Interrater Reliability of Posture Observations

Stephen Bao and **Ninica Howard,** Washington State Department of Labor and Industries, Olympia, Washington, **Peregrin Spielholz,** Sound Transit, Seattle, Washington, **Barbara Silverstein,** Washington State Department of Labor and Industries, Olympia, Washington, and **Nayak Polissar,** The Mountain-Whisper-Light Statistical Consulting, Seattle, Washington

Objective: The aims of this research were (a) to study the interrater reliability of a posture observation method, (b) to test the impact of different posture categorization systems on interrater reliability, and (c) to provide guidelines for improving interrater reliability. **Background:** Estimation of posture through observation is challenging. Previous studies have shown varying degrees of validity and reliability, providing little information about conditions necessary to achieve acceptable reliability. **Method:** Seven raters estimated posture angles from video recordings. Different measures of interrater reliability, including percentage agreement, precision, expression as interrater standard deviation, and intraclass correlation coefficients (ICC), were computed. **Results:** Some posture parameters, such as the upper arm flexion and extension, had ICCs ≥ 0.50. Most posture parameters had a precision around the 10° range. The predefined categorization and 30° posture categorization strategies showed substantially better agreement among the raters than did the 10° strategy. **Conclusions:** Different interrater reliability measures described different aspects of agreement for the posture observation tool. The level of agreement differed substantially between the agreement measures used. Observation of large body parts generally resulted in better reliability. Wider width angle intervals resulted in better percentage agreement compared with narrower intervals. For most postures, 30°-angle intervals are appropriate. Training aimed at using a properly designed data entry system, and clear posture definitions with relevant examples, including definitions of the neutral positions of the various body parts, will help improve interrater reliability. **Application:** The results provide ergonomics practitioners with information about the interrater reliability of a postural observation method and guidelines for improving interrater reliability for video-recorded field data.

INTRODUCTION

Awkward postures are considered one of the risk factors of work-related musculoskeletal disorders (National Institute for Occupational Safety and Health [NIOSH], 1997). Measurement of work postures has been performed in numerous studies (e.g., Aarås, 1994; Holmstrom, Lindell, & Mortiz, 1992; Juul-Kristensen et al., 2002; Punnett, Fine, Keyserling, Herrin, & Chaffin, 2000; Silverstein et al., 2006). Ergonomic practitioners also have the need to collect posture data to perform various job risk evaluations with the use of different ergonomic risk assessment tools (e.g., Buchholz, Paquet, Punnett, & Lee, 1996; Hignett & McAtamney, 2000; McAtamney & Corlett, 1993; Moore & Garg, 1995).

Different posture measurement methods are available for use. They can be grouped into three main categories: direct measurement, self-reporting, and observation. Direct measurement, such as the use of electrogoniometers and potentiometers, has been used in several ergonomics studies (e.g., Aarås, 1994; Hansson et al., 2001). This method usually requires mounting sensors on the body parts of interest and recording their motions

Address correspondence to Stephen Bao, Washington State Department of Labor and Industries, SHARP Program, P.O. Box 44330, Olympia, WA 98504-4330; baos235@lni.wa.gov. ***HUMAN FACTORS,*** Vol. 51, No. 3, June 2009, pp. 292-309. DOI: 10.1177/0018720809340273.

during work. Validity and reliability of this type of measurement can be high if the technique is used properly.

However, it takes a large amount of effort to prepare test participants, calibrate instruments, analyze data, and interpret results. Because of technological limitations (e.g., the number of available input channels of an instrument and available data storage spaces), this method is usually limited to a small number of body parts and participants. Improper use of this method, such as improper placement of sensors, may introduce systematic errors, negatively affecting the measurement validity. Costs associated with direct measurement are usually high, making its use impractical in large musculoskeletal epidemiological studies or by most ergonomics practitioners.

The method of asking workers by questionnaires or interview about their postures during work (self-reporting method) has been used in a number of musculoskeletal epidemiological studies (Balogh, et al., 2001; Punnett, 1998). It is relatively economical and can usually be used among large study populations. In addition, this may be the only way to get historical exposure data. However, the specificity, validity, and reliability of self-reporting methods are generally considered low (Hansson et al., 2001). Ergonomics practitioners seldom use this method in job evaluations.

Posture observation methods, applied either on site or from recorded videos, have been used widely by both researchers and ergonomics practitioners (Hignett & McAtamney, 2000; Juul-Kristensen, Fallentin, & Ekdahl, 1997; Punnett et al., 2000). However, estimating posture angles can be challenging. Validity, which measures how close the estimated posture angle is to the true angle, and reliability, which measures how well different raters agree to each other, are two important qualities of a posture observation method. Varying degrees of validity and reliability have been reported in studies (Burt & Punnett, 1999; Landis & Kock, 1977; Leskinen et al., 1997).

A recent validity study on a posture observation method found that posture observation by trained users can provide reasonable 3-dimensional data to calculate cumulative low back loads with a biomechanical model (Sutherland, Albert, Wrigley, & Callaghan, 2008). It has been reported that better estimation validity is achieved from observing larger body parts (shoulder and elbow; Lowe, 2004a) than smaller body parts (wrists and forearm; Lowe, 2004b). Validity studies are commonly conducted in laboratory settings or among a small number of participants in the field when a posture observational method is used in combination with a direct measurement method. Validity studies are seldom conducted in the field among large study populations.

Interrater reliability of posture observations is another important property of posture observation methods. It measures how well different raters are in agreement about a posture angle they see. Although the reliability of posture observations has been studied by several researchers (e.g. Burt & Punnett, 1999; Landis & Kock, 1977; Leskinen et al., 1997), most studies stopped at simply reporting the degree of reliability and failed to provide users guidelines on improving reliability and its implication in risk quantification for specific applications. Denis, Lortie, and Rossignol (2000) reviewed 72 papers on observation procedures characterizing occupational physical activities and found little information about the conditions necessary to achieve good reliability. It was our hypothesis that a newly developed proprietary posture observational tool (Bao, Howard, Spielholz, & Silverstein, 2007), which was based on still video frames rather than on-site observations, is reliable and therefore useful in large-scale epidemiological studies quantifying work postures.

There is little consensus about what statistical methods are best to analyze interrater agreement. Several common methods include kappa statistic, percentage of agreement between raters, and intraclass correlation coefficient (ICC). For example, Burt and Punnett (1999) found interrater agreement of 26% to 99%, depending on body parts. But agreement was only moderate when using kappa statistic.

Researchers have used different posture categorization systems of varied angle widths (e.g., Fransson-Hall, et al., 1995; Juul-Kristensen, et al., 1997; McAtamney & Corlett, 1993).

Dichotomized posture categories have been used in many epidemiological studies (NIOSH, 1997). The impact of different posture categorizations on interrater reliability has not been well studied.

The present study aimed at evaluating the reliability of a posture observational method that was used in a large epidemiological field study of upper extremity musculoskeletal disorders. Validity issues are not addressed in the present study because of limitations of the data set, although they were studied in a separate laboratory study using the same posture observational method (unpublished data). Specific goals of the present study were

- to study the interrater reliability of a posture observational method that was used in a large field study,
- to test the impact of different posture categorization systems on interrater reliability, and
- to provide user guidelines on improving interrater reliability in posture observations for field epidemiological study use.

MATERIALS AND METHODS

Seven experienced raters participated in an exercise of observing work postures of four different jobs on video. Three of the raters were professional ergonomists, and the remaining were laboratory technicians. The ergonomists have an extensive theoretical background in terms of posture definitions (i.e., the definitions of posture motions at the standard anatomic positions) compared with the technicians, who received training from the ergonomists. However, the technicians had more practical experience in using the posture observational tool and had more opportunity than the ergonomists to discuss, in the normal course of their jobs, the posture observations and definitions under various circumstances.

Four video-recorded jobs performed by different workers were chosen from a large field study data pool. These jobs were very different in nature with large variations in work postures and workplace environment. The jobs were a laundry handler in a commercial laundry facility (work was performed in multiple locations in the same facility), a lumber handler in a sawmill (work was performed in a single location with very fast pace), an assembler in an electronics plant (a job with a well-defined cyclic pattern), and a pharmacist in a large hospital pharmacy (a long-cycle job without a defined cyclic pattern).

Each job was videotaped from two angles during a typical workday using two synchronized camcorders for a period of 15 min. Ideally, the two cameras should have been placed orthogonally to each other. However, in practice, this was not always possible because of constraints of the workplace layout and work activities. In situations such as these, the camera crews were instructed to coordinate with each other to position the cameras in two distinct angles from the worker. The 15-min video recording was digitized in a sampling frequency of 30 frames per second. A specially designed data entry program (Bao et al., 2007) enabled the two synchronized video clips to be viewed simultaneously on a data entry computer screen.

Posture angles were estimated from 37 or 38 randomly selected frames in each of the four video clips by each rater. We selected the frames using the random-number function in the Microsoft Excel program. This frame selection method was used in the large field study (Bao, Silverstein, Howard, & Spielholz, 2006). Posture angles of 20 different body parts were estimated for each video frame. All 7 raters estimated the posture angles of the same selected video frames; however, they were blinded to each other's estimations. The raters estimated the approximate joint angles of the various body parts by clicking on a point on a posture diagram displayed on a computer screen instead of entering a numerical angle value in degrees. An angular value (in degrees), equal to the body position, was automatically entered into a database after the click. If a rater considered a posture angle not visible, he or she was instructed to enter a *missing* data code in the data-recording program. Details about the click-on-screen method were reported previously (Bao et al., 2006, 2007).

Based on the raters' estimated posture data, three different posture categorization strategies were used in postdata processing. This process

TABLE 1: A Predefined Posture Categorization (PDC) Definition (in degrees)

	Posture Category				
Posture Parameter	1	2	3	4	5
Left (L) elbow flexion	<–5	[–5, 20)	[20, 60)	[60, 100)	≥100
Right (R) elbow flexion					
L forearm supination (–)/pronation (+)	<–45	[–45, 45)	≥45		
R forearm supination (–)/pronation (+)					
Neck flexion (+)/extension (–)	<–5	[–5, 20)	≥20		
Neck lateral flexion	[0, 10)	[10, 30)	≥30		
Neck twisting	[0, 10)	[10, 45)	≥45		
Trunk flexion/extension	<–5	[–5, 20)	[20, 60)	≥60	
Trunk lateral flexion	[0, 10)	[10, 30)	≥30		
Trunk twisting	[0, 10)	[10, 45)	≥45		
L upper arm flexion (+)/extension (–)	<–5	[–5, 20)	[20, 45)	[45, 90)	≥90
R upper arm flexion (+)/extension (–)					
L upper arm abduction (+)/adduction (–)	<–5	[–5, 30)	[30, 60)	[60, 90)	≥90
R upper arm abduction (+)/adduction (–)					
L upper arm inward (+)/outward rotation (+)	<–5	[–5, 15)	[15, 45)	≥45	
R upper arm inward (+)/outward rotation (+)					
L wrist flexion (+)/extension (–)	<–45	[–45, –15)	[–15, 15)	[15, 45)	≥45
R wrist flexion (+)/extension (–)					
L wrist ulnar (+)/radial (–) deviation	<–15	[–15, 20)	≥20		
R wrist ulnar (+)/radial (–) deviation					

Note. Neutral posture category in shaded cells. [- including the angle,) – not including the angle

did not require the raters to reestimate the postures. In this stage, the estimated posture data were grouped into categories according to three different categorization strategies: two fixed-width categorization strategies (10° and 30° intervals) and a predefined categorization (PDC) strategy (Table 1). The PDC strategy was designed according to several commonly used posture observational methods (Buchholz et al., 1996; Fallentin et al., 2001; Fransson-Hall et al., 1995; Hignett & McAtamney, 2000; Keyserling, Stetson, Silverstein, & Brouwer, 1993). These different categorization strategies were used to study the impact of adjusting widths of categorization intervals on interrater reliability. Percentages of observed postures in the different angular categories for each job were calculated from the different categorization strategies.

To discover if there were patterns of good and poor agreements, a systematic qualitative review of posture angle estimates of selected video frames was conducted. The video frames of the smallest and the largest between-rater standard deviations (i.e., frames corresponding to the best and worst agreement among the raters) for each of the studied posture parameters were selected. In this qualitative review, issues regarding camera angles, quality of videos, individual observation variations, human coding mistakes, posture definitions, and other possible factors were noted by two of the professional ergonomists.

Statistics

To examine characteristics of the estimated postural data, the mean estimate of the 7 raters for each of the 20 posture parameters was computed for each video frame. Using the data set of the means per frame, we calculated descriptive statistics (median, the 5th and 95th percentiles) for each job. Percentages of times that postures falling in the neutral and non-neutral positions using the PDC strategy (Figure 1) were also calculated.

A four-factor analysis of variance (ANOVA) was performed for all posture parameters to compare mean posture angle between the four different jobs and between ergonomists and technicians. The ANOVA model included random effects (for the individual rater and the frame being viewed) and fixed effects (for the job and the contrast between ergonomists vs. technicians). The random-effects ANOVA model appropriately addressed the repeated measures (multiple observations by each rater and multiple ratings of each frame). A significance level of $p = .05$ was used.

ANOVAs were also conducted to obtain the between-frames mean square, the between-raters mean square and the residual mean square for each posture parameter. As a measure of interrater consistency, an ICC was computed for each of the different postural parameters in each job. The ICC measured the variability among raters in comparison to the variability of the posture angles across all of the frames rated. ICCs were also computed separately for posture estimations made by the ergonomists and technicians. The ICC designated by Shrout and Fleiss (1979) as ICC(2,1) was used. In our context, ICC(2,1) was based on the assumption that a set of raters was chosen at random from a population of raters and that each of the raters estimated a given posture in all of the frames. The numerical interpretation and validity of the ICC did not depend on having a random sample of raters. The ability to generalize the findings to other settings required that the results from these raters were typical of raters doing this kind of work.

Standard deviation of posture estimations was computed for each video frame among all raters. This is a measure of precision, that is, the random error of posture estimates by the raters (Rothman & Greenland, 1998), and is related to the quality of reliability. Unbiased standard deviation estimates were calculated by pooling the standard deviations across all frames using the standard root mean square method. Similar calculations were performed for postures (average postures at each video frame) in the predefined neutral and nonneutral posture categories for each of the 20 postures (Table 1). The unbiased standard deviations were also computed for the ergonomists and technicians separately.

Using the categorized posture data by the three different posture categorization strategies (PDC, 10° intervals, and 30° intervals), we computed percentages of agreement in category assignment. The agreement percentage was calculated for each possible pair of raters, and then the mean of these percentages was computed across all pairs of raters. The mean of these agreement percentages was computed across all pairs of raters separately for each of the four jobs, and then the mean of these job-specific means was calculated across all four jobs. These agreement percentages were also presented separately for the neutral and nonneutral posture categories to examine degree of agreement in different posture ranges.

To determine percentage agreement for neutral and nonneutral categories separately, each frame was placed into the neutral or nonneutral posture category on the basis of the mean of all of the raters' estimations for the given frame. As before, percentage agreement was calculated for each pair of raters across all frames falling into the neutral category, and then the mean of these percentages was calculated for all pairs of raters to yield one summary percentage value for the neutral category. A similar percentage was calculated for the frames falling into the nonneutral category. These percentages of agreement were also computed separately for the ergonomists and technicians and then were compared. All statistical analyses were performed using the SAS statistical program (Version 9E; SAS Institute, Cary, NC).

RESULTS

Figure 1 shows the median, 5th and 95th percentiles of average postures across the 7 raters for the four different jobs. Different jobs had varied posture deviation patterns. For example, the laundry handler had larger estimated angular deviations of upper arm flexion and extension (Figure 1c) and forearm supination and pronation (Figure 1a) compared to the other jobs. The sawmill lumber handler had larger trunk and neck flexion and extension but less neck lateral flexion angular deviations compared with the other jobs (Figure 1b). The electronics assembly

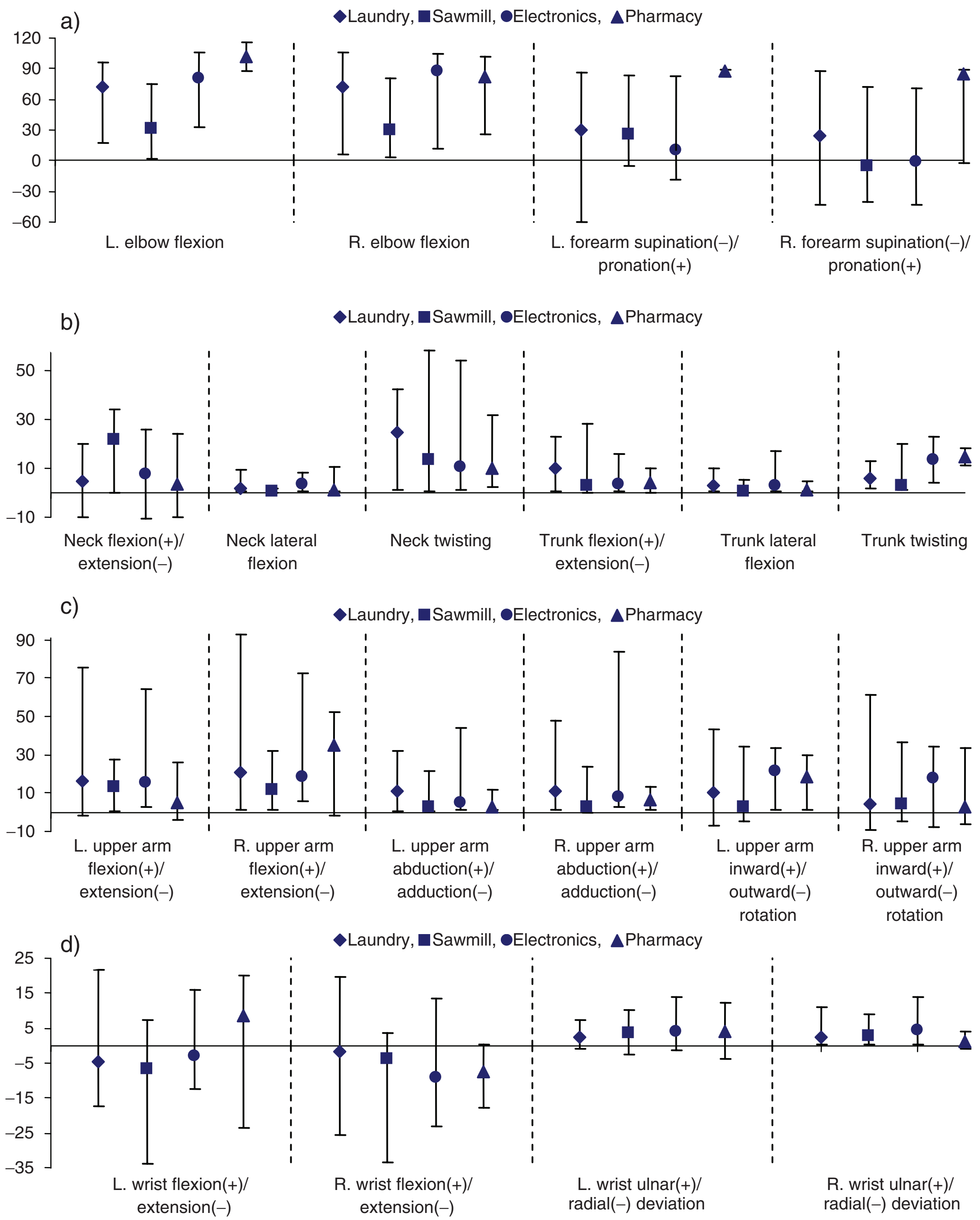

Figure 1. Median and corresponding 5th and 95th percentiles of frame-specific mean posture angles from the 7 raters of the four different jobs: (a) elbow flexion and forearm supination and pronation of left and right body side; (b) neck flexion and extension, neck lateral flexion, neck twisting, trunk flexion and extension, trunk lateral flexion, and trunk twisting; (c) flexion and extension, abduction and adduction, and inward and outward rotation of left and right upper arm; and (d) flexion and extension and ulnar and radial deviation of left and right wrist.

TABLE 2: Percentage of Observed Postures in Neutral And Nonneutral Posture Categories and Percentage of Missing Observations (missing observations were not included in the analyses)

	% of Observed Postures		
Posture Parameter	Neutral Posture Category	Nonneutral Posture Category	% of Missing Observations
Left (L) elbow flexion	52	48	1.6
Right (R) elbow flexion	44	56	1.8
L forearm supination/ pronation	42	58	2.6
R forearm supination/ pronation	50	50	2.7
Neck flexion/extension	62	38	0.0
Neck lateral flexion	89	11	0.1
Neck twisting	44	56	0.3
Trunk flexion/extension	91	9	0.0
Trunk lateral flexion	86	14	0.0
Trunk twisting	71	29	0.1
L upper arm flexion/ extension	63	37	0.7
R upper arm flexion/ extension	51	49	1.0
L upper arm abduction/ adduction	94	6	0.9
R upper arm abduction/ adduction	90	10	1.6
L upper arm inward/ outward rotation	50	50	1.1
R upper arm inward/ outward rotation	57	43	1.2
L wrist flexion/extension	64	36	3.7
R wrist flexion/extension	68	32	4.2
L wrist ulnar/radial deviation	97	3	4.2
R wrist ulnar/radial deviation	97	3	4.2

job had larger angular deviations of trunk lateral flexion (Figure 1b), upper arm abduction and adduction (Figure 1c), and wrist ulnar and radial deviation (Figure 1d) compared with other jobs. According to posture estimations, the job of the pharmacist required less postural deviations on most postures, such as trunk postures (Figure 1b), left elbow flexion, and left forearm supination and pronation (Figure 1a).

The ANOVAs showed significant differences in means between the jobs for all posture parameters ($p < .05$). There were no significant differences between the ergonomists and technicians in terms of the mean posture angles of the different jobs in 12 out of the 20 different postures ($p > .05$). However, the technicians estimated higher flexion angles for the trunk, upper arm flexion and extension (both body sides), and left wrist flexion and extension postures than did the ergonomists ($p < .05$). Overall, the ergonomists appeared to estimate more neck flexion, trunk twisting, and wrist ulnar deviation (both body sides) postures than did the technicians ($p < .05$).

Table 2 shows the distribution of postures in neutral and nonneutral posture categories according to the PDC strategy. Some posture parameters (neck lateral flexion, trunk flexion and extension, trunk lateral flexion, upper arm abduction and adduction, and wrist ulnar and radial deviation) fell predominantly in the neutral posture categories (>85% of observed postures). There were fewer missing observations (when raters determined the posture parameter could not be observed at the frame location) in the trunk, neck, and upper arm posture parameters compared with hand and wrist posture parameters (Table 2).

The ICCs were better for some posture parameters than for others (Table 3). The posture parameters of left and right elbow flexion, left and right forearm supination and pronation,

TABLE 3: Intraclass Correlation Coefficients (ICC) by Job

Posture Parameter	Laundry	Sawmill	Electronics	Pharmacy	Mean ICC
Left (L) elbow flexion	.66	.64	.70	.21	.55
Right (R) elbow flexion	.85	.67	.76	.80	.77
L forearm supination/pronation	.75	.32	.63	.41	.53
R forearm supination/pronation	.60	.41	.51	.82	.59
Neck flexion/extension	.40	.46	.49	.51	.46
Neck lateral flexion	.28	.02	.12	.35	.19
Neck twisting	.42	.65	.64	.30	.50
Trunk flexion/extension	.55	.78	.44	.18	.49
Trunk lateral flexion	.22	.37	.49	.07	.29
Trunk twisting	.17	.43	.02	.00	.15
L upper arm flexion/extension	.75	.41	.78	.42	.59
R upper arm flexion/extension	.77	.46	.56	.80	.65
L upper arm abduction/adduction	.29	.69	.34	.25	.39
R upper arm abduction/adduction	.46	.50	.74	.10	.45
L upper arm inward/outward rotation	.59	.64	.40	.39	.51
R upper arm inward/outward rotation	.75	.58	.52	.66	.63
L wrist flexion/extension	.33	.42	.33	.63	.43
R wrist flexion/extension	.51	.31	.33	.25	.35
L wrist ulnar/radial deviation	.06	.16	.36	.20	.20
R wrist ulnar/radial deviation	.18	.05	.27	.07	.14

neck twisting, left and right upper arm flexion and extension, and left and right upper arm inward and outward rotation had average ICCs ≥ 0.50. The worst ICCs were seen among left and right wrist ulnar and radial deviation, neck lateral flexion, and trunk twisting postures (ICCs ≤ 0.20). There were large variations between ICCs of the same posture parameters across the different jobs. For example, the ICC for trunk flexion and extension was 0.78 for the sawmill job, compared with only 0.18 for the pharmacy job, and the ICCs of upper arm flexion and extension were greater than 0.70 for the laundry job but less than 0.50 for the sawmill job (Table 3). This seemed to correspond to the ranges of postural deviations of that job (Figure 1). In general, ICCs between the technicians were better than those between the ergonomists (17 out of the 20 posture parameters).

The pooled between-rater standard deviations for all posture parameters are shown in Table 4. Standard deviations for the posture parameters indicated varied precisions for the different posture parameters. Most standard deviations were less than 15°, except for trunk twisting and forearm supination and pronation (Table 4). The standard deviations of posture estimates between the neutral and nonneutral posture categories were different (Table 4). The standard deviations were generally smaller in the posture categories, including 0° postures. The technicians had smaller between-rater standard deviations compared to the ergonomists in 18 out of the 20 posture parameters (excluding the left and right wrist flexion and extension postures).

Table 5 shows mean percentages of agreement between raters for the four jobs when the three posture categorization strategies were used. The percentages of agreement were different between jobs and between posture parameters. For example, the percentage of agreement of elbow flexion was higher for the laundry job (72% and 71% for the left and right body side, respectively) than for the sawmill job (55% and 65% for the left and right body side, respectively). The agreement was better for the trunk flexion and extension posture compared with the wrist flexion and extension posture (Table 5). The percentages of agreement also differed between posture categorization strategies. The larger the posture category width was,

TABLE 4: Precision of Measurements of Posture Parameters (in degrees as measured by between-rater standard deviation)

Posture Parameter	All Data	Within Neutral Posture Categories[a]	Within Nonneutral Posture Categories
Left (L) elbow flexion	14.5	21.8	11.7
Right (R) elbow flexion	14.3	18.3	11.8
L forearm supination/pronation	24.2	29.1	18.0
R forearm supination/pronation	27.8	31.5	20.8
Neck flexion/extension	11.1	10.0	12.9
Neck lateral flexion	3.8	3.7	5.7
Neck twisting	12.2	7.6	14.6
Trunk flexion/extension	5.5	5.4	7.1
Trunk lateral flexion	4.2	3.8	7.5
Trunk twisting	19.6	5.4	25.8
L upper arm flexion/extension	10.6	8.8	14.3
R upper arm flexion/extension	12.8	9.5	16.5
L upper arm abduction/ adduction	9.6	8.6	24.4
R upper arm abduction/ adduction	11.2	9.2	26.4
L upper arm inward/ outward rotation	10.6	8.7	12.4
R upper arm inward/ outward rotation	10.4	7.8	13.2
L wrist flexion/extension	12.2	10.6	15.3
R wrist flexion/extension	11.9	10.6	15.6
L wrist ulnar/radial deviation	6.0	5.6	9.3
R wrist ulnar/radial deviation	5.3	5.0	8.5

Note. Precision measured by pooled unbiased standard deviations across all frames estimated by the raters

a. Posture categories are defined by the predefined posture categorization strategy.

the better the percentages of agreement were. The PDC and 30° posture categorization strategies showed substantially better agreement among the raters than did the 10° categorization strategy.

Several posture parameters (wrist ulnar and radial deviation, upper arm abduction and adduction, and trunk lateral flexion postures) appeared to have higher overall percentages of agreement. When stratified into neutral and nonneutral categories, there was generally better agreement among the neutral postures (Table 6). However, for some postures, the agreement of posture estimations in the nonneutral posture categories was much lower (at least 24% less for the wrist ulnar and radial deviation, upper arm abduction and adduction, and trunk lateral flexion postures). The overall agreement for wrist ulnar and radial deviation was 95% in the neutral posture categories, compared with 58% in the nonneutral categories (Table 6). For other posture parameters, the differences of agreement between neutral and nonneutral posture categories were less than 13% (Table 6).

The differences in percentage of agreement among pairs of ergonomists versus pairs of technicians were less than 10% for all posture parameters, except the upper arm flexion and extension and wrist flexion and extension postures (15% to 19%).

The systematic qualitative review of the selected video frames by the 2 ergonomists revealed distinctive factors for obtaining good or poor agreement results for the different posture parameters (Table 7). The issues differed by body parts but were similar between the left and right side of a body segment. In general, poor agreement occurred when cameras were in the movement plane of a particular posture (e.g., in the sagittal plane for trunk flexion and extension, in the elbow flexion movement plane for elbow flexion, and from the same side as the hand with wrist flexion and extension). For some postures (elbow flexion, forearm

TABLE 5: Mean Percentage Agreement Between Pairs of Raters for Three Posture Categorization Strategies, by Posture and Job

	Laundry			Sawmill			Electronics			Pharmacy		
Posture	PDC	10°	30°	PDC	10°	30°	PDC	10°	30°	PDC	10°	30°
Left (L) elbow flexion	72	26	58	55	23	49	69	35	66	59	21	52
Right (R) elbow flexion	71	32	61	65	25	56	64	30	59	63	29	62
L forearm supination/pronation	77	24	46	59	16	32	81	23	52	100	72	98
R forearm supination/pronation	76	25	50	60	16	34	79	21	45	94	49	78
Neck flexion/extension	61	31	64	66	38	64	61	30	66	54	26	66
Neck lateral flexion	83	67	96	100	96	100	70	60	91	86	72	94
Neck twisting	58	29	53	62	40	61	56	36	58	55	33	56
Trunk flexion/extension	81	47	69	95	69	95	90	51	86	98	52	84
Trunk lateral flexion	74	63	85	95	84	98	73	59	83	90	76	98
Trunk twisting	60	46	69	77	65	88	53	43	56	70	64	73
L upper arm flexion/extension	64	30	59	63	34	60	64	25	53	69	36	74
R upper arm flexion/extension	56	26	54	74	35	59	59	24	54	60	30	66
L upper arm abduction/adduction	83	37	68	96	56	86	87	43	76	99	63	93
R upper arm abduction/adduction	79	35	67	94	55	89	87	34	62	90	44	69
L upper arm inward/outward rotation	48	27	55	70	54	77	58	26	60	61	29	64
R upper arm inward/outward rotation	57	33	64	68	49	73	59	28	63	72	43	80
L wrist flexion/extension	58	34	58	54	36	54	60	32	60	45	23	45
R wrist flexion/extension	65	36	65	55	37	55	52	23	52	66	30	66
L wrist ulnar/radial deviation	99	58	92	93	52	84	92	51	81	94	43	83
R wrist ulnar/radial deviation	96	57	92	94	59	86	90	44	74	99	60	98

Note. PDC = predefined posture categorization (see Table 1).

TABLE 6: Mean Percentage of Agreement of Posture Estimates Within Neutral and Nonneutral Posture Categories

Posture Parameter	All Data	Within Neutral Posture Category	Within Nonneutral Posture Category
Left (L) elbow flexion	64	65	54
Right (R) elbow flexion	66	70	60
L forearm supination/ pronation	79	74	80
R forearm supination/ pronation	77	73	79
Neck flexion/extension	60	66	55
Neck lateral flexion	84	85	91
Neck twisting	58	66	53
Trunk flexion/extension	91	92	86
Trunk lateral flexion	83	86	62
Trunk twisting	65	64	65
L upper arm flexion/extension	65	68	57
R upper arm flexion/extension	62	67	56
L upper arm abduction/ adduction	91	93	67
R upper arm abduction/ adduction	88	90	54
L upper arm inward/ outward rotation	59	59	53
R upper arm inward/ outward rotation	64	66	57
L wrist flexion/extension	54	47	57
R wrist flexion/extension	60	53	62
L wrist ulnar/radial deviation	95	95	58
R wrist ulnar/radial deviation	95	95	58

Note. Posture was categorized by the predefined posture categorization strategy.

supination and pronation, neck flexion and extension, and trunk flexion and extension), one camera view was often sufficient for good agreement as long as it was not in the corresponding movement plane.

However, for the remaining posture parameters, two cameras were usually needed to achieve good agreement results. Some posture positions caused confusion in posture definitions. This problem worsened when cameras were placed in unfavorable positions (Table 7). Figure 2 shows an example of an upper arm flexion and extension posture as seen by two synchronized cameras. The cameras were not ideally positioned. The upper arm was abducted almost 90° and was close to the frontal plane (although not in the frontal plane). As a result, disagreement between the raters occurred. Three raters estimated the upper arm flexed to 99° to 109°, whereas the other four estimated it to be 0° to 12°.

DISCUSSION AND CONCLUSION

Methodology Considerations

Several different methods are available for interrater agreement evaluations. Percentage of agreement between raters and ICC were used in the present study. Raw percentage agreement is a simple and important descriptive statistic, providing a straightforward interpretation of agreement between raters. However, it does not distinguish between agreements on neutral versus nonneutral range postures. To remedy this limitation, many researchers have used kappa statistic in posture reliability studies (e.g., Burt & Punnett, 1999; de Bruijn, Engels, & van der Gulden, 1998; Pan et al., 1999). However, the kappa statistic is influenced by posture distributions. Highly agreed-upon posture estimates may still receive a low kappa statistic.

In the current study, we used another remedy for the limitation of raw percentage agreements by simply calculating percentages of specific agreement for different posture ranges

TABLE 7: Posture-Specific Factors Causing Poor Agreement

Posture Parameter	Factors That Contribute to Poor Agreement
Elbow flexion	• Rater's coding error (coded 0° instead of 90° or vice versa) • Poor visibility on the elbow
Forearm supination/ pronation	• Rater's coding error (coded supination instead of pronation or vice versa) • Difficult posture parameter with large between-rater variations
Neck flexion/extension	• Participant's hair blocking camera's view
Neck twisting	• Rater's coding error attributable to reference-neutral position in the analysis tool • Difficult posture parameter with large between-rater variations
Trunk lateral flexion	• Camera in the frontal plane
Trunk twisting	• Rater's coding error attributable to reference-neutral position in the analysis tool • Difficult posture parameter with large between-rater variations
Upper arm flexion/ extension	• Obvious rater's coding error • Upper arm close to the frontal plane, resulting in difficult upper arm flexion/extension definition • Upper arm away from the sagittal plane, resulting in larger between-rater variations
Upper arm abduction/ adduction	• Obvious rater's coding error • Upper arm close to the sagittal plane resulted in difficult upper arm abduction/adduction definition • Upper arm away from in the frontal plane resulted in larger between-rater variations
Upper arm inward/ outward rotation	• Obvious rater's coding error • Difficult to observe when elbow close to 0° resulted in larger between-rater variations • Blocked view when camera was behind participant
Wrist flexion/extension	• Large between-rater variations with only dorsal camera view • Blocked view to the hand because of work object • Poor view attributable to background lighting and/or worker's clothing color
Wrist ulnar/radial deviation	• Large between-rater variations with only a side (ulnar/radial) camera view • Blocked view to the hand because of work object • Poor view attributable to background lighting and/or worker's clothing color

(Cicchetti & Feinstein, 1990). For example, the overall percentage of agreement of right upper arm abduction and adduction posture estimates between the raters was 88% but only 54% for postures within the nonneutral posture categories (Table 6). This has practical importance in epidemiological studies when people are more interested in the interrater reliability of categorizing risky versus nonrisky postures.

A second problem with the percentage agreement is that it is quite dependent on the number of categories used in the rating system. The percentage agreement is likely to be higher with a two-category system than with a five-category system, making comparison across rating methodologies difficult. For a given posture parameter, the wider the posture category is, the fewer the number of categories. One practical solution is to examine the agreement for a group of hazardous categories versus nonhazardous categories, thus focusing the evaluation to only these two important angular ranges. Dichotomizing

Figure 2. Large discrepancy on right upper arm flexion and extension occurred on this video frame among raters because of camera positions and definition issues.

exposure parameters is also a common practice in epidemiological studies. Labeling work posture between neutral and nonneutral is relatively easy, and the data can often be used directly in epidemiological analyses.

Whereas percentage of agreement and kappa statistic can be used for categorical posture data, the ICC is used for continuous data. In the present study, the ICC is the ratio of the variance of the posture estimates between the raters to the total variance of the ratings (Shrout & Fleiss, 1979). This also gives an indication of the relative differences between the raters' estimates of the variations in the work postures. A problem with the ICC is that given the same magnitude of variations among raters' estimates for individual frames, a posture with smaller frame-to-frame variation will result in poorer ICCs compared with postures with larger variations. In the current study, left elbow flexion and trunk flexion postures, which had smaller variations, had lower ICCs (0.21 and 0.18, respectively) than the other postures (Figure 1 and Table 3). This problem is similar to that for kappa statistic, namely, that the measure depends heavily on the overall variation in the phenomenon being studied and is, therefore, difficult to compare from study to study.

The level of agreement may differ substantially according to the measure of agreement used. For example, the percentage of agreement between raters for trunk flexion and extension posture was high (91% overall agreement, 92% within neutral posture category, and 86% within nonneutral posture categories; Table 6), and high precision was observed in the measure of a low between-rater standard deviation (5.5°; Table 4). However, the ICC was only in the moderate range (Table 3) because of the smaller variations of the trunk flexion and extension posture during the job performances (Figure 1 and Table 2).

The measures of agreement used in this study have different qualities. Among them, the percentage agreement for the more hazardous (nonneutral) and less hazardous (neutral) posture categories, as presented in Table 6, is the preferred measure. This approach allowed the specific examination of agreement for high-risk postures. It was also an approach that had something in common with the "proportion of specific agreement," which in our context was the probability that a second randomly chosen rater would agree with the first randomly chosen rater who assigned a specific hazardous category to a video frame (Fleiss, Levin, & Paik, 2003).

Agreement of Posture Observations Between Raters

Agreement of posture observations between raters is a complicated issue. Many factors can influence that agreement. These may include, but are not limited to, (a) specific posture parameters, (b) posture variation of a study population or in a job, (c) posture distribution between different angular categories, (d) posture categorization strategy, (e) rater training and experiences, (f) rater position relative to a study participant (or quality of a video image), (g) types of jobs, (h) posture definition and rater estimation error, and (i) the tools used, such as interface designs and instructions. The current study was limited to the available data already obtained from the large field study. Further research is needed to systematically study the various influences on interrater reliability and validity of posture observations.

Some posture parameters were easier to observe than others. In general, postures of larger body parts, such as the trunk and the arms, were easier to observe than smaller body parts (e.g., wrists). In general, this was reflected by larger ICCs (Table 3), smaller between-rater standard deviations (Table 4), and higher percentages of agreement (Table 5 and 6) in some posture parameters of the larger body parts (e.g., trunk flexion and extension and upper arm flexion and extension). There were more missing data (unobservable postures attributable to video quality) for the smaller body parts (wrists and forearms; Table 2) compared with the larger body parts. However, some exceptions were found in the current study.

For example, the ICC for the trunk flexion and extension of the pharmacists was only 0.18 (Table 3), although the percentages of agreement were quite high (Table 5). This was because the ICCs were influenced by the amount of posture variation of a particular body part, as discussed previously. The pharmacists had little variation in trunk flexion and extension postures (Figure 1). This finding illustrates that posture variation is related to the activities the person performs and/or individual work techniques. However, between different posture parameters, it is also related to the range of motion of the joints. For example, the average ranges of motion are 188° and 27° for the upper arm flexion and wrist radial deviation, respectively (Chaffin & Andersson, 1991). The 95th percentile estimated angles of the right upper arm flexion and right wrist radial deviation was 94° and 14°, respectively, for the laundry handler, compared with 29° and 15°, respectively, for the sawmill lumber handler (Figure 1).

When posture data were categorized according to the PDC strategy, the between-rater reliability varied for data within different angular categories. With the exception of elbow flexion and forearm supination and pronation, posture parameters had better precisions for postures within the neutral posture category compared with the nonneutral category (as measured by the between-rater standard deviation; Table 4). However, percentages of agreement were not always better in the neutral posture categories compared with the nonneutral posture categories (Table 6).

It might be expected that the increase in the width of posture categories could increase the percentages of agreement. This was found to be true in the present study. Larger posture category widths (e.g., the PDC and 30° posture categories; Table 5) showed higher percentages of agreement than posture categories with smaller width (10° posture categories; Table 5). In general, the 10° category width resulted in lower percentages of agreement for most of the posture parameters. On average, it was about 95% lower than the PCD and 78% lower than the 30° category. For some posture categories, angle interval width of 30° or larger may be needed to achieve a higher percentage of between-rater agreement. This was reflected in the precision measures for forearm supination and pronation, which had larger between-rater standard deviations (Table 4).

This finding has an important practical implication that posture evaluation based on observation may not be able to involve the use of a posture category width of smaller than 30°. Many existing posture observation systems (e.g., PATH (Buchholz et al., 1996), OWAS (de Bruijn et al., 1998), and RULA (McAtamney & Corlett, 1993)) involve the use of wide posture category widths for different postures. Pan et al. (1999) found adequate or good reliability between raters (kappa statistics ranged from 0.5 to 0.63) for the PATH parameters, including postures. PATH posture parameters are fairly crude categorical measures of the large body parts (Buchholz et al., 1996). Similar kappa statistic results were found in a study using OWAS parameters (percentage agreement > 85% and kappas > 0.6), on which some of the PATH posture parameters are based (de Bruijn et al., 1998). These good reliability results might have been the product of observing only larger body parts and using wide width posture categories.

Rater training and experience seems to have played an important role in some of the results. The technicians tended to have better estimation precisions (as measured by the between-rater standard deviations) compared with the ergonomists for 18 of the 20 posture parameters. They also tended to have higher ICCs than the ergonomists for most of the posture parameters (17 out of 20). Compared with the ergonomists, who had better theoretical knowledge of posture definitions, all the technicians in this study had similar training on this particular type of posture analysis, spent more time discussing the various posture definitions under different circumstances during the performance of video analysis, and had many more hours in performing posture estimations.

These experiences seemed to have helped the technicians get more consistent estimations in more complicated situations. For example, the theoretical definition of upper arm flexion is the movement of the humerus anteriorly in the sagittal plane and is usually measured in the standard anatomical position (erect stance, face forward, and arms down at the side with palms facing forward, etc.). In reality, upper arms may not be in the sagittal plane, and the worker may not be in the standard anatomical position. The technicians were able to compile a list of complicated situations (not in the experiment of the current study but in the course of the large field study in which the present posture observation tool was used) and seek solutions to standardize their posture estimations.

Covalla (2003) reported that male observers had better posture estimation capacity than female observers. The male:female ratio in the present study was 1:3 for laboratory technicians and 2:1 for professional ergonomists. Although no formal gender effect analysis was performed in the present study, our finding regarding the technicians and ergonomists may still hold true. Technicians, with their more extensive experience in using this specific posture observational tool, seemed to have better agreement in posture estimation than the ergonomists. Given the small sample size, the comparisons between ergonomists and technicians needs to be further investigated.

In spite of a better precision in posture estimation among the technicians, when posture data were categorized, the percentages of agreement between technicians and ergonomists were quite similar (less than 10% difference for most of the posture parameters). Both ergonomists and technicians were able to detect between-job posture differences according to the two-way ANOVA results.

The quality of the video recording appeared to be another important influence in the reliability of posture observations, especially for certain posture parameters (Table 7). Several factors related to poor video recording quality that might have linked to low interrater reliability were identified. The first factor was poor camera positioning in relation to the posture being observed. For example, the camera was only in the sagittal plane for trunk and neck flexion or in the frontal plane for trunk and neck lateral flexion. Wrist flexion and extension posture was filmed only from the dorsal side, wrist ulnar and radial deviation posture from the ulnar or radial side, and elbow flexion posture from the same plane as the elbow flexion and extension movement plane.

The second factor was objects blocking the views, for example, hair blocking neck postures or working object blocking the wrist and forearm postures. This more often happened for the hand and wrist postures (up to about 4%; Table 2) than for the trunk, neck, and upper arm postures (up to about 1%; Table 2). The third factor was poor background lighting, insufficient lighting, and poor contrast because of the color of the workers' clothing.

Camera angle has been found to have a significant impact on the accuracy of posture estimations, particularly on small body parts, such as the wrists (Lau & Armstrong, 2007), although this impact may be small on large body parts, such as the trunk and upper arms (Sutherland, Albert, Wrigley, & Callaghan, 2007). In comparison, for on-site observations, raters can adjust their positions to obtain an adequate view. However, on-site posture observation has its drawbacks as well. The most significant one is that the information-processing capacity of human beings is limited, and people are not able to process multiple items of information simultaneously. Van der Beek, van Gaalen, and Frings-Dresen (1992) found that observing multiple exposure parameters simultaneously decreased interrater reliability. Estimating postures on recorded video gives raters an opportunity to view video recordings multiple times and/or in slow motion. Training of ergonomists or technicians to take quality video recordings should result in better posture estimations from recorded videos.

One of the limitations of the present study was that some important parameters (e.g., different lighting, camera angles, and camera distances) were not well controlled compared with those in a laboratory study. However, one of the advantages that field-based studies can offer is that they can reveal problems of posture observations in a real work environment. This in turn can be used for the design of future laboratory studies.

Tables 3 and 5 show that interrater reliability was different among the jobs. This might be explained by two factors described previously: (a) posture variability for a job (e.g., smaller left elbow flexion angle range of the pharmacist job [Figure 1] resulted in worse ICCs [Table 3] compared with other jobs) and (b) video recording quality attributable to the nature and location of the different jobs.

The left and right hand data were analyzed and presented separately, although some of the results appear to be similar for the same joints on both body sides. One explanation of the similar findings might be that each side is similar in terms of the ease or difficulty of observation. However, they are independent—meaning that posture of one side may not necessarily mirror the posture of the opposite side. This is dependent on what the operator is doing. In terms of interrater reliability of the left and right body side, camera angles, which often differed on each side of the body, may have been one of the most important factors. This could have resulted in different estimation reliability for the left and right body sides.

Obvious human error in posture estimation and confusion with posture definitions contributed to some of the largest disagreements between raters (Table 7). At times, there appears to have been a misinterpretation of the neutral reference posture in our data entry diagrams. Therefore, the layout of the data entry system should be improved to avoid the confusion.

For the elbow flexion and forearm supination and pronation postures, coding 0° elbow flexion instead of 90° or supination instead of pronation (or vice versa) occurred several times. For the neck and trunk twisting postures, one rater coded 0° twisting instead of 90°. This could have been prevented if the click method had been used rather than entering numerical posture data directly. Training on the use of the posture observation program would reduce this type of mistake.

Inconsistent posture definitions occurred in estimating the upper arm flexion and extension postures when the upper arm was close to the frontal plane and in estimating the upper arm abduction and adduction postures when the upper arm was close to the sagittal plane. Clear definitions and training with examples may help reduce this type of error. During the course of using the posture observation program, a list of difficult situations was compiled. Solutions have been developed and may be used in the future to improve the posture observational tool. In the example used in the results, the upper arm close to the frontal plane caused confusion among the raters (Figure 2). We emphasized in the training to the technicians that the upper arm flexion and extension and abduction and adduction are projections of the upper arm in the sagittal and frontal planes. Therefore, the upper arm posture used in the example should be about 90° abduction and 90° flexion. From the physiological point of view, this coding is also more reasonable than coding it to 0° flexion, as the later would not represent the high risk at that upper arm posture (Figure 2).

On the basis of our results and the discussion earlier, we conclude the following:

- Different measures of interrater reliability (percentage agreement, precision expressed as the interrater standard deviation, and ICC) described different aspects of agreement of posture observations among multiple raters and resulted in substantially different levels of agreement.
- In general, larger body parts were easier to observe and resulted in better reliability. However, reliability also depended on such factors such as variability of the posture parameters, camera positions, video quality, and complicated work postures.

- Wider posture category widths usually resulted in better proportions of agreement compared with smaller posture category widths. For most postures, 30° posture categorization was appropriation, except for forearm pronation and supination, which required even wider widths. Categorized postures with wide categorization widths could be used in epidemiological studies, even though we found that the posture observations among multiple raters had relatively low precision and ICCs.
- Targeted training aimed at avoiding common errors together with an improved data entry system should help improve interrater reliability.

ACKNOWLEDGMENTS

We acknowledge the important contributions of Ruby Irving, Cindy Orr, Jessica Keller, and Larry Taing in posture estimations. Blazej Neradilek provided statistical assistance. This research was funded in part by the U.S. National Institute for Occupational Safety and Health (OH1007316) and the Washington State Department of Labor and Industries.

REFERENCES

Aarås, A. (1994). The impact of ergonomic intervention on individual health and corporate prosperity in a telecommunications environment. *Ergonomics*, *37*, 1679–1696.

Balogh, I., Orbaek, P., Winkel, J., Nordander, C., Ohlsson, K., & Ektor-Andersen, J. (2001). Questionnaire-based mechanical exposure indices for large population studies: Reliability, internal consistency and predictive validity. *Scandinavian Journal of Work Environment and Health*, *27*, 41–48.

Bao, S., Howard, N., Spielholz, P., & Silverstein, B. (2007). Two posture analysis approaches and their application in a modified Rapid Upper Limb Assessment evaluation. *Ergonomics*, *50*, 2118–2136.

Bao, S., Silverstein, B., Howard, N., & Spielholz, P. (2006). The Washington State SHARP approach to exposure assessment. In W. S. Marras & W. Karwowski (Eds.), *Fundamentals and assessment tools for occupational ergonomics* (pp. 44-1–44-22). Boca Raton, FL: Taylor & Francis.

Buchholz, B., Paquet, V., Punnett, L., Lee, D., & Moir, S. (1996). PATH: A work sampling-based approach to ergonomic job analysis for construction and other non-repetitive work. *Applied Ergonomics*, *27,* 177–187.

Burt, S., & Punnett, L. (1999). Evaluation of interrater reliability for posture observations in a field study. *Applied Ergonomics*, *30*, 121–135.

Chaffin, D. B., & Andersson, G. B. J. (1991). *Occupational biomechanics*. New York: Wiley.

Cicchetti, D. V., & Feinstein, A. R. (1990). High agreement but low kappa: II. Resolving the paradoxes. *Journal of Clinical Epidemiology*, *43*, 551–558.

Covalla, E. D. (2003). *Visual posture observation error and training*. Master's thesis, North Carolina State University, Raleigh.

de Bruijn, I., Engels, J. A., & van der Gulden, J. W. (1998). A simple method to evaluate the reliability of OWAS observations. *Applied Ergonomics*, *29*, 281–283.

Denis, D., Lortie, M., & Rossignol, M. (2000). Observation procedures characterizing occupational physical activities: Critical review. *International Journal of Occupational Safety and Ergonomics*, *6*, 463–491.

Fallentin, N., Juul-Kristensen, B., Mikkelsen, S., Andersen, J. H., Bonde, J. P., Frost, P., & Endahl, L. (2001). Physical exposure assessment in monotonous repetitive work: The PRIM study. *Scandinavian Journal of Work Environment and Health*, *27*, 21–29.

Fleiss, J. L., Levin, B., & Paik, M. C. (2003). *Statistical methods for rates and proportions.* New York: Wiley.

Fransson-Hall, C., Gloria, R., Kilbom, A., Winkel, J., Karlqvist, L., Wiktorin, C., & Stockholm MUSIC I Study Group. (1995). A portable ergonomic observation method (PEO) for computerized on-line recording of postures and manual handling. *Applied Ergonomics*, *26*, 93–100.

Hansson, G. A., Balogh, I., Bystrom, J. U., Ohlsson, K., Nordander, C., Asterland, P., Sjolander, S., Rylander, L., Winkel, J., & Skerfving, S. (2001). Questionnaire versus direct technical measurements in assessing postures and movements of the head, upper back, arms and hands. *Scandinavian Journal of Work Environment and Health*, *27*, 30–40.

Hignett, S., & McAtamney, L. (2000). Rapid entire body assessment (REBA). *Applied Ergonomics*, *31*, 201–205.

Holmstrom, E. B., Lindell, J., & Mortiz, U. (1992). Low back and neck/shoulder pain in construction workers: Occupational workload and psychosocial risk factors: Part 1. Relationship to low back pain. *Spine*, *17*, 663–671.

Juul-Kristensen, B., Fallentin, N., & 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., Falletin, N., Hansson, G.-A., Madeleine, P., Andersen, J. H., & Ekdahl, C. (2002). Physical workload during manual and mechanical deboning of poultry. *International Journal of Industrial Ergonomics*, *29*, 107–115.

Keyserling, W. M., Stetson, D. S., Silverstein, B. A., & Brouwer, M. L. (1993). A checklist for evaluating ergonomic risk factors associated with upper extremity cumulative trauma disorders. *Ergonomics*, *36*, 807–831.

Landis, J. R., & Kock, G. G. (1977). The measurement of observer agreement for categorical data. *Biometrics*, *33*, 159–174.

Lau, M. H., & Armstrong, T. J. (2007). Parallax effects on estimating wrist posture: What are we really seeing? [Abstract] In *Proceedings of Sixth International Scientific Conference on Prevention of Work-Related Musculoskeletal Disorders (PREMUS)* (p. 101). Rome: International Commission on Occupational Health.

Leskinen, T., Hall, C., Rauas, S., Ulin, S., Tonnes, M., Viikari-Juntura, E., & Takala, E. P. (1997). Validation of Portable Ergonomic Observation (PEO) method using optoelectronic and video recordings. *Applied Ergonomics*, *28*, 75–83.

Lowe, B. D. (2004a). Accuracy and validity of observational estimates of shoulder and elbow posture. *Applied Ergonomics*, *35*, 159–171.

Lowe, B. D. (2004b). Accuracy and validity of observational estimates of wrist and forearm posture. *Ergonomics*, *47*, 527–554.

McAtamney, L., & Corlett, E. N. (1993). RULA: A survey method for the investigation of work-related upper limb disorders. *Applied Ergonomics*, *24*, 91–99.

Moore, J. S., & 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.

National Institute for Occupational Safety and Health. (1997). *Musculoskeletal disorders and workplace factors: A critical review of epidemiologic evidence for work-related musculoskeletal disorders of the neck, upper extremity, and low back.* Cincinnati, OH: U.S. Department of Health and Human Services, Author.

Pan, C. S., Gardner, L. I., Landsittel, D. P., Hendricks, S. A., Chiou, S. S., & Punnett, L. (1999). Ergonomic exposure assessment: An application of the PATH systematic observation method to retail workers. Postures, Activities, Tools and Handling. *International Journal of Occupational and Environmental Health*, *5*, 79–87.

Punnett, L. (1998). Ergonomic stressors and upper extremity disorders in vehicle manufacturing: Cross-sectional exposure-response trends. *Occupational and Environmental Medicine*, *55*, 414–420.

Punnett, L., Fine, L. J., Keyserling, W. M., Herrin, G. D., & Chaffin, D. B. (2000). Shoulder disorders and postural stress in automobile assembly work. *Scandinavian Journal of Work Environment and Health*, *26*, 283–291.

Rothman, K. J., & Greenland, S. (1998). *Modern epidemiology.* Philadelphia: Lippincott Williams & Wilkins.

Shrout, P., & Fleiss, J. (1979). Intraclass correlations: Uses in assessing rater reliability. *Psychological Bulletin*, *86*, 420–428

Silverstein, B. A., Viikari-Juntura, E., Fan, Z. J., Bonauto, D. K., Bao, S., & Smith, C. (2006). Natural course of nontraumatic rotator cuff tendinitis and shoulder symptoms in a working population. *Scandinavian Journal of Work Environment and Health*, *32*, 99–108.

Sutherland, C. A., Albert, W. J., Wrigley, A. T., & Callaghan, J. P. (2007). The effect of camera viewing angle on posture assessment repeatability and cumulative spinal loading. *Ergonomics*, *50*, 877–889.

Sutherland, C. A., Albert, W. J., Wrigley, A. T., & Callaghan, J. P. (2008). A validation of a posture matching approach for the determination of 3D cumulative back loads. *Applied Ergonomics*, *39*, 199–208.

van der Beek, A. J., van Gaalen, L. C., & Frings-Dresen, M. H. (1992). Working postures and activities of lorry drivers: A reliability study of on-site observation and recording on a pocket computer. *Applied Ergonomics*, *23*(5), 331–336.

Stephen Bao is a senior ergonomist in the Safety and Health Assessment and Research for Prevention (SHARP) Program in the Washington State Department of Labor and Industries in Olympia and an affiliate assistant professor in the Department of Environmental and Occupational Health Sciences at the University of Washington in Seattle. He received his PhD in industrial ergonomics from Luleå University of Technology in Luleå, Sweden, in 1995.

Ninica L. Howard is a senior researcher with the SHARP Program in the Washington State Department of Labor and Industries in Olympia. She received her MSc in work environment from the University of Massachusetts in Lowell in 1997.

Peregrin Spielholz is the chief safety officer with the Sound Transit in Seattle, Washington, and an affiliate assistant professor in the Department of Environmental and Occupational Health Sciences at the University of Washington in Seattle, where he received his PhD in environmental health (ergonomics) in 1999.

Barbara A. Silverstein is the research director for the Washington State Department of Labor and Industries SHARP Program and an affiliate professor at the University of Washington in the School of Public Health, Department of Environmental and Occupational Health Sciences, in Seattle. She received her PhD in epidemiology from the University of Michigan in Ann Arbor in 1985.

Nayak L. Polissar is the principal statistician and owner of the Mountain-Whisper-Light Statistical Consulting firm and an affiliate associate professor in the Department of Biostatistics at the University of Washington in Seattle. He received his PhD in statistics from Princeton University in Princeton, New Jersey, in 1974.

Date received: September 18, 2007
Date accepted: May 10, 2009