Applied Occupational and Environmental Hygiene

ISSN: 1047-322X (Print) 1521-0898 (Online) Journal homepage: https://www.tandfonline.com/loi/uaoh20

Test-Retest Reliability of a Self-Administered Musculoskeletal Symptoms and Job Factors Questionnaire Used in Ergonomics Research

John C. Rosecrance , Kelly J. Ketchen , Linda A. Merlino , Dan C. Anton & Tom M. Cook

To cite this article: John C. Rosecrance , Kelly J. Ketchen , Linda A. Merlino , Dan C. Anton & Tom M. Cook (2002) Test-Retest Reliability of a Self-Administered Musculoskeletal Symptoms and Job Factors Questionnaire Used in Ergonomics Research, Applied Occupational and Environmental Hygiene, 17:9, 613-621, DOI: 10.1080/10473220290095934

To link to this article: https://doi.org/10.1080/10473220290095934

Published online: 30 Nov 2010.

Applied Occupational and Environmental Hygiene
Volume 17(9): 613–621, 2002

1047-322X/02 $12.00 + .00
DOI: 10.1080/1047322029009593 4

Test-Retest Reliability of a Self-Administered Musculoskeletal Symptoms and Job Factors Questionnaire Used in Ergonomics Research

John C. Rosecrance, Kelly J. Ketchen, Linda A. Merlino, Dan C. Anton, and Tom M. Cook
College of Public Health, University of Iowa, Iowa City, Iowa

The purpose of this study was to investigate the test-retest reliability of questionnaire items related to musculoskeletal symptoms and the reliability of specific job factors. The type of questionnaire items described in the present study have been used by several investigators to assess symptoms of musculoskeletal disorders and problematic job factors among workers from a variety of occupations. Employees at a plastics molding facility were asked to complete an initial symptom and jobs factors questionnaire and then complete an identical questionnaire either two or four weeks later. Of the 216 employees participating in the initial round, 99 (45.8%) agreed to participate in the retest portion of the study. The kappa coefficient was used to determine repeatability for categorical outcomes. The majority of the kappa coefficients for the 58 questionnaire items were above 0.50 but ranged between 0.13 and 1.00. The section of the questionnaire having the highest kappa coefficients was the section related to hand symptoms. Interval lengths of two and four weeks between the initial test and retest were found to be equally sufficient in terms of reliability. The results indicated that the symptom and job factors questionnaire is reliable for use in epidemiologic studies. Like all measurement instruments, the reliability of musculoskeletal questionnaires must be established before drawing conclusions from studies that employ the instrument.

Keywords Ergonomics, Musculoskeletal Disorders, Occupational Health, Questionnaire, Reliability, Survey

U.S. industrial workplaces reported 1.7 million injuries and illnesses resulting in time off from work in 1998.[1] Musculoskeletal disorders (MSDs) involving the low back, neck, and upper extremities are a significant contributor to workplace injuries and illnesses, accounting for 70 million physician office visits in the United States each year.[2] Conservative estimates of the economic burden resulting from MSDs, as measured by compensation costs, lost wages, and lost productivity are $50 billion annually.[2] Consistent with the National Occupational Research Agenda of the National Institute for Occupational Safety and Health (NIOSH), epidemiological studies of MSDs are a high priority in occupational health research.[3] An additional research priority involves intervention effectiveness research[3] where it is critical to employ reliable instruments determining prevalence and incidence of work-related MSDs, their signs and symptoms, and the job factors that may be associated with the MSDs.

A frequently used data source for assessing the incidence of work-related MSDs is the *Survey of Occupational Injuries and Illnesses* compiled by the U.S. Bureau of Labor Statistics.[1] However, the OSHA 200 log is a lagging indicator for determining the effectiveness of occupational health programs aimed at preventing MSDs and may not be sensitive to intervention strategies that have been recently implemented. An alternative and potentially useful instrument to assess prevalence and incidence of work-related MSD symptoms and other job factors are questionnaires and surveys administered at the work site. Symptom surveys and questionnaires have been utilized as an active surveillance tool to assess the short- and long-term changes of occupational health programs.[4,5] However, there have been surprisingly few studies addressing the reliability of such questionnaire instruments.[6] In a study establishing the reliability of one such questionnaire, Franzblau and colleagues[6] encouraged researchers to perform reliability studies of musculoskeletal survey instruments as a way of improving the quality of research in occupational health.[6] The purpose of the present study was to determine the test-retest reliability of questionnaire items concerning MSD symptoms and specific job factors. The self-administered questionnaire used in the present study has been used by other investigators[7–11] to assess MSDs and problematic job factors in a variety of occupational groups. The

present study assessed the test-retest reliability (temporal stability) of the questionnaire items and determined whether test-retest agreement was dependent on the duration of time (two or four weeks) between the initial and follow-up tests.

METHODS AND MATERIALS

Procedures

The company involved in this study had a work force of approximately 250 employees that performed tasks associated with the production of large molded plastic containers. The company's goal for this project was to assess the prevalence of MSD symptoms and problematic job factors among its work force. The investigator's goals were to assist the company in this effort and to investigate the repeatability of the questionnaire instrument in a subgroup of their employees. The company is divided into three divisions with each division consisting of two day shifts and one night shift. Employees from all divisions and all shifts were eligible and were invited to participate in this study. Thus, participation was based on self-selection from the entire employee population, rather than on a randomized sample of company employees. Employees were assured that company personnel would not have access to individual data and that only summary results, by company division, would be provided to the management personnel. The study participants signed written informed consent that had been approved by the University of Iowa Human Subjects Review Board.

All questionnaires were self-administered at the plastics company in either a conference room or in the break room depending upon room availability. The questionnaire was completed on company time either during regularly scheduled rest breaks or at the beginning or end of the employee's work shift. Instructions on completing the questionnaire were given to the employees before the questionnaire was administered. If an employee had a specific question regarding the questionnaire, the question was answered at that time. Although members of upper management were present when the questionnaire was administered, management never had access to any individual questionnaire results. The questionnaire required between 10 to 15 minutes to complete.

This study involved two rounds of data collection from the same group of employees, consisting of an initial test and then a retest. The identical questionnaire was used during the first and subsequent testing. For logistic reasons, the first round of data collection was conducted on three separate days, with each division of employees assigned a different testing day. The retest was conducted on a similar schedule and in the same physical location as the initial test. However, the length of time between test and retest varied for each of the three divisions as follows: Divisions 1 and 3 completed the retest 33 days and 27 days, respectively, after the initial test while Division 2 retested 14 days after their initial test. These test-retest intervals used in the present study were consistent with other studies assessing MSD questionnaire reliability.[6,12] The variability of test-retest interval across company divisions provided an opportunity to examine the different effects that interval length may have on test-retest reliability.

The first section of the questionnaire consisted of questions addressing age, height, weight, years in trade, and questions regarding work schedules. The second portion of the questionnaire addressed the period prevalence of work-related MSD symptoms within the last 12 months. This portion of the questionnaire was a modification of the Standardised Nordic Questionnaire[13] and consisted of questions referring to nine anatomical areas. A body diagram with the nine anatomical areas (neck, upper back, low back, shoulders, elbows, wrist/hands, hips, knees, and feet) was included to help the subjects answer "yes" or "no" to the following question: "During the last 12 months have you had a job-related ache, pain, discomfort, etc."—followed by a list and body diagram of the nine different anatomical areas. If the respondent marked "yes" that a work-related MSD symptom had occurred, they were instructed to answer "yes" or "no" to two additional questions. The questions were "During the last 12 months have you been prevented from doing your day's work due to this condition?" and "During the last 12 months have you seen a physician for this condition?"

The third section of the questionnaire addressed the employees' perceptions of 15 different job factors and their potential contribution to MSDs. The participants were given a descriptive list of 15 job factors and asked to indicate on a scale of 0–10 (0 = no problem, 10 = major problem) how much of a problem (if any) each factor contributed to job-related MSD symptoms. A response of 2 or greater was coded as "positive" for that factor, and a response of 0 or 1 was coded "negative" for that factor. Two examples of the job factor statements that were rated included: "Performing the same task over and over" and "Working in awkward or cramped positions."

In the fourth section of the questionnaire, participants were asked whether they strongly disagreed, disagreed, had no opinion, agreed, or strongly agreed with 10 psychosocial statements regarding work and supervisor support. These statements were a modification of five questions from the NIOSH Generic Job Stress Questionnaire[14] and five statements used by the Gallup organization[15] regarding work-related psychosocial variables. An example of a psychosocial statement from the NIOSH instrument that was used in the present study included "My supervisor goes out of his/her way to make my work life easier." For the purpose of analyses, the categories of "strongly disagree" and "disagree" were combined, as were the categories of "strongly agree" and "agree." There was also a third category of responses, labeled "no opinion."

The fifth section of the questionnaire consisted of a hand diagram and questions regarding the presence of specific hand symptoms. Employees were asked to indicate on a 0–10 scale (0 = none, 10 = severe) if they had had hand symptoms that included numbness, tingling, soreness, or tightness within the last year, and to shade the hand diagram where they had the symptoms. The hand symptom diagram used in the present study is

similar to other survey tools that evaluate hand symptoms for the classification of carpal tunnel syndrome.[16,17] For purposes of analyses, any symptom severity marked as greater than or equal to 2 on the scale was coded as "positive" for symptoms, and responses of 0 or 1 were coded as "negative" for symptoms. If the respondent indicated symptoms in the median nerve distribution of the hand (palm side of the first through third fingers and lateral side of palm) the response was considered a "positive" response for the location variable. All other areas of the hand diagram were considered a "negative" response for the symptom location variable. Nocturnal symptoms were also evaluated. A "yes" response was coded for nocturnal symptoms if the respondent indicated their hand symptoms were either "worse at night" or if "hand discomfort woke them from their sleep."

Statistical Analyses

The kappa coefficient,[18] which is commonly used to assess agreement between two or more observations, was used to analyze test-retest agreement. The kappa coefficient is calculated by subtracting expected agreement (EA) from the observed agreement (OA), and dividing by 1 minus the expected agreement [$\kappa = (\text{OA-EA})/(1\text{-EA})$].[19] As opposed to using the percent of agreement between the test and retest to assess reliability, the kappa coefficient makes adjustments for chance agreement and allows for comparison of results between studies.[19] T-tests and chi-square analyses were conducted to determine whether there were significant differences in subject characteristics between study participants completing both rounds of testing and those only completing the first round. Analysis of variance and chi-square analyses were conducted to test whether there were significant differences in demographic variables between company divisions completing the retest. Logistic regression was used to assess differences in test-retest agreement for retest interval. We also examined whether the length of time between test and retest was a significant factor in predicting participants' retest agreement. Using logistic regression, we compared responses from Division 2 (retested 2 weeks later) with those from Divisions 1 and 3 (retested 4 weeks later) combined. Logistic regression analyses were performed for each question on the survey, with the dependent variable of "agreement" equal to 1 if there were identical responses for each question between the two rounds of testing, and 0 if the responses were not identical. All analyses were performed using the SAS System for Windows '98, Version 8 (SAS Institute, Cary, NC).

RESULTS

Of the 237 employees eligible for participation in the first round of the questionnaire, 216 volunteered and completed the questionnaire (91% response rate). Approximately one-half (N = 99) of the employees agreed to participate in the retest portion of the study. Table II displays the demographic characteristics of all 216 participants (117 employees who completed the initial test only, and the 99 participants who completed both the initial and retest). There were no significant differences in mean age, body mass index (BMI), hours worked per week, weeks worked per year, or years worked at the company between employees participating in the initial test only and those completing both the initial and retest. Although the numbers were small, there was a significantly smaller percentage of women that took the retest (5%) than women that only participated in the initial test (11%) ($p = 0.01$). There were no significant differences between the three company divisions for retest participation. Four participants who completed both rounds were excluded from the final analyses because they had changed job titles between rounds of data collection. This resulted in a total of 95 employees included in the test-retest analyses.

Table III displays kappa coefficients for job-related MSD symptoms in each anatomical region. The symptom related questions produced kappa coefficients that ranged between 0.13 and 1.00. The questions pertaining to being "seen by a physician" for MSD symptoms had generally higher kappa coefficients than "experiencing job-related symptoms" during the previous 12 months. The higher kappa coefficients for items regarding physician visits may be related to the respondent's better recall of actual events such as traveling to the physician's office versus the recall of subjective symptoms in the preceding 12 months. The questions related to seeing a physician for neck, upper back, and low back symptoms had the highest ($\kappa = 0.73$–0.82) test-retest kappa coefficients. Kappa values were undefined for several of the questions due to an initial period prevalence (P_1) or retest period prevalence (P_2) of zero.

The kappa coefficient pertaining to the 15 job factors that the employees may have perceived as contributing to job-related aches and pains are listed in Table I. The kappa coefficients for the job factor questions ranged between 0.46 and 0.68 with the

TABLE I

Demographic characteristics of participants completing the initial test only and those completing both initial and follow-up test (retest)

	Initial testers only (N = 117)	Retesters (N = 99)	p-value
Gender			
Male	104 (88.9%)	94 (94.9%)	0.01
Female	13 (11.1%)	5 (5.1%)	
	Mean (SD)		
Age	33.7 (8.8)	33.4 (9.7)	0.76
Body mass index	25.8 (4.4)	25.8 (4.5)	0.93
Hours worked per week	43.3 (7.1)	44.1 (6.5)	0.39
Weeks worked per year	50.3 (1.9)	50.1 (1.5)	0.49
Years worked at company	4.2 (4.0)	4.0 (4.3)	0.63

TABLE II
Agreement between test and retest in response to questions regarding job-related musculoskeletal symptoms in each body region

Questionnaire item by body area	κ	95% CI[A]	P_1[B]	P_2[C]
"In the last 12 months, have you:"				
Neck				
Experienced job-related symptoms	0.55	0.31–0.80	0.13	0.15
Been prevented from doing a day's work	Undefined		0.01	0
Seen a physician	0.82	0.58–1.00	0.07	0.05
Upper back				
Experienced job-related symptoms	0.50	0.28–0.72	0.19	0.20
Been prevented from doing a day's work	Negative		0.02	0.01
Seen a physician	0.78	0.55–1.00	0.09	0.06
Low back				
Experienced job-related symptoms	0.61	0.45–0.77	0.44	0.43
Been prevented from doing a day's work	0.49	0.00–1.00	0.01	0.03
Seen a physician	0.73	0.53–0.94	0.14	0.14
Shoulders				
Experienced job-related symptoms	0.50	0.28–0.72	0.17	0.22
Been prevented from doing a day's work	Undefined		0	0.02
Seen a physician	0.64	0.35–0.93	0.07	0.08
Elbows				
Experienced job-related symptoms	0.71	0.55–0.88	0.24	0.24
Been prevented from doing a day's work	1.00	1.00–1.00	0.01	0.01
Seen a physician	0.48	0.05–0.91	0.05	0.03
Wrist/hand				
Experienced job-related symptoms	0.64	0.48–0.79	0.47	0.44
Been prevented from doing a day's work	0.49	0.00–1.00	0.01	0.03
Seen a physician	0.53	0.21–0.86	0.07	0.07
Hips/thighs				
Experienced job-related symptoms	0.13	0.00–0.45	0.05	0.06
Been prevented from doing a day's work	Undefined		0	0
Seen a physician	Negative		0.01	0.01
Knees				
Experienced job-related symptoms	0.71	0.53–0.90	0.18	0.18
Been prevented from doing a day's work	Undefined		0	0.01
Seen a physician	0.56	0.11–1.00	0.03	0.04
Feet				
Experienced job-related symptoms	0.21	0.00–0.44	0.18	0.22
Been prevented from doing a day's work	Undefined		0	0.01
Seen a physician	0.48	0.05–0.91	0.05	0.03

[A]95% confidence interval for $\kappa = \kappa \pm 1.96(SE_\kappa)$.
[B]Prevalence of "yes" responses on initial test.
[C]Prevalence of "yes" responses on retest.

majority above a kappa coefficient of 0.55. The highest kappa was recorded for "insufficient breaks or pauses during the workday" ($\kappa = 0.68$) and the lowest kappa coefficient for "continuing to work while injured or hurt" ($\kappa = 0.46$).

Kappa coefficients from questions related to psychosocial variables ranged between 0.25 and 0.55 and were generally lower than the kappa coefficients for the symptom or job factor items (Table IV). The highest kappa value ($\kappa = 0.55$) was for the statement regarding the worker's "influence over the quality of work." The lowest kappa values were for questions regarding management's concern for "employee health and safety" ($\kappa = 0.25$), and the receipt of "recognition and praise for good work" ($\kappa = 0.36$).

When queried about the presence of hand symptoms, symptom location, and if the symptoms were present at night (nocturnal signs), employees' responses demonstrated good

TABLE III
Agreement between test and retest in response to the question "Do any of the following contribute to your job-related pain and injury?"[A]

Job factor	κ	95% CI[B]	P_1[C]	P_2[D]
Performing the same task over and over	0.61	0.45–0.77	0.45	0.33
Working very fast for short periods	0.64	0.49–0.80	0.48	0.47
Having to handle or grasp small objects	0.49	0.31–0.66	0.41	0.34
Insufficient breaks or pauses during the workday	0.68	0.54–0.83	0.56	0.53
Working in awkward or cramped positions	0.49	0.32–0.67	0.54	0.54
Working in the same position for long periods	0.64	0.49–0.80	0.54	0.53
Bending or twisting your back in an awkward way	0.63	0.48–0.79	0.58	0.57
Working near or at your physical limits	0.56	0.39–0.72	0.46	0.52
Reaching or working over your head or away from your body	0.49	0.32–0.67	0.52	0.52
Hot, cold, humid, wet conditions	0.47	0.29–0.65	0.59	0.63
Continuing to work when injured or hurt	0.46	0.28–0.65	0.61	0.63
Carrying, lifting, or moving heavy materials or equipment	0.56	0.39–0.72	0.55	0.52
Work scheduling	0.60	0.44–0.76	0.51	0.54
Using tools	0.48	0.31–0.66	0.44	0.40
Training on how to do the job	0.54	0.37–0.72	0.36	0.28

[A]Response of ≥ 2 on scale of 0–10 indicates a "positive" response.
[B]95% confidence interval for $\kappa = \kappa \pm 1.96\ (SE_\kappa)$.
[C]Prevalence of "positive" responses on initial test.
[D]Prevalence of "positive" responses on retest.

reproducibility, with kappa coefficients ranging between 0.50 and 0.72 (Table V). The purpose of determining reliability of hand symptoms was to assist in the surveillance of carpal tunnel syndrome.

Division 2 completed the retest approximately two weeks after the first test, while Divisions 1 and 3 were retested approximately four weeks (27 and 33 days) after the first test. Thus, an additional comparison of the test-retest reliability was performed between Division 1 and the combined responses of Divisions 2 and 3 to investigate the effect of duration between the test and retest. Using a logistic regression model, there was only one questionnaire item whose agreement was statistically different between a two week and four week test-retest duration. Participants retested after two weeks had better agreement than those retested after four weeks when asked to respond to the statement "I have a great deal of influence over the amount of work I do."

DISCUSSION

Reliability is defined as the degree to which questionnaires, tests, or other instruments can be replicated and the kappa statistic is frequently used as an index to assess the quality of the measurement tool.[21,22] The purpose of the present study was to assess the reliability of MSD symptom prevalence indicated on an initial test as compared to the prevalence determined on a follow-up of the same test items several weeks later. Self-administered questionnaires assessing MSDs are commonly used in

TABLE IV
Agreement between test and retest response to statements concerning psychosocial factors at work[A]

Psychosocial item	κ	95% CI[B]
I have a great deal of influence over the amount of work that I do.	0.49	0.34–0.63
I have a great deal of influence over the quality of work that I do.	0.55	0.34–0.76
My supervisor goes out of his or her way to make my work life easier.	0.53	0.39–0.67
Employee health and safety is very important to management.	0.25	0.05–0.45
I know what is expected from me at work.	0.50	0.24–0.77
I have the materials and equipment I need to do my work right.	0.43	0.26–0.60
I have the opportunity to do what I do best every day.	0.43	0.27–0.58
In the last 7 days, I have received recognition or praise for good work.	0.36	0.21–0.51
My supervisor or someone at work seems to care about me as a person.	0.42	0.27–0.57
Someone at work encourages my development.	0.41	0.26–0.56

[A]Response categories of "agree," "neutral" or "disagree."
[B]95% confidence interval for $\kappa = \kappa \pm 1.96\ (SE_\kappa)$.

TABLE V
Agreement between test and retest for questions regarding hand symptoms

	κ	95% CI[A]	P_1[B]	P_2[C]
Symptoms present[D]				
Left hand	0.63	0.48–0.79	0.44	0.41
Right hand	0.64	0.49–0.79	0.47	0.44
Symptom location				
Left hand	0.55	0.36–0.75	0.24	0.26
Right hand	0.50	0.30–0.70	0.26	0.24
Nocturnal signs				
Left hand	0.72	0.52–0.91	0.14	0.17
Right hand	0.58	0.35–0.80	0.15	0.18

[A]95% confidence interval for $\kappa = \kappa \pm 1.96\ (SE_\kappa)$.
[B]Prevalence of "positive" responses in initial test.
[C]Prevalence of "positive" responses in retest.
[D]Numbness, pain, and/or burning localized in two or more of the first four fingers corresponding to the median nerve distribution.

occupational settings to investigate MSD symptom prevalence and to assess changes in symptom prevalence following intervention programs. Questionnaire reliability is critical if the instrument is used to assess intervention effectiveness. If the items on the questionnaire do not yield a similar MSD prevalence from week to week, they cannot be used to assess changes in MSD symptom prevalence from year to year.

Values of kappa greater than 0.75 are considered excellent agreement, values between 0.40 and 0.75 are considered fair to good, and values of less than 0.40 are considered poor.[20] The majority of questionnaire items assessed in the present study had kappa coefficients in the fair to good range. The majority of kappa coefficients for MSD symptom variables exceeded 0.55, which indicates acceptable test-retest repeatability for epidemiological studies. However, for some isolated variables, kappa values were in the "poor" range (less than $\kappa = 0.4$). The low kappa coefficient determined for several MSD symptom questions were likely related to the low prevalence of the observed "yes" responses on the initial tests and retests. As the true prevalence of MSD symptoms approach zero or 100 percent in either the initial test (P_1) or the retest (P_2), the kappa coefficients approach a value of zero.[19] Although the agreement between the test and retest answers could be high, the kappa coefficient is low when the prevalence of "yes" responses is low as in many of the MSD symptom items. If either P_1 or P_2 has a prevalence of 0.0 as seen for several questionnaire items in Table III, the resulting kappa coefficient is undefined. Negative kappa values occur when the probability of observed agreement is less than the probability of expected agreement. Several of the MSD symptom variables resulted in negative kappa values (Table III). Despite a relatively high prevalence of job-related symptoms in the feet as recorded on the initial and retests, a low kappa value ($\kappa = 0.21$) was found. In this case, the low kappa was likely the result of poor recall or questionnaire design rather than a prevalence issue. Interestingly, the majority of kappa values for physician visits were higher than those for job-related symptoms in the same anatomical area (Table III). This finding may indicate that workers have a better recall of physician visits than job-related symptoms during the previous 12 months. However, caution should be exercised when interpreting the differences in these kappa values due to the relatively low prevalence (range 1%–14%) of physician visits by participants in this study sample. Inconsistent answers from only one or two participants for the physician visit item would have a large effect on the kappa values.

The job factors portion of the questionnaire had the largest prevalence of "positive" responses and had consistently higher kappa coefficients than other portions of the survey. The 15 job factors listed in the questionnaire were descriptions of common working situations that could put the worker at increased risk for developing MSDs. Although these factors are not concise measurements of exposure to tasks such as "working in awkward or cramped positions," they give an indication of how much an employee perceives the described task as being problematic. In ergonomics it is not only important to improve work tasks where clear risk factors exist, but also to address job tasks that employees perceive as being problematic.

Hand symptoms identified on the hand diagram of the present study also demonstrated fair to good test-retest reliability with kappa coefficients ranging from 0.50 to 0.72. Rather than analyzing individual hand symptoms such as "tingling" or "numbness" by themselves we chose to combine several symptoms as might be done when assessing the hand for neurological signs and symptoms. This approach has been used previously for assessing test-retest reliability of hand symptoms.[5] The kappa coefficients determined for test-retest reliability for right and left hand symptoms in the present study were 0.64 and 0.63, respectively. The test-retest reliability was also fair to good for the location of hand symptoms as well as the nocturnal hand symptoms in the present study.

Other MSD questionnaires and survey instruments have yielded kappa values in a similar range to those found in the present study.[6,12] In a test-retest reliability study of an upper extremity discomfort questionnaire among 148 industrial workers, Franzblau and colleagues reported kappa coefficients ranging between 0.39 and 0.86 for neck and upper limb symptoms.[6] They reported kappa values of 0.65 and 0.74 for their symptom questions in the right and left hands, respectively, which are very close to those found in the present study for hand symptoms. The authors concluded that the MSD discomfort questionnaire was suitable for epidemiological studies. Franzblau and colleagues also suggested that for "reassurance of the robustness" of their results, similar studies should be performed in other populations and that other MSD questionnaire instruments should be evaluated for reliability.[6] In a study of questionnaire reliability assessing MSD symptoms and work history in a sample of 49 construction workers, Booth-Jones and colleagues reported

kappa values between 0.46 and 0.77 depending upon the section of the questionnaire.[12] The authors collapsed their data by questionnaire section and reported one kappa coefficient for each section. Questionnaire items referring to acute injuries to a body region had the highest kappa coefficient ($\kappa = 0.77$) whereas MSD symptoms had the lowest kappa ($\kappa = 0.46$). Psychosocial items on the questionnaire had a kappa coefficient of 0.61[12] as compared to our range of 0.25 to 0.55 for the psychosocial items. The different kappa values between items in the present study and previous studies may be attributed to methodological differences such as the population under study, testing conditions, instructions to participants, specific wording of questionnaire items, and methods used to determine observed prevalence (i.e., collapsing data, deriving binomial data, etc.). However, despite all of the possible differences in methodological approaches, it appears the MSD symptom questionnaires show adequate reliability to use in ergonomic studies.

There are several limitations inherent in the design of test-retest reliability studies as well as limitations in the use of the kappa coefficient to assess reliability. Changes in MSD symptom prevalence between the initial test and retest may influence the kappa values. If the interval between tests is long, lower kappa values may reflect true changes in prevalence that have occurred for the subjects, rather than incorrect disagreement between testing rounds. If the time interval is too short, it is possible that agreement between testing rounds may simply reflect the subjects' short-term recall of previous answers. Additionally, the initial questionnaire administration may "sensitize"[6] an employee such that he or she may be more or less likely to report MSD symptoms on the retest. We attempted to gain insight to the "sensitization" by investigating test-retest reliability with different time durations between the initial test and retest. Only one questionnaire item had significantly different agreement between the two test-retest durations. Thus, we suspect that there was little difference in the degree of sensitization when using a two- versus four-week test-retest duration.

When interpreting the findings in the present study and other reports of reliability, it is important to understand the limitations of the kappa coefficient. Although the kappa coefficient is relatively simple and widely used to test agreement, several authors have described disadvantages of using the kappa with marginal proportions.[19,22] For example, when the prevalence of "yes" or "no " responses is very low, the kappa statistic is likely to be very low even when the agreement between the initial test and re-test is very high. This effect, or paradox, has been termed "prevalence effect on kappa."[19] The prevalence effect likely influenced the kappa coefficients in the MSD symptom section of the present study where the prevalence of "yes" responses was marginal.

An important consideration for any study involving a questionnaire instrument is practicality. The questionnaire must be acceptable to the employees and to those representing the interests of the employees such as the union or employer. The length of time it takes to administer the questionnaire has implications for feasibility, cost, respondent fatigue, and the quality of data. The questionnaire used in the present study was acceptable to the management team but we did not seek approval from the employees before administering the instrument. We recommend that this type of questionnaire instrument be approved by an employee team or by union representatives if applicable. Prior employee approval will likely increase the participation rate. Additionally, the use of monetary incentives or prize drawings would also likely enhance participant motivation and the response rate. The questionnaire required approximately 15 minutes to complete, which management indicated was acceptable given that employees were being paid their regular or overtime wage at the time. Lengthier questionnaire instruments that require additional time may be difficult to cost justify to the employer.

Other practicality issues such as the setting and environment where the questionnaire instrument is administered should be considered. The testing conditions and environment in the present study represented conditions that are typical of many workplaces where the questionnaire may be administered. The employees in the present study were either beginning or ending their shift or were on a work break and in a hurry to complete the survey. The majority of participants had relatively little interest or incentive to complete the survey, and a few were illiterate and needed the questionnaire read to them. Thus, the conditions under which the present test-retest reliability study was conducted were less than ideal for demonstrating reliability.

It is likely that more controlled testing conditions (such as in a quiet office setting) would have resulted in higher test-retest agreement for some questionnaire items. However, health and safety researchers have little control over actual workplace operations and often administer questionnaires under conditions similar to those in the present study. We expect that the instrument employed in the present study would have a similar temporal stability in other occupational settings involved in production operations and manufacturing.

In ergonomic field studies, questionnaires can be useful for assessing the presence of work-related MSD symptoms, characterizing job factors associated with the development of these disorders, and for assessing symptoms and associated outcomes before and after ergonomic intervention. Identification of work-related MSD symptoms at an early stage in the disease process could trigger the implementation of control measures that subsequently reduce the severity or further development of MSDs. Musculoskeletal symptoms by definition are subjective complaints and are commonly described as pain, tingling, numbness, and aching. Since symptoms are often the earliest clinical manifestations of MSDs, establishing the presence and severity of symptoms is critical to developing prevention programs.

Although many U.S. employers are required to maintain records of all occupational illnesses and injuries, it is generally agreed that these records (OSHA 200 logs) underestimate the true prevalence of work-related MSDs. Additionally, the

federally mandated records are a lagging indicator of injuries and illnesses. Symptoms, however, may be an indication of current or future disease. Thus, symptom and job factor questionnaires can be utilized as an active surveillance tool in the prevention of work-related MSDs. Musculoskeletal disorder questionnaire instruments also provide workers with the opportunity to communicate and express their concerns regarding MSDs. Allowing workers to express their concerns regarding MSDs is an important element in a participatory ergonomics process.

The identification of problematic job factors can assist with prioritizing ergonomic interventions. It is often difficult to determine where resources should be directed or how they should be prioritized when there are many ergonomic concerns. Identifying the departments or business units with the most problematic job factors would provide additional data for targeting interventions.

Data derived from symptom and job factor questionnaires can also be utilized as an additional measure of ergonomic intervention/program effectiveness. Although it is often difficult to control for other factors (economic conditions, other safety programs, changes in management, changes in the manufacturing process) that may affect intervention effectiveness, questionnaires can help document changes in MSD symptoms and perceptions of job factors from year to year. It is important to establish a baseline measurement of symptoms and job factors before the initiation of ergonomic programs. It is also critical that the questionnaire instrument have adequate temporal stability if it is used in the assessment of intervention effectiveness.

The construction and evaluation of health status questionnaires is often performed by applying the methods of psychometrics. The psychometric literature provides criteria for assessing the performance of measurement instruments.[21] The most commonly used criteria are practicality, reliability, internal consistency, validity, and responsiveness to change.[21] The primary purpose of the present study was to assess the reliability of the questionnaire items. Issues regarding the practicality were also addressed. In the present study we did not attempt to evaluate validity including content validity, face validity, and construct validity, or issues related to internal consistency and responsiveness. These issues should be addressed in future studies involving questionnaire instruments in health and safety research.

CONCLUSIONS

The questionnaire used in the present study demonstrated fair to good test-retest reliability for the majority of MSD symptoms, job factors, psychosocial items, and specific hand symptoms. The results indicate that the questionnaire items on the instrument are reliable and suitable for assessing MSD symptoms and job factors in the workplace. The reliability of MSD symptom questionnaires should be established before they are employed to assess the effectiveness of ergonomic or health and safety interventions in the workplace.

ACKNOWLEDGMENTS

This study was supported by grants CDC/DHHS/NIOSH 5KO1/OH00182-02 and CDC/DHHS/NIOSH Cooperative Agreement No. U60/CCU317202.

REFERENCES

1. U.S. Department of Labor: Lost-Worktime Injuries and Illnesses: Characteristics and Resulting Time Away from Work, 1998. Bureau of Labor Statistics, Washington, DC (2000). Web Page available at http://stats.bls.gov/oshahome.htm.
2. National Research Council and Institute of Medicine: Musculoskeletal Disorders and the Workplace, pp. ES1. National Academy Press, Washington, DC (2001).
3. National Institute for Occupational Safety and Health (NIOSH): National Occupational Research Agenda, pp. 1–6. DHHS (NIOSH) Pub. No. 96-115. NIOSH, Cincinnati, OH (1996).
4. Deakin, J.M.; Stevenson, J.M.; Vail, G.R.; et al.: The Use of the Nordic Questionnaire in an Industrial Setting: A Case Study. Appl Ergon 25(3):182–185 (1994).
5. Franzblau, A.; Werner, R.; Valle, J.; et al.: Workplace Surveillance for Carpal Tunnel Syndrome: A Comparison of Methods. J Occup Rehab 3(1):1–14 (1993).
6. Franzblau, A.; Salerno, D.F.; Armstrong, T.J.; et al.: Test-Retest Reliability of an Upper-Extremity Discomfort Questionnaire in an Industrial Population. Scand J Work Environ Health 23(4):299–307 (1997).
7. Bork, B.E.; Cook, T.M.; Rosecrance, J.C.; et al.: Work-Related Musculoskeletal Disorders Among Physical Therapists. Phys Ther 76:827–835 (1996).
8. Cook, T.M.; Hatiar, K.; Zimmermann, C.L.; et al.: Differences in Musculoskeletal Symptoms and Job Factor Perceptions Between Slovak and U.S. Construction Workers. In: Implementation of Safety and Health on Construction Sites: Proceedings of the Second International Conference of CIB, Honolulu, 1999, A. Singh, J. Hinze, R.J. Coble, eds., pp. 705–710. A.A. Balkema Publishers, Rotterdam, Netherlands (1999).
9. Zimmermann, C.; Cook; T.M.; Rosecrance, J.: Operating Engineers: Work-Related Musculoskeletal Disorders and the Trade. Appl Occup Environ Hyg 12:670–680 (1997).
10. Rosecrance, J.; Cook, T.M.; Zimmermann, C.L.: Work-Related Musculoskeletal Disorders Among Construction Workers in the Pipe Trades. Work 7:13–20 (1996).
11. Goldsheyder, D.; Schecter, S.W.; Nordin, M.; et al.: Self-Reported Musculoskeletal Symptoms Among Mason Tenders and Perceived Job-Related Characteristics. Proceedings of the 2000 International Ergonomics Association/Human Factors and Ergonomics Society Congress, San Diego, 2000, pp. 636–639 (vol V). Human Factors and Ergonomics Society Publishers, Santa Monica, CA (2000).
12. Booth-Jones, A.D.; Lemasters, G.K.; Succop, P.; et al.: Reliability of Questionnaire Information Measuring Musculoskeletal Symptoms and Work Histories. Am Indus Hyg Assoc J 59:20–24 (1998).
13. Kuorinka, I.; Jonsson, B.; Kilbom, A.; et al.: Standardised Nordic Questionnaires for the Analysis of Musculoskeletal Symptoms. Appl Ergon 18(3):233–237 (1987).

14. Hurrell, J.J., Jr.; McLaney, M.A.: Exposure to Job Stress: A New Psychometric Instrument. Scand J Work Environ Health 14:27–28 (1988).
15. Buckingham, M.; Coffman C.: First, Break All The Rules, pp. 37–41. Simon & Schuster, New York (1999).
16. Katz, J.N.; Stirrat, C.S.; Larson, M.G.; et al.: A Self-Administered Hand Symptom Diagram for the Diagnosis and Epidemiologic Study of Carpal Tunnel Syndrome. J Rheumatol 17:1495–1498 (1990).
17. Franzblau, A.; Werner, R.; Albers, J.W.; et al.: Workplace Surveillance for Carpal Tunnel Syndrome Using Hand Diagrams. J Occup Rehab 4:185–198 (1994).
18. Cohen, J.A.: A Coefficient of Agreement for Nominal Scales. Educ Psychol Meas 20:37–46 (1960).
19. Hoehler F.H.: Bias and Prevalence Effects on Kappa Viewed in Terms of Sensitivity and Specificity. J Clin Epidemiol 53:499–503 (2000).
20. Fleiss, J.L.: Statistical Methods for Rates and Proportions, 2nd ed. John Wiley & Sons, New York (1981).
21. MacDowell, I.; Newell, C.: Measuring Health, A Guide to Rating Scales and Questionnaires, pp. 31–33. Oxford University Press, New York (1987).
22. Thompson, W.D.; Walter, S.D.: A Reappraisal of the Kappa Coefficient. J Clin Epidemiol 41:949–958 (1988).