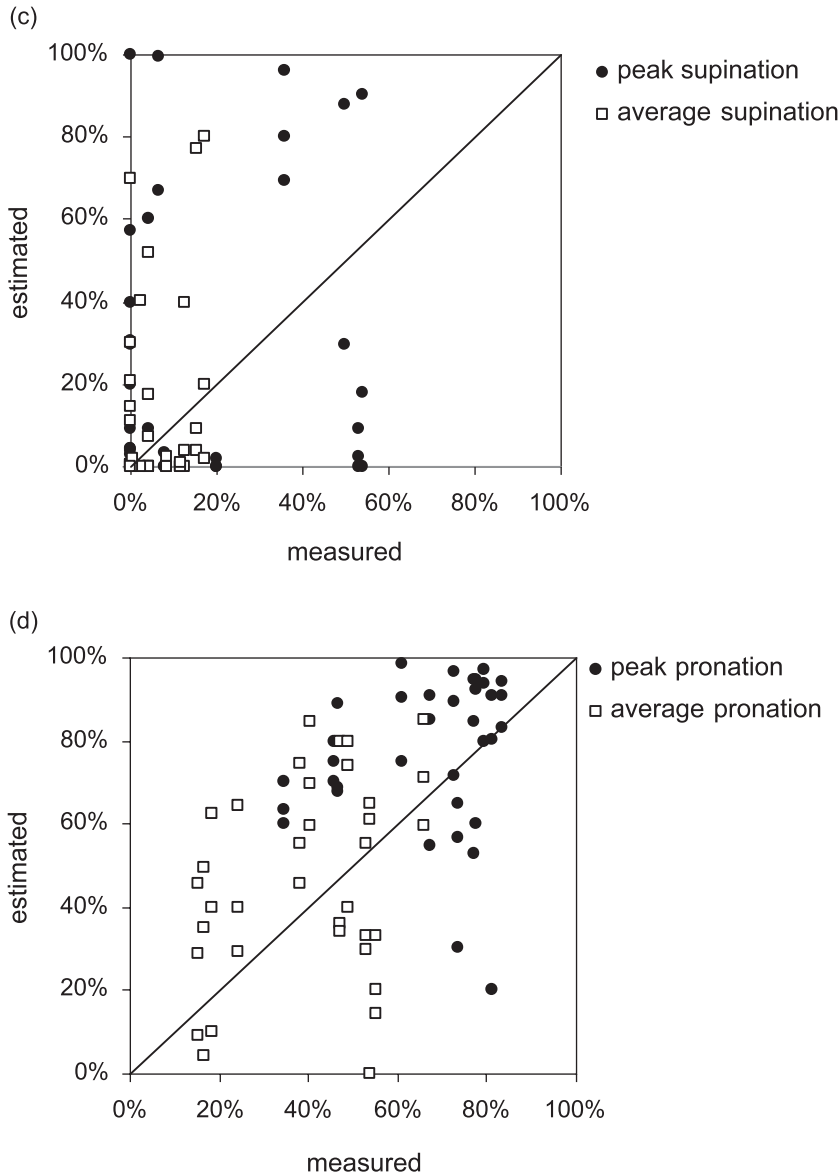


Figure 5. Scatterplots of visual analogue scale (VAS) values-estimated vs. measured postures - for (a) peak and average wrist flexion, (b) peak and average wrist extension, (c) peak and average forearm supination, and (d) peak and average forearm pronation.

can be mapped directly to a percentage of the joint ROM. Students *t*-tests were conducted on the mean observational error under the assumption that observational error was distributed normally with mean 0 and sample variance $\hat{\sigma}^2$. A two-tailed null hypothesis that $\bar{X} = 0$ was tested against the alternative hypothesis of $\bar{X} \neq 0$. When the null hypothesis was rejected a bias in observers' ratings, in the exhibited negative or positive direction, was supported. Posture estimates for which the $\bar{X} = 0$ hypothesis was rejected are designated by statistical significance in table 4.

Table 4. Mean analyst error for estimates of peak and average wrist posture using the continuous visual analog scale.

$n = 36$ (nine ergonomists \times four jobs)		Wrist extension	Wrist flexion	Forearm pronation	Forearm supination
Peak	Peak ($\mu \pm \sigma$)*	-0.294 ± 0.246	0.001 ± 0.250	0.099 ± 0.223	0.115 ± 0.393
	Peak ($t_{x/2}$)	-7.17^{**}	ns	2.67^{**}	ns
	Regression [†] - β_0, β_1	0.051, 0.548	0.053, 0.891	0.642, 0.187	0.262, 0.215
	R ²	0.023	0.311 ^{***}	0.025	0.017
Average	Average ($\mu \pm \sigma$)*	-0.105 ± 0.167	0.018 ± 0.170	0.069 ± 0.245	0.077 ± 0.227
	Average ($t_{x/2}$)	-3.77^{**}	ns	ns	ns
	Regression [†] - β_0, β_1	0.239, -0.023	$-0.027, 1.217$	0.295, 0.432	0.098, 0.676
	R ²	0.001	0.279 ^{***}	0.094	0.033

** $P < 0.01$, *** $P < 0.001$.

*Units are the fraction of the joint ROM.

[†]Regression equations of the form $estimated\ posture = \beta_0 + \beta_1(measured\ posture)$.

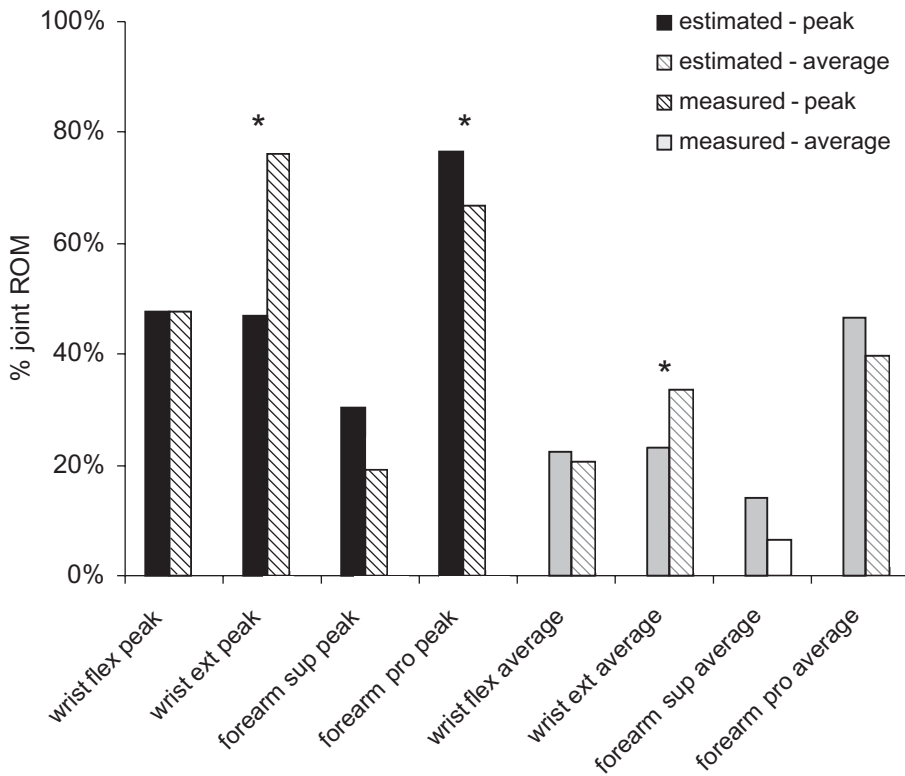


Figure 6. Mean VAS-estimated and measured values for posture variables. *Denotes means that are statistically different ($p < 0.05$).

Analysts' estimates of the temporal distribution of posture for the typical work cycle are shown in figures 7 (three-category scale) and 8 (six-category scale). The analysts' mean estimated distributions are shown with the measured distribution as averaged over the four jobs presented. These figures illustrate a trend towards a central tendency, in which the percentage of the work cycle in the more neutral

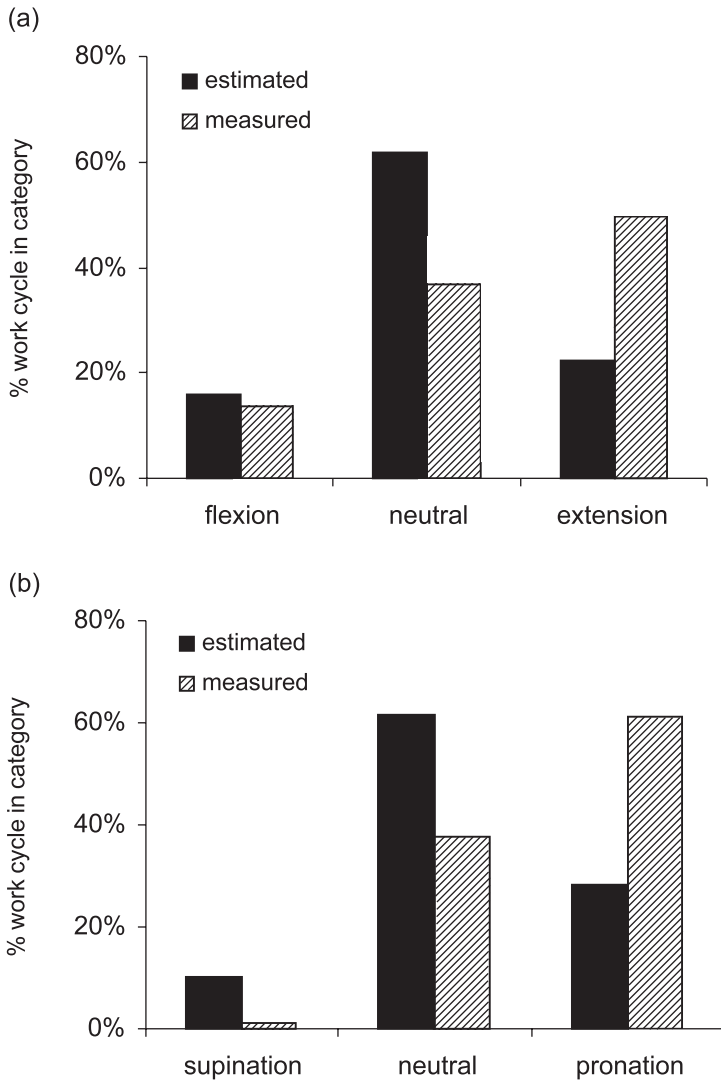


Figure 7. Estimated and measured distributions of posture as percentages of the work cycle using three posture categories for (a) flexion/extension, and (b) supination/pronation.

posture categories are overestimated and the percentage of the work cycle in the more extreme categories (namely for extension and pronation) are underestimated. Two measures of deviation between the estimated and measured postural distributions were calculated. These were the difference between the estimated and measured percentage of the work cycle in the neutral category (middle category with the three-category scale and middle two categories with the six-category scale) and the RMS error between the estimated and measured distributions. A relationship between the duration of the work cycle and the inaccuracy of the posture distribution estimate was hypothesized but neither measure of inaccuracy was related to the duration of the work cycle.

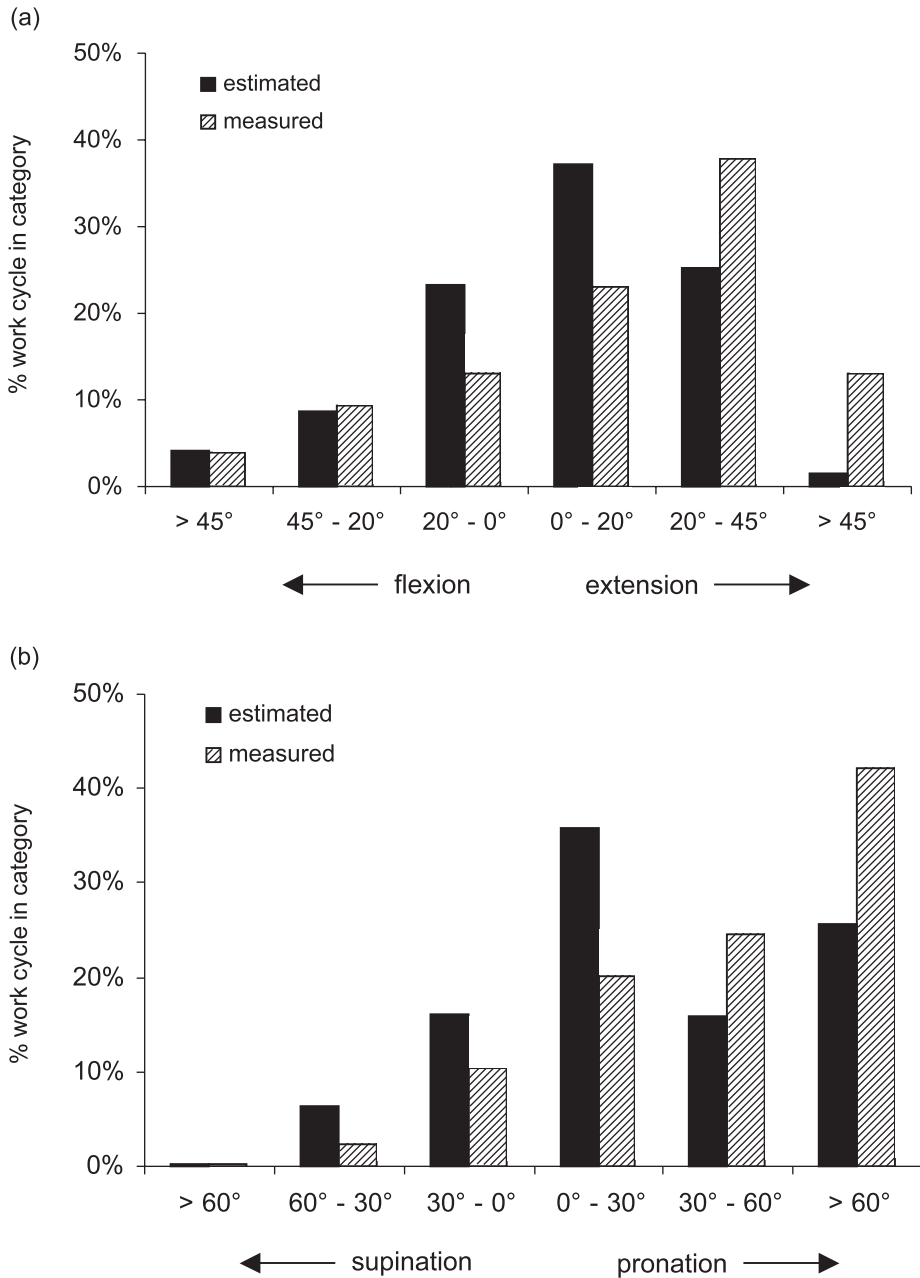


Figure 8. Estimated and measured distributions of posture as percentages of the work cycle using six posture categories for (a) flexion/extension, and (b) supination/pronation.

3.2. Radial/ulnar deviation—inter-observer agreement

Ergonomists' estimates of radial/ulnar deviation were examined only in terms of inter-observer agreement since the electrogoniometer calibrations for radial/ulnar deviation did not meet an acceptable level of reliability. Inter-observer agreement

was expressed using both the intraclass correlation coefficient (ICC) and Kendall's coefficient of concordance, W , which is applicable for ordinal scale data of two or more sets of rankings (Siegel and Castellan 1988). Ergonomists' estimates of radial/ulnar deviation posture were converted to rank orderings of the four jobs according to radial/ulnar deviation severity and the Kendall coefficient of concordance for the rankings of the jobs was calculated among the analysts. Table 5 lists the Kendall coefficients and ICCs for all wrist posture variables. Measures of inter-observer agreement are also shown for flexion/extension and pronation/supination posture for the purpose of comparison. Measures of inter-observer agreement for radial/ulnar deviation posture were slightly lower than those for flexion/extension and supination/pronation.

3.3. Effects of work cycle variability and analyst experience

The hypothesis that the accuracy of ergonomists' estimates of wrist/forearm posture is affected by postural variability in the work cycles observed was tested by generating ANCOVA models predicting the absolute inaccuracy in the posture estimate (absolute value of inaccuracy expressed as the difference between the estimated and measured posture levels). Work cycle variability was defined by the standard deviation of the measurements of the posture variable over the work cycles presented on the video recording and was entered as a covariate in the ANCOVA model. The models also included random effects terms for *worker* and *analyst nested within worker*. Separate models were generated for each posture variable.

Among the analyses using the continuous visual analogue scale only one of the eight individual models (peak and average for wrist flexion, wrist extension, forearm supination, forearm pronation) resulted in a statistically significant ($p < 0.05$) relationship between work cycle variability in the posture and analyst error. This model was for inaccuracy of the estimate for peak forearm supination, which was associated with a coefficient estimate of 0.012. This coefficient expresses a change in the error in the posture estimate (in percentage of the joint ROM) per change in

Table 5. Kendall coefficients of Concordance, W , and intraclass correlation coefficients (ICC) for agreement among analysts

	Kendall, W		ICC	
	VAS	Six-category	Three-category	Six-category
Peak radial deviation	0.112	0.197	0.497	0.335
Peak ulnar deviation	0.024	0.219	0.111	0.033
Average radial deviation	0.278	—	—	—
Average ulnar deviation	0.191	—	—	—
Mode radial/ulnar deviation	—	0.017	0.044	0
Peak flexion	0.678	0.344	0.418	0.201
Peak extension	0.223	0.219	0.172	0.392
Average flexion	0.694	—	—	—
Average extension	0.236	—	—	—
Mode flexion/extension	—	0.472	0.096	0.434
Peak supination	0.533	0.415	0.590	0.366
Peak pronation	0.546	0.044	0.056	0.429
Average supination	0.321	—	—	—
Average pronation	0.328	—	—	—
Mode supination/pronation	—	0.129	0	0.129

standard deviation in the measured posture variable among the presented work cycles. The work cycle standard deviation for peak forearm supination ranged from 0° to 33° among the 12 conditions—three workers performing four jobs. Thus, for estimates of peak forearm supination obtained with the VAS, the expected increase in error due to between work cycle variability was 40% of the joint ROM (33 degree SD \times 0.012 per degree SD).

Among analyses using the six-category scale, three of six individual models (peak flexion, extension, supination, and pronation; mode flexion/extension and supination/pronation) resulted in statistical significance ($p < 0.05$) for the work cycle variability covariate. These were the models for peak wrist extension, peak forearm supination, and mode supination/pronation. The coefficient estimates for these covariates were 0.1283, 0.0456, and 0.2102, respectively. These coefficients express a change in the number of category units of error in the posture estimate per change in standard deviation in the measured posture variable among the presented work cycles. The differences between minimum and maximum work cycle standard deviations for the postures above were 5.9°, 33°, and 4.4°, respectively, which correspond to 0.76, 1.5, and 0.92 category units of error. This finding diminishes concerns about a high degree of variability in the postural dynamics of the ‘workers’ in the present study, and the adverse effect this might have on the accuracy of posture estimates.

Analysts’ self-reported years of experience in the field of ergonomics exhibited no meaningful correlation with posture estimation accuracy. ANCOVA models for individual posture variables revealed two models in which the model coefficient for the covariate *years experience* was significant ($p < 0.05$) among analysts using the six-category scaling method and two models in which the covariate was significant among analysts using the VAS method. The significant coefficient was actually positive in one case (indicating that *increasing* experience *increased* analyst inaccuracy) and negative in three cases. The significant negative coefficient estimates for years experience were -0.032 for peak supination and -0.039 for peak pronation using the six-category scale and -0.0184 for peak flexion with the VAS. For peak supination and pronation, these correspond to a decrease in error by approximately one category unit for 30 years experience. Thus, there was minimal evidence to suggest that more experienced ergonomists estimated posture more accurately, insofar as experience was broadly defined by self-reported years working in the field of ergonomics.

3.4. Completion time and accuracy

Since the analyses of these video recorded jobs were conducted by ergonomists at their own institutions on their own time they had a great deal of freedom in terms of strategies used to evaluate the jobs. Ergonomists were asked to self-time their analyses and to report the time to completion of the posture analysis for each of the four jobs. The self-reported time to completion was documented for each job, which included analyses of both the wrist/forearm and elbow/shoulder posture. (The validity of observational posture analysis of the elbow and shoulder is reported elsewhere.) Times to completion of the posture analyses averaged 33.9 ± 20.5 and 27.8 ± 9.2 min per job with the three- and six-category scales, respectively. This unanticipated trend was primarily due to two analysts in the three-category scale condition who were clearly outliers in that they took an average of 59 and 70 min respectively to complete the analysis of each job—more than double the times of the other analysts. With these outliers removed the mean times to completion were 25.2 ± 11.5 and 27.8 ± 9.2 min for the three- and six-category scales, respectively.

Differences in time to completion between the three- and six-category scale methods were not statistically significant, with or without the inclusion of the outliers. The posture analyses using the visual analogue scale were completed in an average of 20.0 min (SD = 9.7 min) per job.

A relationship between analysis time to completion and accuracy was expected. However, models regressing the magnitude of posture classification error against analysis completion time revealed no such relationship. Since no constraints on analysis completion time were imposed, a true speed-accuracy tradeoff could not be examined. The present data do suggest that time to completion of the analysis was not predictive of the resulting accuracy of the analysis.

4. Discussion

The design of the electrogoniometer calibration fixture used in this study resulted in an unacceptably low reliability of radial/ulnar deviation measurements. This may have been a result of some play in the mechanism for strapping the proximal phalangeal segment to fix the angle of radial/ulnar deviation. Because the range of motion available in radial/ulnar deviation is much smaller than in flexion/extension and pronation/supination, and the absolute error in the electrogoniometric calibrations in this axis were much larger, the decision was made to abandon the electrogoniometric measurements as a standard against which to express observational error for radial/ulnar deviation posture estimates. The results for radial/ulnar deviation were thus limited to examination of inter-observer agreement, rather than accuracy. The results indicated that agreement among analysts in terms of the rank orderings among the jobs for peak, average and most frequently occurring radial/ulnar deviation were clearly lower than those for flexion/extension and pronation/supination. Intraclass correlation coefficients appeared to be slightly lower for the radial/ulnar deviation postures. The Kappa statistics reported by Burt and Punnett (1999) for agreement between raters on the presence or absence of wrist postures indicated no clear differences between radial/ulnar deviation and flexion/extension or supination/pronation in terms of interrater agreement. However, in the present study, the lowest agreement among analysts appeared to be for estimates of radial/ulnar deviation posture. Radial/ulnar deviation may be the most difficult wrist posture to estimate visually because of the relatively small available range of angular displacement in this axis. The present data, in combination with those of other studies (Buchholz and Wellman 1997, Smutz *et al.* 1994, Spielholz 1998), also suggest that electrogoniometric measurements of radial/ulnar deviation are often less reliable than those of wrist flexion/extension and forearm pronation/supination.

A difficulty in calculating the accuracy of posture estimates made with the VAS is that the endpoint must be anchored to the extreme of the joint range of motion (ROM). This is problematic for several reasons. First, there is individual variability in joint range of motion and reported average joint ROMs vary widely from study to study. Second, joint ROMs reported in the clinical literature have been measured statically and neglect interdependence of joint postures. As an example, forearm pronation/supination is dependent upon elbow flexion angle (Schoenmarklin and Marras 1993) so that the ROM values reported for forearm rotation at a 90° elbow angle may underestimate the range of motion in this joint when the elbow is extended (Bonebrake *et al.* 1990). Third, normative clinical joint ROMs do not account for passive influence of other joints that can occur in occupational situations. An example of this was seen in Job C of the present study, involving

manual screwdriver use in which peak forearm pronation postures were measured in excess of 100°. Had the VAS scale been anchored to the forearm pronation ROM published by the American Academy of Orthopaedic Surgeons' (1965), 80°, the peak forearm pronation angle associated with the job would have exceeded 100% of the ROM. The ROM limits that were presented to analysts in this study to anchor the VAS extreme were taken from the study of Bonebrake *et al.* (1990) using their reported sample mean plus one standard deviation. This selection was made largely to insure that no worker/job would exhibit a peak wrist/forearm deviation in excess of the ROM limit to which the visual analogue scale was referenced. The study of Bonebrake *et al.* (1990) measured ROM in complex postures, combining wrist and elbow angles, which made them more appealing for application in occupational situations.

The choice of ROM limit used to anchor the extreme pole of the visual analogue scale clearly affects the magnitude of the measured posture when it is expressed as a percentage of the ROM. In the present study analysts were instructed to base their estimates of posture on the same ROMs as the measured posture values. Thus, the choice of the ROM would be expected to have no influence on the nature of the relationship between the estimated and measured postures. However, in real job analysis situations, where the magnitude of the postural exposure is of primary interest, the choice of the limit to the ROM used to anchor the visual analogue scale will influence the estimate of posture exposure.

An advantage of the VAS over the categorical scaling methods is that the latter are prone to ceiling effects. The apparent biases towards underestimation of peak posture, and in some cases the most frequently occurring (mode) posture, using the categorical scales are largely attributable to a ceiling effect introduced when the true posture level is in or near the highest posture category, so that errors in the posture estimate are mostly constrained towards the direction of underestimation. Comparison of the analysts' probabilities of misclassification by under-/over-estimation with those expected by chance confirms this ceiling effect, particularly with the higher precision, six-category scale. As an example, peak forearm pronation was underestimated in 32.5% and overestimated in 15% of occurrences, suggesting a bias towards underestimation. When these probabilities were expressed in relation to the probabilities of under-/over-estimation expected by chance, overestimation was more prevalent (2.24 times more likely than by chance) than was underestimation (0.54 times more likely than by chance). All instances of apparent biases towards underestimation were diminished or eliminated completely when misclassification probabilities were expressed relative to those expected by chance performance. The high proportion of instances in which the measured posture levels fell in the most extreme, or a near extreme, category resulted in ceiling effects that contributed heavily to the higher relative probabilities of underestimation.

A second disadvantage of categorical posture scales which the VAS avoids is that they require a priori selection of categorical boundaries to define levels of postural risk. These categorical boundaries have, in many cases, been selected rather arbitrarily. Juul-Kristensen *et al.* (1997) noted that several of the observational posture assessment methods which delineate specific categorical boundaries reference studies in which few or no specific categories are given. For instance, the 45° boundary common to several methods for categorization of back, neck, shoulder, and wrist angles has few references to support this as a value with any external

validity and may have been chosen mostly because ‘... it is an easily recognizable angle’ (Juul-Kristensen *et al.* 1997).

Posture misclassifications appear to be influenced by the boundary zone problem (Keyserling 1986) which occurs when the actual posture level falls near the boundary between two adjacent categories making the choice between the adjacent categories more difficult. This boundary zone problem might be expected to increase in proportion to the number of categories included in the scale since increasing the number of scale categories necessitates a proportionate reduction in category width. However, a simple relationship between the likelihood of misclassification and the number of scale categories was not evident. This can be seen in table 3 in which the probability of misclassifying either the extreme or the most frequently occurring posture was not a simple function of the number of scale categories. In addition, while the higher precision, six-category, scale was associated with greater likelihood of posture misclassification, the probability that a misclassification will be of more than one category error (i.e., ± 2 , ± 3 , ± 4 , or ± 5) is generally low. The fact that misclassifications by two or more categories are unlikely, combined with the fact that the effect of a posture misclassification between adjacent categories is lessened when more categories are available, indicates that there are several advantages of the higher precision six-category framework over that of three categories. Qualitatively, a six-category scale affords a low/medium/high registration in both directions for the bi-directional wrist/forearm postures (i.e., three categories in flexion, three in extension), while the three-category scale affords only a low/high, or neutral/non-neutral scoring in both directions. Thus, a trade-off appears to exist between the acceptability of a smaller likelihood of misclassification of a lower precision scale vs. a greater likelihood of misclassification of a higher precision scale. Results for time to completion suggest no advantage of the lower-precision categorical scale as analysis completion time with six categories was the same as completion time with only three categories.

Estimates of the temporal distribution of posture over the work cycle exhibited a central tendency error in which analysts appeared to be biased in favour of more neutral wrist/forearm postures. Juul-Kristensen *et al.* (2001) reported a bias towards non-neutral posture in which the percentage of the work cycle in the neutral posture was underestimated and the percentages in mild extension and ulnar deviation were overestimated. Figure 9 shows results for the temporal distribution of flexion/extension posture compared between the present study and that of Juul-Kristensen *et al.* (2001). Differences between the measured posture distributions of the two studies are relatively small (note the slight differences in posture scaling described in the figure caption). However, there is a marked difference between the studies in terms of a reversal of the over- and underestimations of neutral posture and mild extension posture. Results of the present study indicated that ergonomists tended to *underestimate* the severity of the posture exposure in terms of its duration as a percentage of the work cycle, while the results of Juul-Kristensen *et al.* (2001) indicated that this duration severity was *overestimated*. The Juul-Kristensen *et al.* findings should probably be interpreted more cautiously as their results were compiled from what appears to be a single analyst.

Baluyut *et al.* (1995) examined the performance of students with no formal training in posture analysis and found that approximately 70% of these students correctly classified wrist flexion/extension postures (with five posture categories)

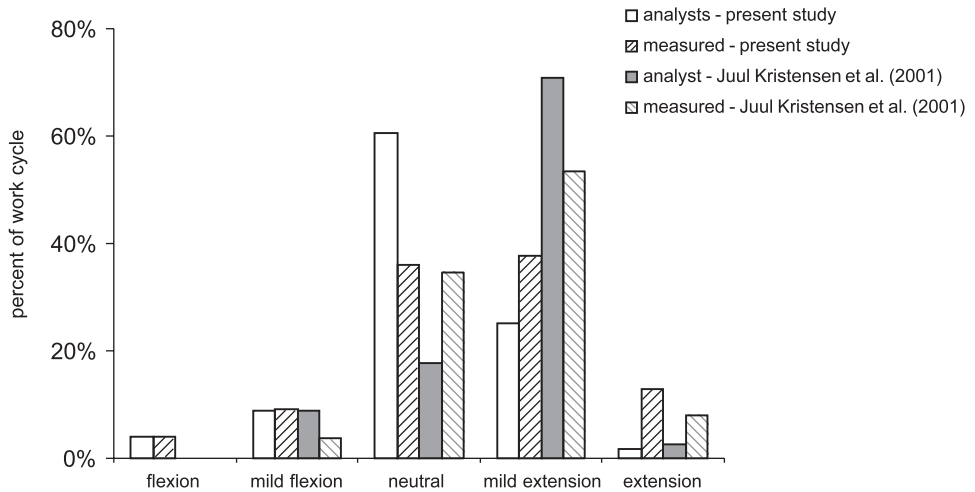


Figure 9. Comparison of temporal posture distributions between the present study and that of Juul-Kristensen *et al.* (2001). In the present study the 'neutral' category combined the percentages of the work cycle in the 0° – 20° flexion and 0° – 20° extension categories. Juul-Kristensen *et al.*'s (2001) neutral category was between 15° flexion and 15° extension. The 45° boundary was used to define mild flexion/flexion and mild extension/extension in both studies.

presented for 10 s as static images on a video monitor. These authors concluded that a training program was needed to eliminate difficulties in the evaluation of non-neutral postures encountered by the untrained observers, though the authors had no trained group of observers against whom the untrained observers' performance could be compared. Ketola *et al.* (2001) reported high sensitivity for two observers with 12 h of training and experience identifying non-neutral wrist posture when the reference standard was the estimate of these physical factors made by an expert. However, when comparing observers' estimates of posture to an instrumentation-based standard, that of electrogoniometric readings, sensitivity was shown to be low. The sensitivity of the untrained observers' estimates of non-neutral wrist posture appeared to be comparable to that of the expert. In the present study only ergonomists who reported having professional experience conducting job analyses participated, so the effect of a training program could not be evaluated. Untrained observers would seem to be less likely to conduct postural analyses for the purpose of ergonomic evaluation, so the effect of experience among trained ergonomists already possessing a high level of proficiency may be a more relevant variable to examine than is the difference in performance between trained and untrained observers. There was no clear effect of analysts' experience on accuracy of posture estimates.

There are many qualitative aspects of videographing industrial work that may result in a medium that affects an ergonomist's ability to accurately estimate wrist posture by observation. The goal of the present study was not necessarily to reflect the *best* performance that can be realized in video-based observational job analysis but, rather, it was intended to represent the *typical* performance that can be expected from ergonomists using pencil and paper-based approaches. In the present investigation a single digital camera was used to video record the jobs and analysts

were required to classify posture manually, using only features standard to VHS format video playback. There is likely to be some gain in posture classification accuracy when analysts are provided with either multiple simultaneous video views and/or the benefit of a more sophisticated computerized task/posture analysis system (e.g. Keyserling 1986, Yen and Radwin 2002). Another limitation of the present study was that the analysts were not involved in the acquisition of the video recording and did not have the opportunity to observe these jobs at first hand. The opportunity to observe the jobs at first hand in the work environment would be expected to improve analysts' abilities to accurately estimate aspects of working posture.

Standardization of methods for scaling wrist/forearm posture for the purpose of observational analysis would be beneficial for epidemiological dose-response studies because the effect of working posture could be more readily compared between studies. Such standardization should consider the external validity of the scaling method (i.e., the validity of the categories in terms of their association with disease prevalence) as well as internal validity (i.e., the reliability of the observational exposure measures relative to a physical measurement). Epidemiologic exposure-response studies are needed to determine the external validity of exposure assessment methods. The present laboratory-based study was intended to outline capabilities of ergonomists conducting video-based observational posture analysis that should be considered in the development and application of methods for observational posture analysis of the wrist and forearm.

5. Conclusions and future research direction

The objective of this study was to quantify the accuracy of ergonomists' video-based observational estimates of working posture of the wrist/forearm obtained with representative methods for scaling and documenting posture as a risk factor for WMSDs. The results of this investigation revealed possible limitations of these methods that may be summarized as follows:

- Ergonomists made errors in the classification of peak and most frequently occurring postures. The likelihood of a posture misclassification error was dependent upon the number of categories in the posture scale and was as high as a 77% probability of misclassification for extreme wrist extension in the three categories of 0°–20°, 20°–45°, > 45° extension, and as low as a 14% probability of misclassification for peak forearm pronation in the two categories of 0°–40°, > 40°.
- Ergonomists' estimates of peak and average wrist flexion/extension and forearm supination/pronation posture using a continuous visual analogue scale exhibited low correlation with these measured postures. Visual analogue scales may be problematic for estimating absolute levels of posture because they must be anchored to a single value selected as a limit to the range of motion for the particular joint posture.
- Inter-analyst agreement among ratings of jobs according to the severity of radial/ulnar deviation posture were lower than those among ratings according to flexion/extension and supination/pronation. This finding suggests that estimates of radial/ulnar deviation may be less reliable than those of wrist flexion/extension and forearm supination/pronation.

- Use of a categorical scale with six categories took no longer for posture analysis than did a categorical scale with only three categories. No relationships between analysis completion time and analysis accuracy were evident with any posture of the scaling methods.
- Estimates of the temporal distribution of working posture for the typical work cycle exhibited a central tendency error. The percentages of the work cycle in the neutral postures tended to be overestimated and the percentages in more extreme postures tended to be underestimated.

The present study evaluated the accuracy of posture analysis from video observation in which analysts were presented with a single video view and a 'paper and pencil' method of analysis. Future studies will evaluate the accuracy of posture analyses when the analysts are provided with more sophisticated techniques to aid the analysis, such as multiple synchronous video views and/or the benefit of a computerized task/posture analysis system. A limitation of the present study was the fact that jobs were simulated in a laboratory environment and were necessarily somewhat contrived. Evaluation of posture analysis validity in realistic work environments should prove fruitful but challenging.

Acknowledgements

The author wishes to acknowledge the assistance of Dan Habes and Ahmed Khalil with data collection and of Ed Krieg with statistical support and consultation.

References

- AMERICAN ACADEMY OF ORTHOPAEDIC SURGEONS 1965, *Joint Motion. Method of Measuring and Recording* (Chicago, IL: American Academy of Orthopaedic Surgeons).
- ARMSTRONG, T. J., FOULKE, J. A., JOSEPH, B. S. and GOLDSTEIN, S. A. 1982, Investigation of cumulative trauma disorders in a poultry processing plant, *American Industrial Hygiene Association Journal*, **43**(2), 103–116.
- BALUYUT, R., GENAIDY, A. M., DAVIS, L. S., SHELL, R. L. and SIMMONS, R. J. 1995, Use of visual perception in estimating static postural stresses: magnitudes and sources of error, *Ergonomics*, **38**(9), 1841–1850.
- BATY, D., BUCKLE, R. W. and STUBBS, D. A. 1986, Posture recording by direct observation, questionnaire assessment, and instrumentation: a comparison based on a recent field study, in N. Corlett, J. Wilson and I. Manenica (eds), *The Ergonomics of Working Postures* (Philadelphia: Taylor and Francis), 283–292.
- BONEBRAKE, A. R., FERNANDEZ, J. E., MARLEY, R. J., DAHALAN, J. B. and KILMER, K. J. 1990, A treatment for Carpal Tunnel Syndrome: Evaluation of objective and subjective measures, *Journal of Manipulative and Physiological Therapeutics*, **13**(9), 507–520.
- BUCHHOLZ, B. and WELLMAN, H. 1997, Practical operation of a biaxial goniometer at the wrist joint, *Human Factors*, **39**(1), 119–129.
- BURT, S. and PUNNETT, L. 1999, Evaluation of interrater reliability for posture observations in a field study, *Applied Ergonomics*, **30**, 121–135.
- DE LOOZE, M. P., TOUSSAINT, H. M., ENSINK, J., MAGNUS, C. and VAN DER BEEK, A. J. 1994, The validity of visual observation to assess posture in a laboratory-simulated, manual material handling task, *Ergonomics*, **37**(8), 1335–1343.
- DRURY, C. G. 1987, A biomechanical evaluation of the repetitive motion injury potential of industrial jobs, *Seminars in Occupational Medicine*, **2**, 41–49.
- GENAIDY, A. M., SIMMONS, R. J., GUO, L. and HIDALGO, J. A. 1993, Can visual perception be used to estimate body part angles?, *Ergonomics*, **36**(4), 323–329.

- HOLZMANN, P. 1982, ARBAN – A new method for analysis of ergonomic effort, *Applied Ergonomics*, **13**(2), 82–86.
- JUUL-KRISTENSEN, B., FALLENTIN, N. and EKDAHL, C. 1997, Criteria for classification of posture in repetitive work by observation methods: A review, *International Journal of Industrial Ergonomics*, **19**, 397–411.
- JUUL-KRISTENSEN, B., HANSSON, G.-A., FALLENTIN, N., ANDERSEN, J. H. and EKDAHL, C. 2001, Assessment of work postures and movements using a video-based observation method and direct technical measurements, *Applied Ergonomics*, **32**, 517–524.
- KEMMLERT, K. 1995, A method for the identification of ergonomic hazards – PLIBEL, *Applied Ergonomics*, **26**(3), 199–211.
- KETOLA, R., TOIVONEN, R. and VIIKARI-JUNTURA, E. 2001, Interobserver repeatability and validity of an observation method to assess physical loads imposed on the upper extremities, *Ergonomics*, **44**(2), 119–131.
- KEYSERLING, W. M. 1986, Postural analysis of the trunk and shoulders in simulated real time, *Ergonomics*, **29**(4), 569–583.
- KILBOM, A. 1994, Assessment of physical exposure in relation to work-related musculoskeletal disorders – what information can be obtained from systematic observations?, *Scandinavian Journal of Work Environment and Health*, **20** (special issue), 30–45.
- KILBOM, A., PERSSON, J. and JONSSON, B. G. 1986, Risk factors for work-related disorders of the neck and shoulder – with special emphasis on working postures and movements, in N. Corlett, J. Wilson and I. Manenica (eds), *The Ergonomics of Working Posture* (London: Taylor and Francis), 44–54.
- LATKO, W. A. 1997, *Development and evaluation of an observational method for quantifying exposure to hand activity and other physical stressors in manual work*. Unpublished doctoral dissertation. The University of Michigan, Ann Arbor, Michigan.
- LESKINEN, T., HALL, C., RAUAS, S., ULIN, S., TONNES, M., VIIKARI-JUNTURA, E. and TAKALA, E. 1997, Validation of portable ergonomic observation (PEO) method using optoelectronic and video recordings, *Applied Ergonomics*, **28**(2), 75–83.
- MCAATAMNEY, L. and CORLETT, E. N. 1993, RULA: a survey method for the investigation of work-related upper limb disorders, *Applied Ergonomics*, **24**(2), 91–99.
- MOORE, J. S. and GARG, A. 1995, The strain index: a proposed method to analyze jobs for risk of distal upper extremity disorders, *American Industrial Hygiene Association Journal*, **56**, 443–458.
- OCHIPINTI, E. 1998, OCRA: a concise index for the assessment of exposure to repetitive movements of the upper limb, *Ergonomics*, **41**(9), 1290–1311.
- SCHOENMARKLIN, R. W. and MARRAS, W. S. 1993, Dynamic capabilities of the wrist joint in industrial workers, *International Journal of Industrial Ergonomics*, **11**, 207–224.
- SETH, V., WESTON, R. L. and FREIVALDS, A. 1999, Development of a cumulative trauma disorder risk assessment model for the upper extremities, *International Journal of Industrial Ergonomics*, **23**, 281–291.
- SIEGEL, S. and CASTELLAN, N. J. 1988, *Nonparametric Statistics for The Behavioral sciences*, 2nd edn. (New York: McGraw-Hill).
- SMUTZ, P., SERINA, E. and REMPEL, D. 1994, A system for evaluating the effect of keyboard design on force, posture, comfort, and productivity, *Ergonomics*, **37**(10), 1649–1660.
- SOROCK, G. S. and COURTNEY, T. K. 1996, Epidemiologic concerns for ergonomists: illustrations from the musculoskeletal disorder literature, *Ergonomics*, **39**(4), 562–578.
- SPIELHOLZ, P. 1998, Development of an electrogoniometer calibration procedure for measurement of wrist angle and forearm rotation, in S. Kumar (ed.), *Advances in Occupational Ergonomics and Safety* (Washington, DC: IOS Press), 499–502.
- SPIELHOLZ, P., SILVERSTEIN, B., MORGAN, M., CHECKOWAY, H. and KAUFMAN, J. 2001, Comparison of self-report, video observation and direct measurement methods for upper extremity musculoskeletal disorder physical risk factors, *Ergonomics*, **44**(6), 588–613.
- STETSON, D., KEYSERLING, W. M., SILVERSTEIN, B. A. and LEONARD, J. A. 1991, Observational analysis of the hand and wrist: a pilot study, *Applied Occupational and Environmental Hygiene*, **6**(11), 927–937.
- STOCK, S. R. 1991, Workplace ergonomic factors and the development of musculoskeletal disorders of the neck and upper limbs: a meta-analysis, *American Journal of Industrial Medicine*, **19**, 87–107.

- VAN DER BEEK, A. J. and FRINGS-DRESEN, M. H. 1998, Assessment of mechanical exposure in ergonomic epidemiology, *Occupational and Environmental Medicine*, **55**(5), 291–299.
- WINKEL, J. and MATHIASSEN, S. E. 1994, Assessment of physical work load in epidemiologic studies: concepts, issues, and operational considerations, *Ergonomics*, **37**(6), 979–988.
- WINKEL, J. and WESTGAARD, R. 1992, Occupational and individual risk factors for shoulder-neck complaints. Part II – The scientific basis (literature review) for the guide, *International Journal of Industrial Ergonomics*, **10**, 85–104.
- YEN, T. Y. and RADWIN, R. G. 2002, A comparison between analysis time and inter-analyst reliability using spectral analysis of kinematic data and posture classification, *Applied Ergonomics*, **33**, 85–93.

Copyright of Ergonomics is the property of Taylor & Francis Ltd and its content may not be copied or emailed to multiple sites or posted to a listserv without the copyright holder's express written permission. However, users may print, download, or email articles for individual use.