AMERICAN JOURNAL OF INDUSTRIAL MEDICINE 29:609-617 (1996)

Evaluation of Symptom Surveys for Occupational Musculoskeletal Disorders

Sherry Baron, MD, MPH, Thomas Hales, MD, and Joseph Hurrell, PhD

Symptom surveys have been used extensively as part of workplace ergonomic screening programs and epidemiologic assessments of musculoskeletal disorders in groups of workers. This paper examines the reliability and validity of two musculoskeletal symptom surveys, the Nordic Musculoskeletal Questionnaire (NMQ) and a survey used in conjunction with epidemiologic assessments by the National Institute for Occupational Safety and Health (NIOSH). Journal articles assessing the validity and reliability of the NMQ were reviewed. A retrospective assessment combining two NIOSH cohorts with a total of 852 workers assessed the reliability and validity of that survey. Reliability was assessed through test-retest methods and interitem correlations between similar questions. Validity was assessed by comparison with results from physical examination assessments of workers and self-reports of workers seeking medical care. Both reliability and validity were found to be acceptable for the purposes of workplace ergonomics programs. Implications for use of these surveys for prevention and treatment outcomes research are discussed. © *1996 Wiley-Liss, Inc.**

KEY WORDS: work-related musculoskeletal disorders, Nordic Musculoskeletal Questionnaire, outcome research, ergonomics

INTRODUCTION

Work-related musculoskeletal disorders of the upper extremities encompass a variety of conditions, including disorders of muscles, tendons, and nerves. Although the underlying pathology of these conditions may differ, their symptom complex is often similar. Recognizing this, Corlett and Bishop [1976] introduced the body map discomfort diagram as a reliable method for assessing ergonomic design changes. Building upon that research, many epidemiologic studies have also used reports of discomfort to determine the prevalence of musculoskeletal disorders in a variety of industries. The most widely used survey tool, particularly in Europe, is the Nordic Musculoskeletal Questionnaire [NMQ; Kuorinka et al., 1987]. In the United States, in addition to the NMQ, a similar symptom survey has been developed [Silverstein et al., 1986] and further adapted by the National Institute for Occupational Safety and Health (NIOSH). An assessment of the reliability and validity of symptom surveys used for epidemiologic purposes may provide useful information for the future development of symptom surveys for the purposes of both workplace screening programs and prevention and treatment outcome research.

The NMQ was developed by a team of Nordic researchers whose goal was to create a simple standardized questionnaire that could be used as a screening method for musculoskeletal disorders as part of ergonomic programs and for epidemiologic studies of musculoskeletal disorders. It was based upon more than 100 projects where one or more of the questions were administered to over 50,000 individuals. It consists of a general questionnaire and a more detailed body part-specific questionnaire. The general questionnaire shows a body map diagram divided into nine an-

Hazard Evaluation and Technical Assistance Branch, National Institute for Occupational Safety and Health, Cincinnati, OH.

Address reprint requests to Sherry Baron, Hazard Evaluation and Technical Assistance Branch, NIOSH, 4676 Columbia Parkway, MS R-10, Cincinnati, OH 45226.

Accepted for publication December 15, 1995.

atomic regions and asks about the presence of "trouble (ache, pain, discomfort)" during the previous 12 months in each of the areas and if the trouble has occurred in the past 7 days. It also includes an indication of severity by using a measure of functional status: "Have you at any time during the last 12 months have been prevented from doing normal work (at home or away from home) because of the trouble?" All answers are in the form of a dichotomous yes/no response.

The NIOSH symptom survey uses a similar body part diagram but adds a series of questions to determine the severity of discomfort by quantifying the frequency, duration, and intensity of symptoms for each body part area using a scaled response. The case definition for a musculoskeletal disorder used in NIOSH studies has incorporated a minimum severity level of musculoskeletal discomfort, e.g., discomfort lasting at least 1 week or of at least moderate intensity. This case definition, which excludes less severe discomfort, such as pain lasting less than a week, was chosen in order to reduce disease misclassification for the purposes of epidemiological analysis.

An assessment of the utility of symptom surveys for research on the outcome of preventive or treatment interventions in the workplace should address several important issues. These include: validity—Does the questionnaire measure what you think it is supposed to measure? Is it practical—Is the questionnaire acceptable to workers, brief and are the results easily analyzed? Is it reproducible—When health status has truly not changed, do the same results appear on retesting? Is it sensitive enough to detect clinically important change over time, even when the change is small [Guyatt et al., 1989]? And is it capable of measuring responses in both a positive as well as a negative direction, or is it hampered by floor or ceiling effects? Some of these issue—validity, reproducibility (reliability), practicality—can be addressed by studies that have used the Nordic and NIOSH questionnaires in the workplace, as described here. Almost all of these studies are cross-sectional in design, or have a limited test-retest component, and therefore cannot address questions of responsiveness to change or floor and ceiling effects, that are important in a longitudinal study. However, these tools have been used extensively in this setting, and thus are attractive to outcome researchers.

This paper examines the utility of these two symptom surveys by reviewing the literature on the acceptability, validity, and reliability of the NMQ and by presenting new analyses which assess the validity and reliability of the NIOSH symptom survey. To accomplish the latter, we looked retrospectively at two NIOSH epidemiologic investigations of musculoskeletal disorders and focused on two issues: 1) the interitem correlation between reports of frequency, duration, and intensity of musculoskeletal discomfort of the hand and 2) the predictive validity of hand discomfort for determining those with physical examination evidence of a musculoskeletal disorder of the hand and those reporting a visit to a health care provider.

MATERIALS AND METHODS

The NMQ

A literature review using the computerized bibliographic databases of the National Library of Medicine and NIOSH, NIOSHTIC®, identified peer-reviewed journal articles from the English literature that addressed the validity and/or reliability of the NMQ. Although many investigations have used the survey, only those that were specifically designed to evaluate the validity and/or reliability were selected for review.

The NIOSH Symptom Survey

Data from NIOSH studies of work-related musculoskeletal disorders that included both a symptom survey and a simultaneous physical examination were combined to perform this analysis. Between 1988 and 1993, NIOSH has conducted eight cross-sectional epidemiologic studies that included a similar symptom survey [Hales and Fine, 1989a,b; Hales et al., 1992; Kiken et al., 1990; Burt et al., 1990; Baron et al., 1990; Bernard et al., 1993; Hoekstra et al., 1994]. Of these, four included both a symptom survey and a simultaneous standardized physical examination for all participants. The wording and format of the questionnaires evolved over time so that only two of these four studies were similar enough in design to allow a combined analysis.

The first of these two studies [Baron et al., 1990; Baron and Habes, 1992] included workers from four different supermarkets, all within a single supermarket chain. The evaluations were done during normal work hours in a medical trailer located in the parking lot outside the supermarkets. The questionnaires were administered by trained interviewers and the examinations were performed by two physicians blinded to the symptom status of the workers. The participation rate during data collection in the field was 65%; however, when follow-up questionnaires were administered by telephone and participation was increased to over 80%, it was established that there was no selection bias in regards to symptom prevalence. Those completing phone interviews had similar disease prevalence rates as those who participated during the field data collection phase.

The second study [Hales et al., 1992, 1994] included office workers using video display terminals from one of the major telecommunications companies. All workers from three cities representing five departments were selected for participation in the investigation. Evaluations were conducted during the normal workday in private conference

TABLE I. Text of NIOSH Symptom Survey (Left Columns Study 1, Right Column Study 2)

In the past year, have you had pain, aching, stiffness, burning numbness, or tingling in the areas shown on this diagram? (picture of hand/wrist shown)

If yes, the following additional questions were answered:

How often have you had this hand/wrist problem in the past year?

Every 6 months	_1	Almost always (daily)	_1
Every 2–3 months	_2	Frequently (once a week)	_2
Once a month	_3	Sometimes (once a month)	_3
Once a week	_4	Rarely (every 2–3 months)	_4
Daily	_5	Almost never (every 6 months)	_5

How long does this hand/wrist problem usually last?

One day or less	_1	Less than 1 hr	_1
More than 1 day, but less than 1 week	_2	1 hour to 1 day	_2
One week, but less than 1 month	_3	More than 1 day to 1 week	_3
One week, but less than 1 month	_4	More than 1 week to 2 weeks	_4
Three months or more	_5	More than 2 weeks to 4 weeks	_5
		More than 1 month to 3 months	_6
		More than 3 months	_7

Describe the intensity of the hand/wrist pain.

No pain	_0	No pain	_1
	_1	Mild pain	_2
	_2	Moderate pain	_3
	_3	Severe pain	_4
	_4	Worst pain ever	_5
Worst pain ever	_5		

Have you had this hand/wrist problem during the past 7 days? _yes _no

rooms in the buildings. The questionnaires were self-administered in groups of five to ten, with an investigator present to answer questions and assure completeness. Four physicians, blinded to symptom status, conducted the physical examinations. Overall participation in the study was 93%.

Questionnaire

The questions used in both of these studies ascertained information on demographics, work characteristics, and musculoskeletal discomfort and disability. The sections on musculoskeletal discomfort contained a series of similar symptom severity questions for each of the major musculoskeletal regions (neck, shoulder, elbow, hand/wrist, and back). Although the questionnaires from the two studies were similar, the response scales to the questions varied slightly (Table I). An additional question asked whether the person had seen a health care provider because of the symptoms. This analysis was limited to discomfort involving only the hand/wrist.

In order to combine the data for hand discomfort from the two study cohorts, the symptom scales were reversed scored, if indicated, and then standardized. Since workers reporting no hand discomfort during the previous year did not respond to the symptom severity questions, the standardized scores did not include asymptomatic participants. Additionally, since those without current pain may differentially report pain and are likely to have different physical examination findings, those who did not report pain in the previous 7 days were analyzed separately.

Physical examination

A standardized physical examination was developed based upon work from previous studies [Silverstein et al., 1986]. The examination was completed in about 10 min without disrobing the participant. The physical examination maneuvers related to the hand and wrist included in this analysis were limited to measures of flexor/extensor tendon disorders and median nerve impairment. Flexor/extensor tendon disorders were considered present when the person reported significant pain on resisted flexion or extension of

the fingers or wrist. Significant pain was defined by reported pain greater than 2 on either a 6-point scale (study 1) or a 5-point scale (study 2). Median nerve impairment was defined as having both a positive 60-sec Phalen's and a positive carpal Tinel's sign. For the purposes of analysis, physical examination of the hand was considered abnormal when a person had either an abnormal tendon-related examination finding or an abnormal median nerve examination finding.

Statistics

Standardization of the symptom scale values was done separately for each study by subtracting the mean value from each observation and then dividing by the standard deviation. After standardization, the two study cohorts were combined into a single data set. Correlation coefficients were calculated for the three standardized scales using the software package PC-SAS version 6.03 (1988).

Analysis of the relationship between physical examination findings and symptom scale values was accomplished by examining the prevalence of abnormal findings for each quartile of the standardized scale values. The relationship between symptom severity level and physical examination outcome was assessed with a chi square for trend analysis using the software package Epiinfo version 5.0 (1990).

Odds ratios and 95% confidence limits were used to assess the association between hand symptom severity and reports of seeking medical care. Odds ratios were calculated using the lower symptom severity group as the reference group and both those with higher symptom severity and those with both higher symptom severity and an abnormal examination as comparison groups. Odds ratios were generated using the statcalc option of Epiinfo version 5.0.

RESULTS

The NMQ

The original Nordic committee published the questionnaire in a journal article [Kuorinka et al., 1987]. This initial report included results on the reliability and validity of the questionnaire. They found that the reliability based on a test-retest methodology used on three separate small cohorts of between 17 and 29 workers ranged from 0 to 23% disagreement. Validity, when compared to clinical history in two cohorts of about 20 workers, varied between 0–20% disagreement. They concluded that this was acceptable reliability and validity for a screening tool that would be used to trigger a more in-depth evaluation. In addition to the authors' initial reports, we identified three subsequent journal articles that have focused on assessing the reliability and validity of the NMQ [Andersson et al., 1987: Ohlsson et al., 1994; Dickinson et al., 1992].

Andersson et al. [1987] found that reporting of discomfort was affected by the way the questionnaire was administered. In this study, the NMQ was completed by bus drivers in two different formats: as part of a workplace investigation of musculoskeletal problems, and as part of a general health questionnaire mailed to the workers' home and then used in conjunction with their annual health examination. The investigators compared the prevalence of musculoskeletal symptoms among three overlapping groups of workers: those who completed their general health questionnaire before the workplace study, those who participated in the workplace study, and those who completed the general questionnaire after the workplace study. This report was limited only to their response to the first of the NMQ questions: "Have you at any time during the last 12 months had trouble (ache, pain, discomfort) in . . ."

The authors found that when the workplace study of musculoskeletal disorders occurred after the workers had already completed their health examination questionnaire, the prevalence rates of musculoskeletal symptoms was higher in the workplace study (Table II, columns 1 and 2). Additional statistical analyses, not shown in Table II, using McNemar's test, found that among the 127 bus drivers who completed both the health examination and then the workplace questionnaire, the pattern of prevalence rates was similar to those in Table II columns 1 and 2, with statistically significant increases in the neck and shoulder regions.

In this same report the investigators found that when a third group of 163 bus drivers completed their general health questionnaire after the worksite evaluation, there were no significant differences in the prevalence rates (Table II, columns 2 and 3). Thus, this study seems to suggest a higher level of reporting of symptoms when the questionnaire was introduced in the context of evaluating work-related musculoskeletal disorders. The increased prevalence rates of symptoms found in the second and third groups were greatest for the body areas for which there were the highest levels of complaints (neck and shoulder). The authors suggest that this pattern may occur because those bus drivers experience the heaviest work loads on their neck and shoulder.

Ohlsson et al. [1994] examined the validity of the NMQ by evaluating the sensitivity and specificity of reports of symptoms during the previous 7 days compared to clinical diagnosis of musculoskeletal disorders using a standardized clinical interview and musculoskeletal examination. Among a group of 165 female workers from a variety of occupations, they found the sensitivity ranged between 66% and 92% and the specificity between 71% and 88% (Table III). The lowest specificity was found in the shoulder, which they attributed to workers misclassifying discomfort resulting from neck conditions as shoulder discomfort. The authors also noted that the low sensitivity for several areas

TABLE II. Effect of Administration Method on Disease Prevalence

	PHE First (%)[a]	Workplace Study (%)[b]	PHE Second (%)[c]
Neck	29	34	36
Shoulder	25	33	38
Elbows	8	11	12
Wrist/hand	8	9	12

[a]Periodic health examination administered before the workplace study (N = 409).
[b]Workplace study of work-related musculoskeletal disorders (N = 477).
[c]Periodic health examination completed after the workplace study (N = 163).
(Reproduced from Andersson et al., 1987, with permission of the publisher.)

suggests that workers may underreport symptoms on standardized surveys.

The last study, by Dickinson et al. [1992], reports the experience from the Health and Safety Executive of the United Kingdom when they pilot tested the NMQ on 481 supermarket cashiers. Using a repeat questionnaire 1 week following initial completion, they found that the rate of nonidentical answers on the questions regarding the prevalence of discomfort during the previous year varied between 7% and 26%. This could be compared to 7% who gave conflicting responses on whether they were right or left handed and 2% who recorded different dates of birth.

The NIOSH Symptom Survey

The two NIOSH study cohorts included 852 workers, 319 supermarket workers and 533 telephone workers (Table IV). Of the 852 workers, 461 (54%) reported some hand discomfort during the past year and 287 (34%) reported hand/wrist discomfort during the previous 7 days. Table V shows the prevalence rate of tendon-related and median nerve-related physical examination findings for each of these three subgroups.

Standardized values for frequency, duration, and intensity of hand discomfort were calculated for the 287 workers who complained of hand discomfort during the previous 7 days. Since those who had no discomfort were not asked these questions, the scales could not include a no discomfort value. Table VI shows the correlation coefficients for these three standardized scales. Although all of these correlations were statistically significant, the magnitude of their association was quite modest.

Each of the three standardized scales were then divided into quartiles, and prevalence rates of abnormal tendon or median nerve physical examination findings were calculated (Table VII). Chi square for trend analysis showed a statistically significant trend ($p < 0.01$) of increasing prevalence of examination findings for all three scales.

The sensitivity and specificity of three alternative case definitions compared to those who had an abnormal hand examination finding were then computed, including: 1) the presence of any hand discomfort during the last 7 days; 2) hand discomfort during the previous 7 days that fell in the upper two quartiles of any one of the three symptom scales; and 3) reported hand discomfort during the previous 7 days that fell in the upper two quartiles of all of the three scales. Table VII shows the sensitivities and specificities for these three case definitions.

To further explore the validity of symptom self-reports, the relationship between hand discomfort and having sought

TABLE III. Sensitivity and Specificity of the NMQ

	Prevalence Rate (%)	Sensitivity (%)	Specificity (%)
Neck	33	66	84
Shoulder	39	92	71
Elbow	20	79	88
Hand	27	67	76

(Reproduced from Ohlsson et al., 1994, with permission of the publisher.)

TABLE IV. Demographics of Participants in Two NIOSH Studies

Study	No.	% Female	Mean age (range)
Supermarket workers	319	62	32 (14–72)
Telephone workers	533	78	37 (19–68)

TABLE V. Prevalence of Abnormality on Physical Examination

		Abnormality		
Hand symptoms during previous 12 months	No.	Tendon (%)[a]	Median nerve (%)[b]	Either
No hand pain	391	4	2	5
Hand pain, not current	164	8	4	11
Current hand pain	287	25	6	26

[a]Pain on resisted flexion or extension of wrist or fingers.
[b]Abnormal 60-sec Phalen and carpal Tinel signs.

TABLE VI. Standardized Symptom Severity Scales Correlation Matrix

	Correlation coefficients		
	Pain frequency	Pain duration	Pain intensity
Pain frequency	1.00	0.38[a]	0.27[a]
Pain duration	0.38[a]	1.00	0.29[a]
Pain intensity	0.27[a]	0.29[a]	1.00

[a]$p < 0.001$.

TABLE VII. Prevalence of Abnormal Hand Examinations By Quartile of Symptom Severity

	Prevalence of abnormal hand findings		
Standardized value	Pain frequency (%)	Pain duration (%)	Pain intensity (%)
No pain	5	5	5
1st Quartile	7	18	20
2nd Quartile	22	20	21
3rd Quartile	46	33	22
4th Quartile	31[a]	37[b]	50[a]

[a]Chi square for trend, $p < 0.001$.
[b]Chi square for trend, $p < 0.01$.

TABLE VIII. Sensitivity and Specificity of Hand Symptom Survey Compared to Abnormalities on Physical Examination

Symptoms	Prevalence rate (%)	Sensitivity (%)	Specificity (%)
Any current hand symptoms	34	79	64
Upper half of any symptom scale[a]	27	71	72
Upper half of all symptom scales[a]	7	32	95

[a]Includes frequency, duration, and intensity of symptoms.

medical care was evaluated. Of the 461 workers who reported some hand pain during the previous 12 months, 30% reported seeing some health care provider for that discomfort; and of the 287 who reported symptoms during the previous 7 days, 36% reported seeking medical care. Table IX shows the proportion of the 287 workers with current hand discomfort who reported seeking medical care broken down into those reporting only mild symptoms (the bottom half of all symptom scales), those reporting more severe discomfort (the upper half of any one of the symptom scales), and those with more severe symptoms who also had an abnormal hand examination. Odds ratios and 95% confidence intervals show statistically significant increased odds for seeking health care among those with more severe symptoms compared to those with only mild discomfort. Among those reporting more severe symptoms, the presence of an abnormal physical examination finding increased the likelihood of seeking medical care (chi square $p = 0.01$).

DISCUSSION

An evaluation of the utility of a symptom survey has to be addressed in the context of the goals for that survey. The

TABLE IX. Percent of Participants Reporting Having Sought Medical Care for Hand Discomfort

Symptoms	% Seeking medical care	Odds ratio	95% confidence interval
Mild hand symptoms[a]	23	1.0	—
More severe hand symptoms[b]	39	2.1	1.0–4.5
More severe symptoms and abnormal physical examination finding	51	3.4	1.4–8.3

[a]Lower half of any symptom scale.
[b]Upper half of any symptom scale.

questionnaires discussed above were designed for use as a screening tool for workplace ergonomics programs or for cross-sectional epidemiologic assessment of groups of workers. For both of these purposes, the requirements of a useful survey are simplicity, acceptability, reliability, and validity. However, these data do not address some of the concerns that would arise in longitudinal outcome studies of preventive or treatment interventions, such as stability, sensitivity to a particular anticipated change, or the absence of floor and ceiling effects [Lohr, 1988]. Given these limitations, we can evaluate the NMQ and the NIOSH symptom surveys.

The extensive use of NMQ throughout the world is evidence enough that it is simple to administer and well accepted by workers. The NIOSH survey was initially adapted from work by other researchers [Silverstein et al., 1986] who have used similar surveys in several large cohorts of workers. In total, over 4,000 workers have completed a version of this survey as part of a NIOSH study. Target populations for the questionnaire have ranged from highly literate newspaper writers to less literate rural chicken processing workers. This too suggests that the format is simple, quick, and easily completed either by workers or by an interviewer. A typical questionnaire which includes demographic data, a series of discomfort questions for each of five body parts, and questions related to lost workdays and visits to a health care provider takes on average 30 min to complete.

Review of the literature assessing reliability of the NMQ found that using test-retest methodology there is, at most, about a 25% rate of nonidentical answers. Evaluation of reliability of questions related to musculoskeletal discomfort using a test-retest format is difficult since the severity of hand pain can vary from day to day depending upon the types of activities the person has engaged in. The reliability testing reported here had either a 1-week [Dickinson et al., 1992] or a 3-week [Kuorinka et al., 1987] interval between questionnaire administrations. Given the

dynamic nature of these conditions, this is probably an acceptable level of reliability for this type of screening questionnaire. In an outcome study, however, if the level of test-retest nonidentical answers is 25%, a small but significant intervention effect could easily be obscured. Also, if the purpose of the questionnaire is to determine whether symptoms are improving or getting worse after treatment or other intervention, a small effect that represents the beginning of an important trend (positive or negative) would also be obscured.

Another test of reliability is to examine interitem correlation for similar questions. In our evaluation of the NIOSH questionnaire, we examined the interitem correlation of the three symptom scales, even though they were not intended to measure identical components of pain. We found that these items were statistically correlated but the magnitude of that correlation suggests that they measure overlapping but different aspects of discomfort.

Andersson et al. [1987] found that the reliability of a symptom survey may be influenced by the method of administration. These findings also suggest that reports of symptoms may be greater when the survey is used as part of an evaluation of specific concerns about workplace exposures. This is an example of the commonly recognized problem of any epidemiologic investigation known as recall or reporting bias. The magnitude of this problem depends upon the goals of the investigation. If the symptom survey is being used as a method of determining priority areas for intervention programs, as long as the survey is filled out in a similar manner by all workers, this would not invalidate its results. However, these findings do suggest that care should be used in comparing results from different studies where the administration methods may differ. In the context of an intervention study, reporting bias will be a consideration unless all workers are enrolled in a randomized, blinded trial. Blinding and placebo intervention is nearly impossible in the workplace setting, where subjects have ample opportunity to discuss allocation to a particular study group.

Our evaluation of the validity of both of these symptom surveys focused on evidence for a relationship between reports of discomfort and other objective measures of impairment, such as physical examination maneuvers. One of the complexities of this analysis is that most musculoskeletal disorders lack a diagnostic "gold standard." Physical examination maneuvers such as Phalen's and Tinel's sign are suggestive but not diagnostic of carpal tunnel syndrome [Katz et al., 1990] and even an abnormal finding on nerve conduction velocity in a previously asymptomatic individual must be interpreted cautiously. Furthermore, the physical examination maneuvers that are used to assess flexor/extensor tendinitis rely on patient self-reports of pain during resisted maneuvers and thus are not truly objective. Nonetheless, an analysis that looks at trends in the relationship between reports of the magnitude of discomfort and the presence of abnormal physical examination findings is important in evaluating the validity of these surveys. In our analysis, we found clear trends between increasing reports of severity on all three scales and a greater prevalence of physical examination findings.

The use of sensitivity and specificity measurements is helpful in examining the predictive validity of these surveys. The primary goal of symptom surveys used as part of workplace ergonomics programs is to trigger more in-depth evaluations of the worker and the workplace. Therefore, a symptom survey must have sufficient sensitivity so that those workers and jobs at the greatest risk are identified for further evaluation. However, specificity is also an important consideration, so that workers are not unnecessarily scrutinized, and that limited medical and ergonomic resources can be most efficiently utilized.

The work by Ohlsson et al. [1994] evaluating the sensitivity and specificity of the NMQ and the analysis of the NIOSH data is not completely comparable. The Ohlsson study used as the standard for comparison a clinical diagnosis based upon a combination of symptoms and abnormal examination findings, while the NIOSH analysis used only abnormal examination findings. The latter is limited by the difficulty of interpreting the significance of an abnormal examination finding in the absence of any symptoms and would tend to underestimate the actual sensitivity and specificity of the questionnaire.

Given that limitation, these results suggest that for use as a screening tool, symptom surveys provide reasonable sensitivity and specificity while remaining extremely simple and easy to administer. There are no clear standards for determining when a survey instrument has sufficient sensitivity and specificity since this determination is dependent upon many factors, including the prevalence of the disorder, the goals of the survey, and the utility and simplicity of alternative surveys or screening tools. Our analysis of the three case definitions suggests that the addition of symptom severity scales is useful in increasing the specificity level above 70%. The case definition that required reporting all three indicators of severity improved specificity even further but had a sensitivity level that was much too low for use as a screening tool. In those situations where greater precision is indicated, either a more in-depth or focused questionnaire, clinical interviews and physical examination maneuver, or other objective testing can supplement the initial screening questionnaire. The additional precision without loss of sensitivity may be especially important in longitudinal investigations. However, in the setting of a prospective follow-up study, it might be acceptable to have a high initial rate of reporting, as the important outcome would be change over time, not the absolute level of symptoms.

An important concern in assessing the predictive validity of symptom surveys, whether or not they are associated

with physical examination findings, is how well symptoms correlate with measures of both vocational and avocation function. In the evaluation of a prevention or treatment outcome, it is difficult to interpret the significance of a change in symptoms unless it is accompanied by demonstration that there is a significant effect on function, health perceptions, or quality of life [Ware, 1991]. In the above analysis of NIOSH data, we found that the likelihood of seeking medical care for hand discomfort increased with increasing levels of severity (frequency, duration, or intensity) of symptoms. We also found that the addition of a physical examination increased the likelihood of identifying those who had sought medical care. Future studies may want to further explore the relationship between musculoskeletal discomfort and reports of function in several domains, self-medication for symptoms, and seeking medical care.

It is difficult to determine from this evaluation the potential utility of symptom surveys when applied to treatment and prevention outcomes research. The symptom surveys discussed above were designed for use in epidemiologic studies and as part of ergonomic programs to examine patterns of disorders within groups of workers. They were not intended for use in clinical diagnosis of patients or for longitudinal follow-up of patients undergoing treatment. Furthermore, since these studies were conducted cross-sectionally at the workplace, most of the workers with discomfort were at an early enough stage of the disease process to be able to continue working. A few studies have used reports of musculoskeletal discomfort to longitudinally assess the efficacy of intervention programs. The Corlett-Bishop body map diagram has successfully been used to follow changes in the frequency and intensity of musculoskeletal discomfort following the introduction of a range of types of ergonomic intervention programs including a workplace exercise program [Silverstein et al., 1988] and the redesign of a supermarket checkstand [Orgel et al., 1992]. Both of these documented high prevalence rates of discomfort both before and after interventions but found that measuring changes in the severity of discomfort was useful in evaluating the efficacy of interventions. However, neither of the study designs allowed the authors to evaluate the reliability or stability of the symptom surveys.

Some studies have also used symptom surveys to longitudinally evaluate outcomes among injured patients. Kemmlert et al. [1993] surveyed patients, using the NMQ, 3 years after a work-related injury and found that 95% of those who still had symptoms (n = 152) reported symptoms in the same anatomical region as was initially injured. They also found associations between severity of symptoms, functional difficulties, and the duration of lost work time. Levine et al. [1993] examined the utility of symptom surveys to evaluate the outcome of carpal tunnel surgery and suggest that they are both reliable and valid. They developed two scales, one based on reported symptoms and the other based on functional status. The questionnaire was given to 38 patients who had undergone carpal tunnel release surgery. They found that both the symptom survey questions and the functional questions had excellent internal consistency, were correlated with some objective measures of impairment, and were able to document improvements. Since this study was limited to carpal tunnel syndrome patients who were being treated in a specialty clinic, it does not address the issue of the usefulness of these instruments for workplace prevention research. Also, since it lacked a comparison group of untreated patients, they were unable to assess the stability of the questionnaire.

Further research will be needed to determine the applicability of symptom surveys in treatment and prevention outcomes research. Important issues that must be addressed include reliability, stability, sensitivity to detect small but clinically important changes, and correlations with functional status or other indicators of impairment such as medical visits, lost work time, or likelihood of progression to more severe conditions.

REFERENCES

Andersson K, Karlehagen S, Jonsson B (1987): The importance of variations in questionnaire administration. Appl Ergonomics 18:229–232.

Baron S, Milliron M, Habes D (1990): Health hazard evaluation report 88-344-2092; Shoprite Supermarkets; New York, New Jersey. US Department of Health and Human Services, Public Health Service, Center for Disease Control and Prevention, National Institute for Occupational Safety and Health, NTIS Report No. PB-91-212431.

Baron SL, Habes D (1992): Occupational musculoskeletal disorders among supermarket cashiers. Scand J Work Environ Health 18 Suppl 2:127–129.

Bernard B, Sauter S, Petersen M, Fine L, Hales T (1993): Health hazard evaluation 90-013-2277; Los Angeles Times, Los Angeles, CA. US Department of Health and Human Services, Public Health Service, Center for Disease Control and Prevention, National Institute for Occupational Safety and Health, NTIS Report No. PB-93-188-456.

Burt S, Hornung R, Fine L, Silverstein B, Armstrong T (1990): Health hazard evaluation report 89-250-2046; Newsday, Inc; Melville, NY. US Department of Health and Human Services, Public Health Service, Center for Disease Control and Prevention, National Institute for Occupational Safety and Health, NTIS Report No. PB-91-116-251.

Corlett EN, Bishop RP (1976): A technique for assessing postural discomfort. Ergonomics 19:175–182.

Dickinson CE, Campion K, Foster AF, Newman SJ, O'Rourke AMT, Thomas PG (1992): Questionnaire development: An examination of the Nordic musculoskeletal questionnaire. Appl Ergonomics 23:197–101.

Guyatt GH, Deyo RA, Charlson M, Levine MN, Mitchell A (1989): Responsiveness and validity in health status measurement: A clarification. J Clin Epidemiol 42:403–408.

Hales T, Fine L (1989a): Health hazard evaluation report 88-180-1958; Morrell & Company, Sioux Falls, South Dakota. US Department of Health and Human Services, Public Health Service, Center for Disease Control and Prevention, National Institute for Occupational Safety and Health, NTIS Report No. PB-80-128992.

Hales T, Fine L (1989b): Health hazard evaluation report 89-251-1997; Cargill Poultry Division, Buena Vista, Georgia. US Department of Health and Human Services, Public Health Service, Center for Disease Control

and Prevention, National Institute for Occupational Safety and Health, NTIS Report No. PB-90-183989.

Hales TR, Sauter SL, Peterson MR, Fine LJ, Putz-Anderson V, Schleifer LR, Ochs TT, Bernard BP (1994): Musculoskeletal disorders among visual display terminal users in a telecommunication company. Ergonomics 37: 1603–1621.

Hoekstra E, Hurrell J, Swanson N (1994): Health hazard evaluation report 92-382-2450; Social Security Administration teleservice centers; Boston, MA. Fort Lauderdale, Florida. US Department of Health and Human Services, Public Health Service, Center for Disease Control and Prevention, National Institute for Occupational Safety and Health.

Katz JN, Larson MG, Sabra A, Krarup C, Stirrat C, Sethi R, Eaton H, Fossel A, Liang M (1990): The carpal tunnel syndrome: Diagnostic utility of the history and physical examination findings. Ann Intern Med 112: 321–327.

Kemmlert K, Orelius-Dallner M, Kilbom A, Gamberale F (1993): A three year follow-up of 195 reported occupational over-exertion Injuries. Scand J Rehab Med 25:16–24.

Kiken S, Stringer W, Fine L, Sinks T, Tanaka S (1990): Health hazard evaluation report 89-307-2009; Purdue Farms, Inc, Lewiston and Roberville, NC. US Department of Health and Human Services, Public Health Service, Center for Disease Control and Prevention, National Institute for Occupational Safety and Health, NTIS Report No. PB-91-104620.

Kuorinka I, Jonsson B, Kilborn A, Vinterberg H, Biering-Sorensen F, Andersson G, Jorgensen K (1987): Standardized Nordic questionnaire for the analysis of musculoskeletal symptoms. Appl Ergonomics 18:233–237.

Levine DW, Simmons BP, Koris MJ, Daltroy LH, Hohl GG, Fossel AH, Katz JN (1993): A self-administered questionnaire for assessment of severity of symptoms and functional status in carpal tunnel syndrome. Bone Joint Surg 75A:1585–1592.

Lohr KN (1988): Outcome measurement: Concepts and questions. Inquiry 25:37–50.

Ohlsson K, Attewell RG, Johnsson B, Ahlm A, Skerfving S (1994): An assessment of neck and upper extremity disorders by questionnaire and clinical examination. Ergonomics 37:891–897.

Silverstein BA, Fine LJ, Armstrong TJ (1986): Hand wrist cumulative trauma disorders in industry. Br J Ind Med 43:779–784.

Ware JE (1991): Conceptualizing and measuring generic health outcomes. Cancer 67(Suppl 1):775–779.