

Assessing Agreement of Self-reported and Observed Physical Exposures of the Upper Extremity

ANN MARIE DALE, JAIME STRICKLAND, BETHANY GARDNER, JUERGEN SYMANZIK, BRADLEY ALLEN EVANOFF

Assessment of workplace physical exposures by self-reported questionnaires has logistical advantages in population studies, but is subject to exposure misclassification. This study measured agreement between eight self-reported and observer-rated physical exposures to the hands and wrists, and evaluated predictors of inter-method agreement. Workers ($n = 341$) from three occupational categories (clerical/technical, construction, and service) completed self-administered questionnaires and worksite assessments. Analyses compared self-reported and observed ratings using a weighted kappa coefficient. Personal and psychosocial factors, presence of upper extremity symptoms, and job type were evaluated as predictors of agreement. Weighted kappa values were substantial for lifting (0.67) and holding vibrating tools (0.61), moderate for forceful grip (0.58), and fair to poor for all other exposures. Upper extremity symptoms did not predict greater disagreement between self-reported and observed exposures. Occupational category was the only significant predictor of inter-method agreement. Self-reported exposures may provide a useful estimate of some work exposures for population studies. *Key words:* upper extremity; epidemiologic studies; physical exertion; carpal tunnel syndrome; hand; wrist; work; surveys; questionnaires; bias.

INT J OCCUP ENVIRON HEALTH 2010;16:1-10

INTRODUCTION

The measurement of physical work exposures is critical to studying exposure-response relationships in work-

related musculoskeletal disorders (WRMSDs). WRMSDs affected 333,760 workers in the United States in 2007.¹ To study the factors associated with the development of these disorders, there must be adequate methods available to quantify the physical work exposures. Physical exposures vary widely between jobs due to the differences in types of tasks performed by workers, frequency and duration of task performance, and intensity levels within the tasks.² Several methods have been used to assess physical exposures including worker self-reports, observation, and direct physical measurements. There are logistical trade-offs between different methods;^{3,4} in large epidemiological studies, worker self-report provides the simplest and most cost-effective method for measuring worker physical exposures.⁵

There have been many studies evaluating the validity and reliability of self-reported surveys to assess physical exposures attributed to disorders of the lower back, lower extremities, and upper extremities.⁶ In validity studies, self-reported surveys are often compared to reference methods such as direct observation,⁷⁻⁹ observed videotaped work samples,^{3,10} and direct measurement.^{11,12} Survey items vary by study, addressing issues such as the types of tasks (walking, sitting), the characteristics of the exposures (time, intensity, body posture), and the associated injury risks. Stock et al. reviewed self-reported surveys and found that the reproducibility and validity of a variety of survey items ranged from poor to good.⁶ It is unclear what factors, such as job type or personal or cultural differences, may contribute to this wide range in validity.

Although hand use is common in most jobs and the frequency of upper extremity disorders is high in many occupations, there are few surveys directed toward tasks and exposures involving use of the hands and wrists. Nordstrom and colleagues¹³ conducted a validity study with carpal tunnel patients using survey questions from the 1988 Occupational Health Supplement to the National Health Interview Survey conducted by the US Department of Health and Human Services.¹⁴ The survey addressed several exposures including repetitive hand use, hand/wrist postures, hand force, and use of

Received from: Department of General Medical Sciences, Washington University School of Medicine, St. Louis, MO, USA (AMD, JS, BG, BAE); and Department of Mathematics and Statistics, Utah State University, Logan, UT, USA (JS). This study was funded by the Centers for Disease Control and Prevention, National Institute for Occupational Safety and Health grant no. R01 OH008017-01. Send correspondence to: Ann Marie Dale, Division of General Medical Sciences, Washington University School of Medicine, Campus Box 8005, 660 South Euclid Avenue St. Louis, MO 63110; email: <adale@dom.wustl.edu>.

Disclosures: The authors declare no conflict of interest.

vibrating tools. Surveys in other studies often include a limited number of questions about hand use within a larger questionnaire. For example, assessment of repetitive wrist movement was surveyed by Viikari-Juntura,¹² Pope,⁹ and Hansson.¹¹ Surveys exclusively related to hand and wrist use are limited.

There is a concern that self-reported data can lead to misclassification from biased reporting of exposures. Presence of symptoms has been suggested as one potential source of bias and would lead to spurious exposure-response relationships. Results from past studies have been mixed in terms of detecting overestimated exposures among symptomatic workers, as well as other biases resulting from gender and type of job, but there has been no bias found based on the age of workers.^{11,12,15-18} It may be that there are other confounders affecting this exposure-response association that have not been explored. Psychosocial factors have been associated with WRMSDs, but these factors have not been examined for possible misclassification.¹⁹ Physical characteristics such as individual worker strength may modify the exposure response as stronger workers may under-report exposures compared to weaker workers in the same job. Inclusion of questions about the presence of symptoms or other factors may be important to evaluate the validity of the exposure-response results from a survey for a specific population.

Self-reported surveys remain a necessary element of large scale epidemiological studies so it is advisable to explore the quality of exposure data such instruments provide.⁵ The purpose of this study was to measure agreement between workers' self-reported estimates and observed ratings of daily hand and wrist use in a group of workers from a variety of industries, and to examine predictors of over- and under-estimations of self-reported exposures. In addition, the presence of hand/wrist symptoms was analyzed as a predictor of agreement, after controlling for personal characteristics and psychosocial factors.

METHODS

Design and Study Sample

Data are from an ongoing prospective study examining personal and work factors in a group of newly hired workers ($n = 1108$). Invited subjects were hired into both low and high hand-intensive work from three main job categories: construction, service, and clerical/technical. Study subjects were hired by eight employers and three trade unions. The study design and population has been described by Armstrong and colleagues.²⁰ As part of the activities of the larger study, subjects completed a physical examination with grip strength testing using a dynamometer at baseline as well as a self-administered questionnaire approximately

six months after enrollment. Subjects were also visited at their worksite by a member of the research team. The subjects for the current study ($n = 341$) consisted of a subset of individuals from the larger study that had completed a self-administered questionnaire and a worksite visit for the same job. This study was approved by the Washington University School of Medicine Institutional Review Board; all participants provided written informed consent and received compensation.

Upper Extremity Physical Exposures

Physical exposures were evaluated through eight activities involving hand and wrist use during work. Several of these items were used in studies by Nordstrom and colleagues,^{13,21,22} including: lift/carry or push/pull > 0.91 kilograms ("lift"), work with hand-held or hand-operated vibrating power tools or equipment ("vibrate"), work on an assembly line ("assembly"), bend/twist hands/wrists ("bend"), use hand in a finger pinch grip ("pinch"), twist/rotate or screwing motion of forearm ("rotate"), and use tip of finger/thumb to press/push ("digit press"). An eighth item selected from past research relates to the physical exposure of using the hand in a forceful grip ("grip").²³ The response scale was a seven-point non-equidistant ordinal scale based on duration of daily work time spent performing the work activity, modified from the scale in Nordstrom's study.¹³ The scale categories were: none (1), less than five minutes (2), five to 30 minutes (3), more than 30 minutes but less than one hour (4), one to two hours (5), more than two hours but less than four hours (6), and four or more hours (7) per day. There was no information on reliability of the measure, but Nordstrom¹³ assessed the validity of seven items and showed poor to good results with Cohen kappa values of -0.02 to 0.79 . The same items were assessed in both self-reported and observer-based questionnaires.

Self-reported Questionnaires

The questionnaires asked for demographic information, work history, medical history, presence of upper extremity symptoms, completion of a hand diagram, and psychosocial and functional status. For the eight work activity questions, subjects were instructed to indicate how much time on average was spent each day performing the task or exposure. The presence of upper extremity symptoms was assessed by questions, developed for the Standardized Nordic Musculoskeletal Questionnaire, that asked about symptoms occurring more than three times or lasting more than one week in the neck/shoulder, elbow/forearm, or hand/wrist.²⁴ These symptoms were assessed for three different time frames: symptoms in the past six to 12 months, symptoms in the past 30 days, and current symptoms. Work-related psychosocial factors were assessed using four

summary scales from the Job Content Questionnaire (JCQ): job decision latitude, co-worker support, supervisor support, and job insecurity.²⁵ Many of the questionnaire items were drawn from past research on upper extremity disorders and had previously shown good to excellent test-retest reliability.^{25,26}

Observer-rated Questionnaires

Subjects received a worksite visit by a research team member at least six months post-enrollment. This observer was an occupational therapist trained in ergonomics. The observer was blinded to subject's self-reported work activity ratings and presence of symptoms. Subjects remained with the same employer and job title during completion of both the self-reported and observer questionnaires. The one-hour worksite visit included brief interviews with subjects and supervisors to gather task information and approximately 20 minutes of videotaped recordings of the worker performing tasks. During the interviews, workers were asked to list the work tasks performed during their typical job and to estimate the proportion of daily time per task. The observer asked for additional information about each task including a description of the steps of the task, the number of items or work cycles completed per task or day, and the type and weight of equipment and materials used. Then workers were asked to return to performing their typical work tasks while the observer took videotaped recordings of their work activities. The observer recorded several cycles for tasks of shorter duration and one full cycle for tasks of longer duration. Workers were asked to demonstrate those tasks that were not observed during the brief visit but are typical tasks performed by the worker in order to capture a sample of the task on videotape. Workers were asked whether the observed tasks were representative of their typical day and if not, to describe the differences. Following the worksite visit, two or three team members who were experienced in assessing work exposures, including the observer, evaluated the information gathered at the worksite visit. The team determined the daily time per task based on the recorded time per task on the videotape, worker estimated time from the interview, and from knowledge gained through prior worksite assessments of the same job. Using a consensus method developed by Latko and colleagues,¹⁰ team members jointly assigned ratings for the eight work activity questions using the videotape and interview data. Latko and colleagues showed good reliability ($r^2 = 0.88$) using this consensus method.¹⁰ Prior to proceeding with the current study, we evaluated the inter-rater reliability using this consensus method by three of our team members. We compared independent ratings of a separate sample of 26 subjects and found an overall intra-class correlation coefficient of 0.88.

TABLE 1 Personal and Psychosocial Characteristics of 341 Workers

	n	(%)
Gender		
Male	181	(53)
Female	160	(47)
Race		
Caucasian	188	(55)
Others ^a	151	(44)
Missing	2	(1)
Job category		
Construction ^b	108	(32)
Service ^c	95	(28)
Clerical/technical ^d	138	(40)
Upper extremity symptoms		
Present ^e	150	(44)
	Mean	SD
Age, in years	34	(11)
Psychosocial factors ^f		
Job decision latitude (n = 320)	26	(4)
Co-worker support (n = 337)	12	(2)
Supervisor support (n = 324)	13	(3)
Job insecurity (n = 336)	6	(1)
Right hand grip, in kilograms	41	(12)

^aIncludes African Americans, Asians, Native Americans, and others.

^bIncludes carpenters, floor layers, and sheet metal workers.

^cIncludes housekeepers and food service workers.

^dIncludes clerical, laboratory, computer and workers, hospital technicians, and other types of workers.

^eIncludes symptoms of the neck/shoulder, elbow/forearm, or wrist/hand experienced in the past six to 12 months.

^fInformation missing for 43 workers (13%).

Statistical Analyses

Self-reported and observed responses to the work activity items were evaluated using a weighted kappa coefficient directly comparing the response values. Weighted kappa accounts for partial agreement of responses on an ordinal scale and corrects for chance agreement.^{27,28} Landis and Koch²⁹ categories were used to describe levels of agreement: < 0 = less than probability; 0 = poor; 0.01–0.20 = slight; 0.21–0.40 = fair; 0.41–0.60 = moderate; 0.61–0.80 = substantial; and 0.81–1 = almost perfect. We calculated the weighted kappa and the simulation (bootstrap-based) 95% confidence intervals (95%CI). We repeated the analysis testing for agreement using an intra-class correlation coefficient for each exposure.

Distribution of Agreement

To evaluate whether self-reported responses were systematically overestimated or underestimated with respect to observed responses, the trends were examined graphically. Responses that were within one point

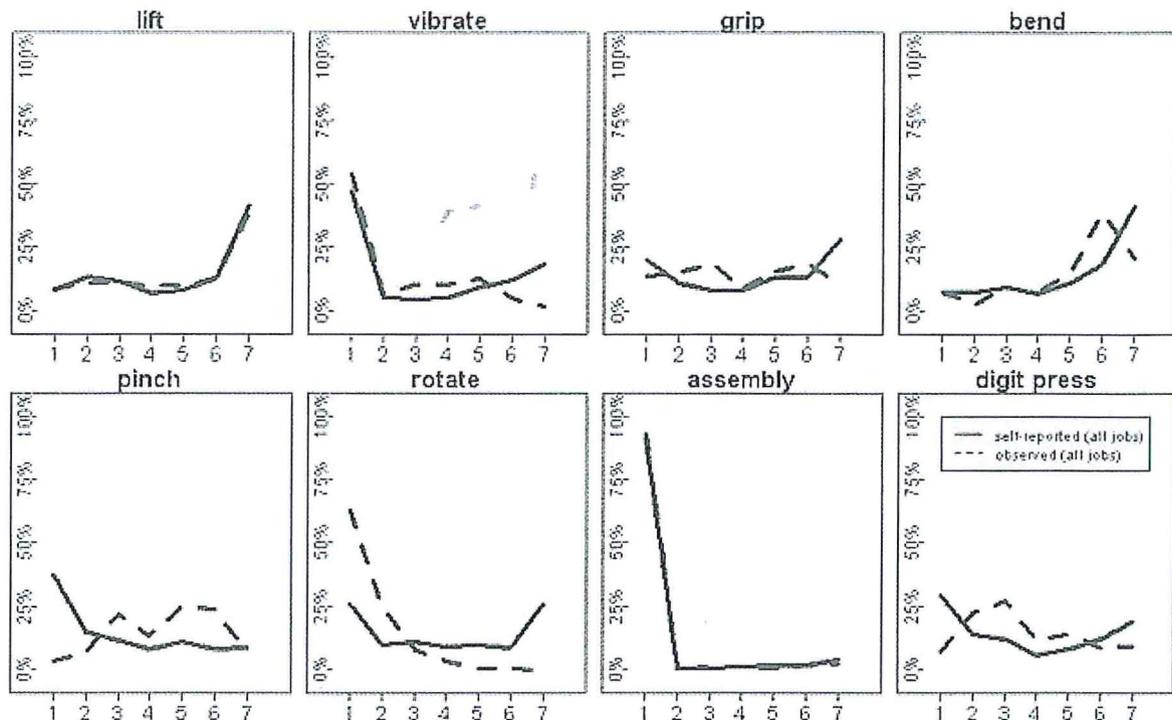


Figure 1—Distribution of self-reported (solid lines) and observed (dashed lines) exposures for daily duration of time. Numbers on the horizontal axis represent time categories for the average time performed on daily basis: none (1), less than five minutes (2), five to 30 minutes (3), more than 30 minutes but less than one hour (4), one to two hours (5), more than two hours but less than four hours (6), four or more hours (7).

on the seven point ordinal scale based on duration of daily work time spent performing the work activity were considered to be in agreement (“near agreement”). Overestimation was defined as a difference of more than one time category on the scale for self-reported compared to observed exposures. Underestimation occurred when self-reported exposures were more than one time category less than observed. Each group of exposure responses was stratified by job categories to observe potential differences by job type.

Predictors of Agreement

Logistic regression analyses were performed to examine potential predictors of agreement between self-reported and observed responses for each of the eight separate items. For this analysis, we used “near agreement” to describe agreement responses that were within one point on the seven-point ordinal scale. We tested the effects of several personal variables on agreement between observed and self-reported exposure: age, gender, race (categorized as Caucasian or other), presence of upper extremity symptoms, mean grip strength, job category, and the psychosocial scales of job decision latitude, coworker support, supervisor support, and job insecurity. Presence of upper extremity symptoms was evaluated by three different definitions in separate analyses: (1) positive response to symptoms occurring

more than three times or lasting more than one week in the neck/shoulder or elbow/forearm or hand/wrist in the past six to 12 months, (2) symptoms occurring in the past 30 days, and (3) current symptoms. Subjects were assigned nominal job categories: the construction group consisted of carpenters, floor layers, and sheet metal workers; the service group consisted of housekeepers and food service workers; and the clerical/technical group consisted of clerical workers, computer workers, laboratory workers, health technicians, and other types of workers. Psychosocial variables were summary scores following recommended calculations described for the JCQ.²⁵ Mean grip strength (in kilograms) was represented as the average of three trials for the right hand. We computed the odds ratios and 95% confidence intervals to show the likelihood of an association between these predictors and the outcome of agreement.

Homogeneity of Self-reported Exposures

To further evaluate whether the presence of symptoms influenced the respondents, we tested the homogeneity of the exposures with the sample stratified by symptom status. We calculated and compared the weighted kappas for the symptomatic and non-symptomatic cases. All analyses for this study were conducted using the statistical software package R.³⁰

Figure 2a

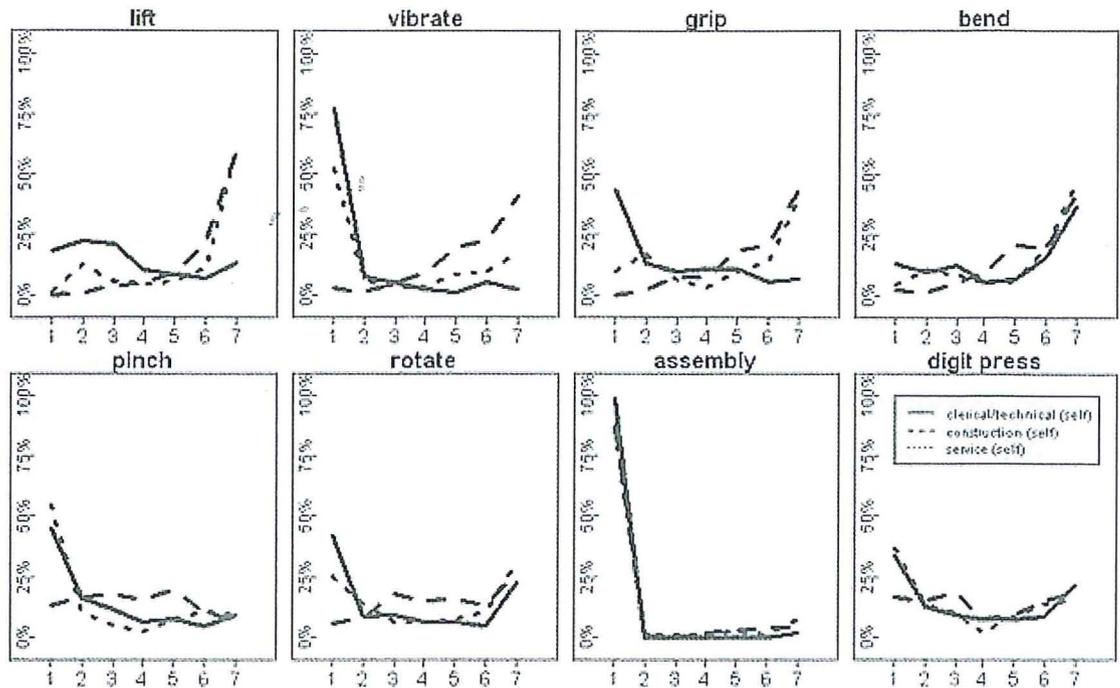
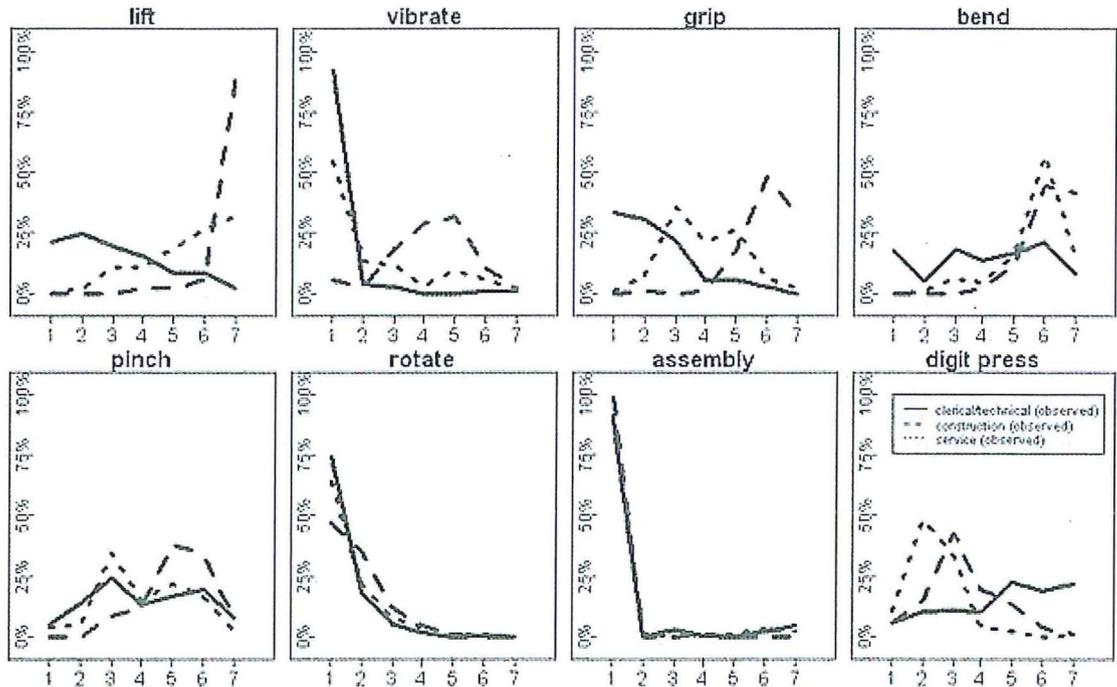


Figure 2b



Figures 2a and 2b—Distribution of self-reported exposures for three job types: clerical/technical (solid lines), construction (dashed lines), and service (dotted lines) worker groups. Figure 2a shows the results from the self-reported exposures; Figure 2b shows the results from the observed exposures. Numbers on the horizontal axis represent the same information as in Figure 1.

RESULTS

Three hundred and forty-one subjects completed the self-reported questionnaire and had observed estimates based on information collected during worksite visits. Subjects had a mean age of 34 years; 53% were male and 55% Caucasian. Table 1 shows the demo-

graphic and psychosocial characteristics of this sample. The time interval from collection of self-reported questionnaires and observed data from the worksite visit had a median difference of slightly less than seven weeks; there was no relationship between a greater time interval and differences in self-reported and observed responses. There were 44 cases with at least

TABLE 2 Percent Agreement and Weighted Kappa Coefficients for Comparison of Self-Reported and Observer-Rated Exposures for 341 Subjects

Physical Exposures	Percent Near Agreement	Weighted Kappa Coefficient (K_w)	Category of Agreement ^a	95%CI of K_w
Lift	68%	0.67	Substantial	0.60–0.73
Vibrate	67%	0.61	Substantial	0.54–0.68
Grip	58%	0.58	Moderate	0.51–0.64
Bend	59%	0.23	Fair	0.11–0.34
Pinch	33%	0.16	Slight	0.08–0.24
Rotate	43%	0.04	Slight	0.003–0.08
Assembly	87%	-0.01	Less than probability	-0.08–0.11
Digit press	36%	-0.07	Less than probability	-0.18–0.04

^aUsing definitions of categories by Landis and Koch.

one missing psychosocial factor or value for race, although chi-square and t-test analyses showed no meaningfully different results between cases with missing data and those with complete data (44 versus 297 cases) for age, gender, presence of symptoms, and grip strength.

The distribution of the self-reported and observed values for each item is shown in Figure 1. The graphs show a wide range for duration of time for most items. There was minimal time reported by all subjects for the assembly task, and the greatest daily time estimates were reported for lifting and hand/wrist bending exposures. Self-reported and observed responses showed similar distributions except for finger pinch, forearm rotation, and digit press where considerable differences can be seen.

Stratification of responses by job type shown in Figures 2a and 2b indicate that the construction group had the highest time estimates and the clerical/technical group reported the lowest estimates across most items. In particular, the construction group had higher time estimates for activities involving hand force such as lifting, use of hand-held vibrating tools, and forceful hand grip. Self-reported exposures were more similar across all job types whereas observed exposures showed greater differences.

Agreement of self-reported and observed values was examined with near agreement defined as a difference of one category or less between self-reported and observed values on the ordinal scale. Table 2 shows there was near agreement of 33% to 87% with a lower percent agreement for the physical exposures of finger pinch, forearm rotation, and digit press. Weighted kappa statistics and intra-class correlation coefficients (ICCs) were used to further assess the agreement between self-reported and observed responses. Using the Landis and Koch²⁹ scale to interpret results, weighted kappas showed moderate to substantial agreement (0.41–0.80) for lifting, hand-held vibrating tools, and forceful grip, and fair or less agreement for the other exposures (Table 2). The assembly task showed high agreement between self-reported and observed (87%); the narrow distribution of all responses (Figure

1) would require almost identical answers in order to obtain a higher kappa coefficient ($kappa = -0.01$). Calculations of ICCs produced nearly identical results as the weighted kappa statistics.

Figure 3 illustrates the overestimation or underestimation of self-reported values relative to observed values for the three job types separately. The graphs show that forearm rotation was systematically overestimated by workers across all job types; finger pinch was more frequently underestimated. In addition, construction workers tended to overestimate vibration and digit press exposures, service workers overestimated forceful grip and digit press, and clerical workers underestimated pressing or pushing with the fingers and the thumb and overestimated wrist bend. There were no obvious trends or patterns of overestimation or underestimation across all exposures by job type.

Logistic regression analyses examined possible associations contributing to the misclassification of agreement between self-reported and observed responses. The potential confounders included several personal and psychosocial factors, job type, and the presence of symptoms. Results for the 297 subsets with complete observations shown in Table 3 found few meaningful associations although job type demonstrated a significant effect for six out of the eight exposures. The service group had the greatest number of items with differing agreement. There was less agreement shown for four items and greater agreement shown for two when compared to the clerical/technical group.

To further examine the possibility that the presence of symptoms would lead to biased self-reported exposure estimates, we looked at three separate symptom definitions in different models. The prevalence of symptoms in the past 30 days (29%) and current symptoms (14%) were much lower than for symptoms in the past six to 12 months (44%). In all models, symptoms did not predict greater or lesser agreement between self-reported and observed exposures. The results for presence of symptoms experienced in the past six to 12 months are shown in Table 3.

We further tested the homogeneity of the kappa statistics that originated from the symptomatic cases and

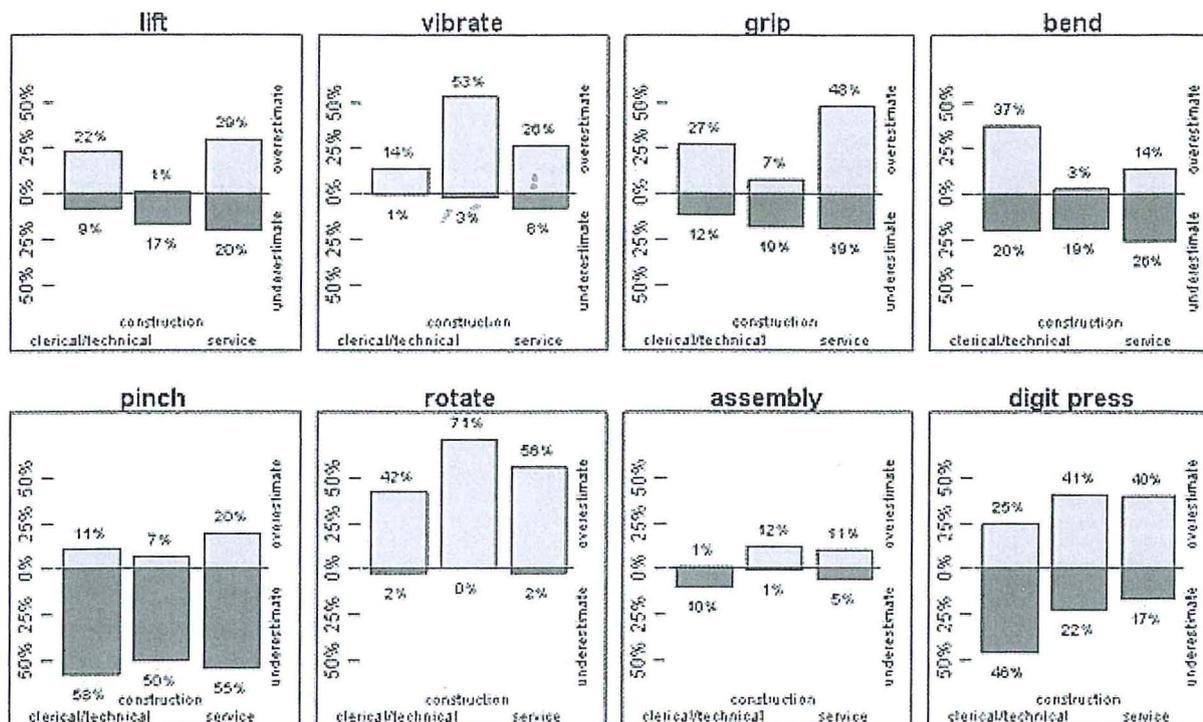


Figure 3—Prevalence of overestimated and underestimated values for self-reported and observed differences for each physical exposure by three job types. Percentage of overestimated values are shown above the horizontal line and underestimated values below.

the non-symptomatic cases of the population. We calculated the weighted kappa values separately for the symptomatic ($n = 150$) and non-symptomatic ($n = 191$) cases (data not shown). Using the recommended 84% confidence interval of two samples³¹ in which the ratio of the square root of the sample size for all eight exposures is close to unity, we found that the 84% confidence intervals for the two groups overlapped for all exposures, indicating homogenous kappa statistics (levels of agreement) for the symptomatic and the non-symptomatic populations.

DISCUSSION

This study explored the agreement of physical work exposures from a self-reported survey focused on hand and wrist activities and showed substantial agreement between self-reported values and observed ratings for some of the physical exposures. Examination of differential misclassification showed no systematic effect for presence of symptoms, gender, race, or psychosocial variables. Type of job was associated with differences in agreement for many exposures, but there was no trend toward overestimation or underestimation of exposures by a single job type. The construction group had the highest exposures on average. The service group, with moderate level work exposure, had the greatest tendency toward low agreement and showed the largest percentage of both overestimated and underestimated

exposures compared to the other two job types. Overall, the clerical/technical work group had the lowest exposures and showed the best agreement between self-reported and observed values. Some exposures were systematically underestimated or overestimated by all job types indicating there was a difference in the perception of exposure between the workers who completed the self-reported surveys and the observers.

There are few surveys that assess work exposures of the hands and wrists. This study used survey items from the 1988 Occupational Health Supplement to the National Health Interview Survey¹⁴ that were used in a validation study by Nordstrom and colleagues.¹³ Our study extended this prior work using a larger and more diverse worker population, including a greater proportion of blue collar and service workers. Both studies used work site observations (approximately 1 hour in length), and our study included video samples rated by two or three observers. Despite these differences, both studies showed similar results. Our study had slightly higher agreement for tasks involving force and vibration and lower agreement for activities involving precise hand movements and hand/wrist postures. This suggests that workers are able to accurately report time spent in general work tasks such as lifting and using vibrating tools, but they have difficulty recognizing exposures within tasks such as intermittent pinching and wrist bending. In addition, we found that the greater the variability of activities and exposures within

TABLE 3 Logistic Regression Results Assessing Agreement of Self-reported and Observed Ratings for Separate Outcome Exposures with Personal and Psychosocial Independent Predictor Variables in Each Model (n = 297)

Independent Variables	Lift	Vibrate	Grip	Bend	Pinch	Rotate	Assembly	Digit Press
	OR 95%CI	OR 95%CI	OR 95%CI	OR 95%CI	OR 95%CI	OR 95%CI	OR 95%CI	OR 95%CI
Age (per year)	1.00 (0.97–1.02)	0.99 (0.96–1.02)	1.00 (0.98–1.03)	0.99 (0.96–1.01)	1.01 (0.99–1.04)	1.03* (1.01–1.06)*	1.00 (0.97–1.03)	1.01 (0.98–1.03)
Race (caucasian ^a)	1.17 (0.54–2.60)	1.33 (0.53–3.45)	0.79 (0.39–1.63)	0.84 (0.41–1.74)	0.84 (0.39–1.76)	1.07 (0.52–2.20)	0.94 (0.35–2.67)	0.49 (0.21–1.08)
Right mean grip (kg)	0.99 (0.96–1.02)	1.00 (0.97–1.03)	1.02 (1.00–1.05)	1.01 (0.98–1.04)	1.04* (1.00–1.07)*	1.03* (1.00–1.07)*	1.01 (0.97–1.06)	1.01 (0.98–1.04)
Gender (male ^a)	0.38* (0.15–0.91)*	0.86 (0.32–2.20)	1.18 (0.52–2.70)	2.06 (0.90–4.79)	1.11 (0.48–2.66)	1.95 (0.86–4.52)	1.23 (0.38–3.71)	0.95 (0.42–2.19)
Job decision latitude	0.94 (0.86–1.02)	0.92 (0.84–1.00)	0.96 (0.89–1.04)	1.01 (0.94–1.10)	0.94 (0.86–1.01)	0.96 (0.89–1.04)	1.08 (0.97–1.20)	1.03 (0.96–1.12)
Coworker support	1.15 (0.98–1.37)	1.08 (0.90–1.29)	1.08 (0.92–1.26)	1.09 (0.93–1.28)	1.04 (0.89–1.23)	0.97 (0.82–1.13)	1.05 (0.85–1.29)	1.02 (0.86–1.19)
Supervisor support	0.87* (0.76–0.99)*	1.04 (0.91–1.19)	0.99 (0.88–1.12)	0.90 (0.79–1.02)	0.95 (0.84–1.08)	1.09 (0.96–1.23)	0.94 (0.79–1.10)	1.05 (0.93–1.19)
Job Insecurity	1.17 (0.87–1.62)	0.81 (0.60–1.08)	1.08 (0.82–1.44)	0.80 (0.60–1.05)	0.90 (0.67–1.17)	1.09 (0.83–1.43)	1.32 (0.91–2.00)	1.01 (0.77–1.33)
Construction group (clerical/technical ^a)	1.17 (0.48–2.85)	0.10* (0.04–0.26)*	1.49 (0.66–3.38)	6.51* (2.88–15.17)*	1.44 (0.65–3.23)	0.35 (0.16–0.79)	0.71 (0.22–2.16)	1.29 (0.59–2.86)
Service group (clerical/technical ^a)	0.38* (0.17–0.82)*	0.22* (0.08–0.52)*	0.40* (0.19–0.83)*	2.73* (1.30–5.87)*	0.81 (0.36–1.79)	0.48* (0.23–1.00)*	0.99 (0.35–2.66)	4.12* (1.85–9.73)*
Presence of symptoms ^b	1.11 (0.65–1.92)	0.77 (0.44–1.34)	0.68 (0.41–1.14)	1.39 (0.83–2.34)	0.89 (0.53–1.48)	0.78 (0.47–1.30)	0.95 (0.47–1.93)	0.81 (0.48–1.34)

*Statistically significant result.

^aReference group.

^bDefined as symptoms in the past six to 12 months; reference: no symptoms.

a single job, the lower the agreement. Past studies, such as that by Stock et al, have shown wide variations in validity with kappa values from -0.07 to 0.81 .⁶ Agreement between self-reported and observed upper extremity exposures seen in this study was somewhat higher than those discussed by Stock et al. Other studies have reported moderate agreement for duration of handling loads of specific weights, fair agreement for use of hand-held vibrating hand tools, and fair to poor agreement for tasks involving hand use.^{9,12,13,17,32}

Several studies have shown an association between physical work exposures and upper extremity case definitions for WRMSDs.^{20,33,34} Determining a dose-response relationship requires accurate quantification of exposures; the presence of exposure misclassification may obscure true relationships or create spurious associations. In the current study, we examined several possible sources of differential exposure misclassification, including personal and psychosocial factors, and the presence of symptoms, and found no relationship with self-reported exposures. The lack of association between symptoms and exposure misclassification is important; if present, such an association could result in a spurious association between physical exposures and symptoms, particularly in a cross-sectional study.

We found that there were differences in agreement between self-reported and observed exposures by work group. This may be related to intermittently performed tasks that are not recognized by workers. Observers use more quantifiable criteria for exposures whereas workers' perceptions are formed by personal knowledge, experience, and, possibly, work-related terminology. Workers and observers have different knowledge about exposures and the tasks of a given job; asking workers to assess time spent in exposures that vary considerably during tasks may be unreasonable.

Most studies that have explored the validity of self-reported exposures have shown adequate classification for levels of exposures when workers report general body postures and work tasks such as standing or walking.^{17,35} Finer dimensions of exposures including joint posture, frequency of movement, intensity, or specific loads have poorer agreement of self-reported with observed or directly measured exposures. This presents a difficult problem when it comes to trying to quantify exposures of the upper extremity, all of which involve precision or posture or generally low loads compared to the exposures on the trunk or legs. This may be one reason researchers to date do not have a set of well-validated upper extremity questions avail-

able for use. Some researchers report customizing upper extremity questions to the tasks within the industry, but little is reported about the nature of the questions and testing prior to use.³⁶ Since self-reported surveys remain the most feasible means of collecting exposure information in large population epidemiologic studies, there must be systematic exploration into the psychometric formulation of questions and response scales, testing with specific industries, and reporting results in the literature.⁶

The primary strength of this study was the extension of a previously used self-reported hand and wrist survey in a large population of workers across several industries.¹³ An important element of this study was our examination of exposure misclassification related to personal and psychosocial factors and job types. We found no effect of symptoms on exposure reporting, even when using different symptom definitions and controlling for other potential associations. The novel exploration for this manuscript was the examination of responses for overestimations and underestimations by self-report. These analyses shed some light on the need for future studies incorporating type of work into item formulation and interpretation.

There were several limitations to this study. The difference in time between completion of the survey and the worksite visit may have affected results in ways that were not detected by our analysis. In responding to the self-reported questions, some workers may consider only recently performed tasks when describing their "usual exposures," but others may reflect upon tasks over a much longer period of time. Finally, the observed method of rating exposures may not have accurately captured all exposures, particularly for the more varied jobs. The duration of time spent in tasks for the observed exposures was based on limited quantitative information from the videotaped samples, on information from worker interviews, and on information gathered from previous worksite visits for similar jobs. This method for determining task time likely provides an appropriate estimate, but is undoubtedly subject to some measurement error. All of these limitations would likely lead to lower agreement between self-reported and observed exposure estimates.

CONCLUSION

Self-reported estimates of time spent on tasks involving hand and wrist use is a useful tool for population based studies. Self-reported estimates of physical exposures may be less accurate for jobs involving variable tasks and intermittent exposures: such jobs are challenging to study by any method. Presence of musculoskeletal symptoms did not cause greater misclassification of self-reported tasks and exposures in our study population. Job category may cause overestimates or underestimates of time spent in work activities, so this potential

bias should be evaluated when comparing exposures across different job groups.

References

1. US Department of Labor. US Bureau of Labor Statistics. Occupational injuries and illnesses: Counts, rates and characteristics. Report 1014. Washington, DC: US Dept of Labor; 2007. Chart 37. Number and incidence rates of injuries and illnesses due to musculoskeletal disorders, by selected occupations, private industry, 2006. Available from: http://www.bls.gov/iif/oshwc/osh/os/osh06_37.pdf.
2. Winkel J, Mathiassen S. Assessment of physical work in epidemiology studies: Concepts, issues and operational considerations. *Ergon*. 1994 Jun;37(6):979-88.
3. Spielholz P, Silverstein B, Morgan M, Checkoway H, Kaufman J. Comparison of self-report, video observation and direct measurement methods for upper extremity musculoskeletal disorder physical risk factors. *Ergon*. 2001 May 15;44(6):588-613.
4. Van der Beek A, Frings-Dresen MH. Assessment of mechanical exposure in ergonomic epidemiology. *Occup Environ Med*. 1998 May;55(5):291-9.
5. Punnett L, Wegman DH. Work-related musculoskeletal disorders: The epidemiologic evidence and the debate. *J Electromyogr Kinesiol*. 2004 Feb;14(1):13-23.
6. Stock SR, Fernandes R, Delisle A, Vezina N. Reproducibility and validity of workers' self-reports of physical work demands. *Scand J Work Environ Health*. 2005 Dec;31(6):409-37.
7. Descatha A, Roquelaure Y, Caroly S, Evanoff B, Cyr D, Mariel J, et al. Self-administered questionnaire and direct observation by checklist: Comparing two methods for physical exposure surveillance in a highly repetitive tasks plant. *Appl Ergon*. 2009 Mar;40(2):194-198.
8. Somville P, van Nieuwenhuysse A, Seidel L, Masschelein R, Moens G, Mairiaux P. Validation of a self-administered questionnaire for assessing exposure to back pain mechanical risk factors. *Int Arch Occup Environ Health*. 2006 Jun;79(6):499-508.
9. Pope DP, Silman AJ, Cherry NM, Pritchard C, Macfarlane GJ. Validity of a self-completed questionnaire measuring the physical demands of work. *Scand J Work Environ Health*. 1998 Oct;24(5):376-85.
10. Latko WA, Armstrong TJ, Foulke JA, Herrin GD, Rabourn RA, Ulin SS. Development and evaluation of an observational method for assessing repetition in hand tasks. *Am Ind Hyg Assoc J*. 1997 Apr;58(4):278-85.
11. Hansson GA, Balogh I, Bystrom JU, Ohlsson K, Nordander C, Asterland P, et al. Questionnaire versus direct technical measurements in assessing postures and movements of the head, upper back, arms and hands. *Scand J Work Environ Health*. 2001 Feb;27(1):30-40.
12. Viikari-Juntura E, Rauas S, Martikainen R, Kuosma E, Riihimaki H, Takala EP, et al. Validity of self-reported physical work load in epidemiologic studies on musculoskeletal disorders. *Scand J Work Environ Health*. 1996 Aug;22(4):251-9.
13. Nordstrom DL, Vierkant RA, Layde PM, Smith MJ. Comparison of self-reported and expert-observed physical activities at work in a general population. *Am J Ind Med*. 1998 Jul;34(1):29-35.
14. US Department of Health and Human Services. National Center for Health Statistics. National Health Interview Survey, 2006 [Computer file]. ICPSR20681-v2. Ann Arbor, MI: Inter-university Consortium for Political and Social Research [distributor], 2009-01-07. doi:10.3886/ICPSR20681
15. Wiktorin C, Vingard E, Mortimer M, Pernold G, Wigaeus-Hjelm E, Kilbom A, et al. Interview versus questionnaire for assessing physical loads in the population-based MUSIC-Norrtaalje Study. *Am J Ind Med*. 1999 May;35(5):441-55.
16. Wiktorin C, Selin K, Ekenvall L, Kilbom A, Alfredsson L. Evaluation of perceived and self-reported manual forces exerted in occupational materials handling. *Appl Ergon*. 1996 Aug;27(4):231-9.
17. Leijon O, Wiktorin C, Harenstam A, Karlqvist L. Validity of a self-administered questionnaire for assessing physical work loads in a general population. *J Occup Environ Med*. 2002 Aug;44(8):724-35.

18. Torgen M, Winkel J, Alfredsson L, Kilbom A. Evaluation of questionnaire-based information on previous physical work loads. Stockholm MUSIC 1 Study Group. Musculoskeletal Intervention Center. *Scand J Work Environ Health*. 1999 Jun;25(3):246-54.
19. Toomingas A, Alfredsson L, Kilbom A. Possible bias from rating behavior when subjects rate both exposure and outcome. *Scand J Work Environ Health*. 1997 Oct;23(5):370-7.
20. Armstrong T, Dale AM, Franzblau A, & Evanoff B. (2008). Risk factors for carpal tunnel syndrome and median neuropathy in a working population. *J Occup Environ Med*. 2008 Dec;50(12):1355-64.
21. Nordstrom DL, Vierkant RA, DeStefano F, Layde PM. Risk factors for carpal tunnel syndrome in a general population. *Occup Environ Med*. 1997 Oct;54(10):734-40.
22. Nordstrom DL. A population-based, case control study of carpal tunnel syndrome [dissertation]. Madison, WI: University of Wisconsin-Madison; 1996. 304 p.
23. Silverstein BA, Stetson DS, Keyserling WM, Fine LJ. Work-related musculoskeletal disorders: Comparison of data sources for surveillance. *Am J Ind Med*. 1997 May;31(5):600-8.
24. Kuorinka I, Jonsson B, Kilbom A, Vinterberg H, Biering-Sorensen F, Andersson G, et al. Standardised Nordic questionnaires for the analysis of musculoskeletal symptoms. *Appl Ergon*. 1987; 18:233-7.
25. Karasek RA. Job Content Questionnaire and user's guide. Lowell: University of Massachusetts Lowell, Department of Work Environment; 1985. Available at: <http://www.uml.edu/dept/WE/research/jcq/jcq.htm>.
26. Franzblau A, Salerno DF, Armstrong TJ, Werner RA. Test-retest reliability of an upper-extremity discomfort questionnaire in an industrial population. *Scand J Work Environ Health*. 1997 Aug; 23(4):299-307.
27. Streiner D, Norman G. Health measurement scales: A practical guide to their development and use. 3rd ed. New York: Oxford University Press; 2003.
28. Cohen J. Weighted kappa: Nominal scale agreement with provision for scaled disagreement or partial credit. *Psychol Bull*. 1986(70):213-20.
29. Landis J, Koch G. The measurement of observer agreement for categorical data. *Biometrics*. 1977;33:159-74.
30. R Development Core Team. R: A language and environment for statistical computing. Vienna, Austria: R Foundation for Statistical Computing; 2007. Available from: <http://www.R-project.org>.
31. Payton ME, Greenstone MH, Schenker N. Overlapping confidence intervals or standard error intervals: What do they mean in terms of statistical significance? *J Insect Sci*. 2003;3:34.
32. Wiktorin C, Karlqvist L, Winkel J. Validity of self-reported exposures to work postures and manual materials handling. Stockholm MUSIC 1 Study Group. *Scand J Work Environ Health*. 1993 Jun;19(3):208-14.
33. Rempel D, Evanoff B, Amadio PC, de Krom M, Franklin G, Franzblau A, et al. Consensus criteria for the classification of carpal tunnel syndrome in epidemiologic studies. *Am J Public Health*. 1998 Oct;88(10):1447-51.
34. Gardner BT, Dale AM, Vandillen L, Franzblau A, Evanoff BA. Predictors of upper extremity symptoms and functional impairment among workers employed for 6 months in a new job. *Am J Ind Med*. 2008 Dec;51(12):932-40.
35. Halpern M, Hiebert R, Nordin M, Goldsheyder D, Crane M. The test-retest reliability of a new occupational risk factor questionnaire for outcome studies of low back pain. *Appl Ergon*. 2001 Feb;32(1):39-46.
36. Fallentin N, Juul-Kristensen B, Mikkelsen S, Andersen JH, Bonde JP, Frost P, et al. Physical exposure assessment in monotonous repetitive work-the PRIM study. *Scand J Work Environ Health*. 2001 Feb;27(1):21-29.