

Expert Consensus Ratings of Job Categories From the Third National Health and Nutrition Examination Survey (NHANES III)

Jennifer C. D'Souza, PhD,¹ W. Monroe Keyserling, PhD,² Robert A. Werner, MD,¹
Brenda Gillespie, PhD,³ and Alfred Franzblau, MD^{1*}

Background A method of occupational physical exposure assessment is needed to improve analyses using large data sets (e.g., national surveys) that provide only job title/category information as a proxy for exposure.

Methods Five ergonomic experts rated and arrived at consensus ratings for job categories used in the Third National Health and Nutrition Examination Survey. Interrater agreement was examined for initial (pre-consensus) ratings. Correlation between consensus ratings and an independent source of ratings (US Department of Labor (DOL)) was used as a basis of comparison.

Results Interrater agreements for the initial ratings were weak. Highest interrater agreement was for sitting (weighted kappa (κ_w) = 0.56). Lowest agreement was for standing (κ_w = 0.07). Consensus ratings were well correlated with DOL ratings (correlation coefficients ≥ 0.6).

Conclusions The correlation between consensus ratings and DOL ratings support the use of expert consensus to assess physical exposures in national data sets. However, validation of expert consensus ratings is required. Am. J. Ind. Med. 50:608–616, 2007.

© 2007 Wiley-Liss, Inc.

KEY WORDS: exposure assessment; expert ratings; NHANES III; occupational epidemiology

INTRODUCTION

Occupational physical exposures at work often result in symptoms and disorders among workers and also in

tremendous costs to both workers and employers [Kelsey and Hochberg, 1988; Yelin, 2003]. Epidemiologic studies that have examined work-related musculoskeletal disorders and symptoms range from studies performed within a single company or hospital/clinic, to postal surveys of communities, to the use of national data sets (e.g., national surveys, record linkages, claims data).

The use of national data has major advantages such as: (1) containing a large number of subjects, (2) being representative of the general population (thus avoiding selection bias), and (3) containing information on important covariates and health outcomes. However, one of the main disadvantages of using national databases/data for occupational epidemiologic studies is the frequent lack of occupational exposure information, which can restrict analyses to associations with job title or crude exposure categories [Anderson and Felson, 1988; Vingard et al., 1991;

¹Department of Environmental Health Sciences, The University of Michigan School of Public Health, Ann Arbor, Michigan

²Department of Industrial and Operations Engineering, The University of Michigan College of Engineering, Ann Arbor, Michigan

³Department of Biostatistics, The University of Michigan School of Public Health, Ann Arbor, Michigan

*Correspondence to: Alfred Franzblau, M6023 SPH II, 109 South Observatory, Ann Arbor, MI 48109-2029.

Accepted 18 May 2007
DOI 10.1002/ajim.20487. Published online in Wiley InterScience
(www.interscience.wiley.com)

Rossignol et al., 2003]. A large portion of studies that used national databases/data have assessed associations between musculoskeletal disorders and particular job categories or titles [Vingard et al., 1991]. When the available job categories/titles were sufficiently precise, exposure information was obtained through other sources (the Dictionary of Occupational Titles) [Anderson and Felson, 1988] or through interviews of the baseline population [Tuchsen et al., 2000]. Some studies have also used expert judgments to categorize occupations as “physically demanding” [de Zwart et al., 1997] or “blue collar” [Rossignol et al., 2003]. Although expert judgments cannot achieve the quality of direct measurements, experts’ experience and knowledge of jobs can provide valuable insight to occupational physical exposures and risks. If the expert ratings of occupational activities were improved, expert ratings could be a practical solution to the lack of exposure information when analyzing very large study populations (e.g., national data/surveys).

A literature search of expert judgments of job titles/categories showed that most were performed in occupational studies of chemical exposures. Thus, the evaluation of the validity and reliability of experts’ ratings has been far more extensive in the area of chemical exposures relative to physical exposures. Although some aspects of assessing chemical exposures are distinct from assessing physical exposures, there are some principles that can be applied to the challenge of rating ergonomic factors. For example, assessments of experts’ ratings of chemical exposures indicated that the following characteristics increased ratings’ validity and reliability: (1) when experts worked as a panel, (2) when experts were familiar with the jobs, and (3) when there were detailed descriptions of the job’s tasks and environment [Siemiatycki et al., 1997].

This article describes a method of assessing selected occupational physical risk factors that incorporates some of the above-mentioned recommendations for expert ratings of job titles/categories, namely (1) using a panel of experts and (2) using experts who are familiar with the jobs’ physical exposures. This method uses job categories from the Third National Health and Nutrition Examination Survey (NHANES III) and exposure to work activities related to knee osteoarthritis (OA).

Objectives

The main objectives of this study were:

- (1) to use experts to develop lower extremity ergonomic exposure ratings for NHANES III job categories that can be used to assess knee OA risk;
- (2) to assess the agreement between expert ratings (i.e., interrater agreement);
- (3) to compare the experts’ ratings with an external source of ratings.

Ultimately, by providing a method that makes greater use of job titles/categories, the results of epidemiologic studies using job titles/categories can be more helpful in identifying specific hazardous exposures and exposure levels associated with work-related musculoskeletal disorders.

BACKGROUND

NHANES III Knee Osteoarthritis (OA) X-Ray Data File

The NHANES III was a national cross-sectional survey of the civilian, non-institutionalized United States population. Subjects received a household interview, which obtained data on variables such as weight and job history. A portion of the subjects were invited for a physical examination. The NHANES III invited all subjects who were aged 60 years and older and who received a physical exam, to have their knees X-rayed while lying down (for safety purposes). The data provided a unique opportunity to study the association of occupational factors and knee OA, while adjusting for important covariates.

The job categories in the publicly available NHANES III data set were based on the 3-digit 1980 US Census Job Codes (approximate $n=900$) that the interviewer used to code subject’s answer to “What kind of work were you doing the longest?” In order to protect the identities of the subjects, the individual job codes/titles were collapsed into 40 job categories that are in the publicly available NHANES III data (Fig. 1).

Since the publicly available job categories encompassed several job codes/titles, some of the job categories had a large amount of physical exposure heterogeneity. For instance, one of the 40 job categories contained writers and athletes, while another job category contained post-secondary and kindergarten teachers. This was an added complication to the exposure assessment. Additionally, there were no adequate physical exposure data (e.g., a database) for the job categories ($n=40$), nor could the subjects be re-contacted and interviewed regarding their occupational activities.

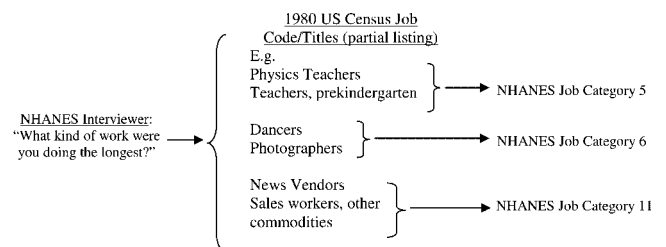


FIGURE 1. Relationship between 1980 US Census Job Code/Titles and NHANES Job category.

Occupational Activities and Exposures for Knee OA

Six occupational activities were chosen based upon the support of previous epidemiologic literature for an association with knee OA: (1) sitting, (2) standing, (3) walking/running, (4) carrying/lifting a heavy load (>10 kg), (5) kneeling, crouching, stooping, crawling, or squatting (abbreviated as “kneeling”), and (6) working in a cramped space. The evidence is strongest for knee bending activities (i.e., kneeling and squatting) and heavy lifting [Anderson and Felson, 1988; Felson et al., 1991; Cooper et al., 1994; Maetzel et al., 1997; Coggon et al., 2000]. These associations have been primarily limited to dichotomous exposure levels. Positive associations, though not always significant, have been observed between knee OA and standing and walking/running [Coggon et al., 2000; Yoshimura et al., 2004]. No studies have examined the association between working in a cramped position and the risk of knee OA. However, “working in a cramped position” may also entail knee bending and thus was also worth further examination.

United States’ Department of Labor (DOL) Ratings

The United States’ DOL had occupational ratings for jobs, in order to give an accurate job characterization for those seeking jobs or careers. These ratings were used to provide a basis for comparison for the expert ratings, and were obtained from the US DOL’s O*Net Analyst Database. This database was developed by the DOL to update the previously published “Dictionary of Occupational Titles.” Some of the variables in the database were related to knee OA: sitting, standing, walking/running, kneeling, working in a cramped position and certain strength variables. These occupational ratings ranged from 1 to 5 for sitting, standing, walking/running, working in a cramped space, and kneeling. The DOL did not have ratings of “heavy lifting” specifically but had ratings of various strength variables with a different scale that ranged from 0 to 7. These data were published

online under the Standard Occupational Classification (SOC) Job Code/Title. There was no data source that directly mapped the SOC Job Titles to the NHANES job category. Instead two data sources were used which indirectly linked the SOC to the NHANES Job Categories: (1) a mapping of SOC to 1980 Census Job Codes (provided by the National Crosswalk Center) and (2) a final mapping of the 1980 Census Job Codes to the NHANES III Job Categories (provided by NCHS).

“Average DOL ratings” were calculated for each NHANES job category using the different mappings. This was performed by taking an unweighted average of the DOL ratings of all the SOC job titles which were mapped to a NHANES III job category. This method is illustrated in Figure 2.

The DOL ratings provided an independent and widely used assessment of physical job stresses related, in part, to the lower extremities.

MATERIALS AND METHODS

Expert Participants

Five ergonomic “experts” were recruited to participate in the study. The ergonomic experts were selected based upon their extensive experience in rating jobs. Three out of the five experts were based in academic settings. One expert was an ergonomic consultant and one expert was an ergonomics researcher at a government agency. The experts’ years of professional experience in the field of ergonomics/occupational health (including job analysis) ranged from approximately 11 years to over 25 years, with an approximate mean of 19 years of experience. All were Certified Professional Ergonomists.

Initial Expert Ratings

The experts were sent a survey that instructed them to rate each of the 40 NHANES III job categories with respect to the proportion of the work day (%) spent in the six

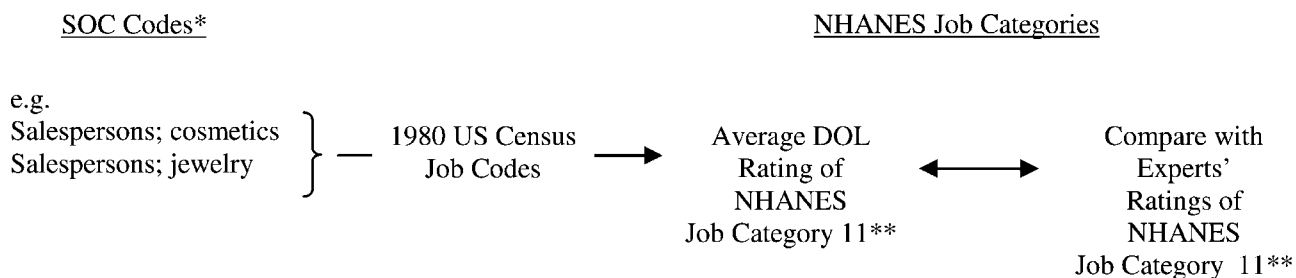


FIGURE 2. Calculating the Average DOL Rating for NHANES Job Category. *Standard Occupation Classification; **NHANES Job Category 11: Sales workers, retail and personal services.

categories of occupational activities. The rating scale was an 11-point, interval scale that ranged from 0% to 100% of the work day. Each unit/interval represented 10% of the work day.

Consensus Rating

Experts were given approximately 1 month to individually complete the survey (referred to as “initial ratings”). Upon completion of the survey, experts convened to discuss their initial ratings and arrive at consensus ratings. Their initial ratings were collected prior to the meeting and entered into a spreadsheet. When the experts met, the spreadsheet of the initial ratings was projected on to a screen so that the

experts could see their initial ratings as well as the ratings of their peers. The initial ratings were not kept anonymous.

Consensus was reached when all of the ratings were within 1-unit of each other. For initial ratings that did not meet the consensus definition, experts with outlying ratings were asked to give the rationale behind their ratings. Then the experts were given the opportunity to change their initial ratings, based upon the discussion. If the changes did not result in a consensus, the process was repeated again and experts were asked to give a rationale for their ratings (Fig. 3).

During these discussions, it was recognized that some of the disagreements in the initial ratings were caused by inconsistent definitions and/or assumptions used by the experts when determining initial exposure scores. In order to

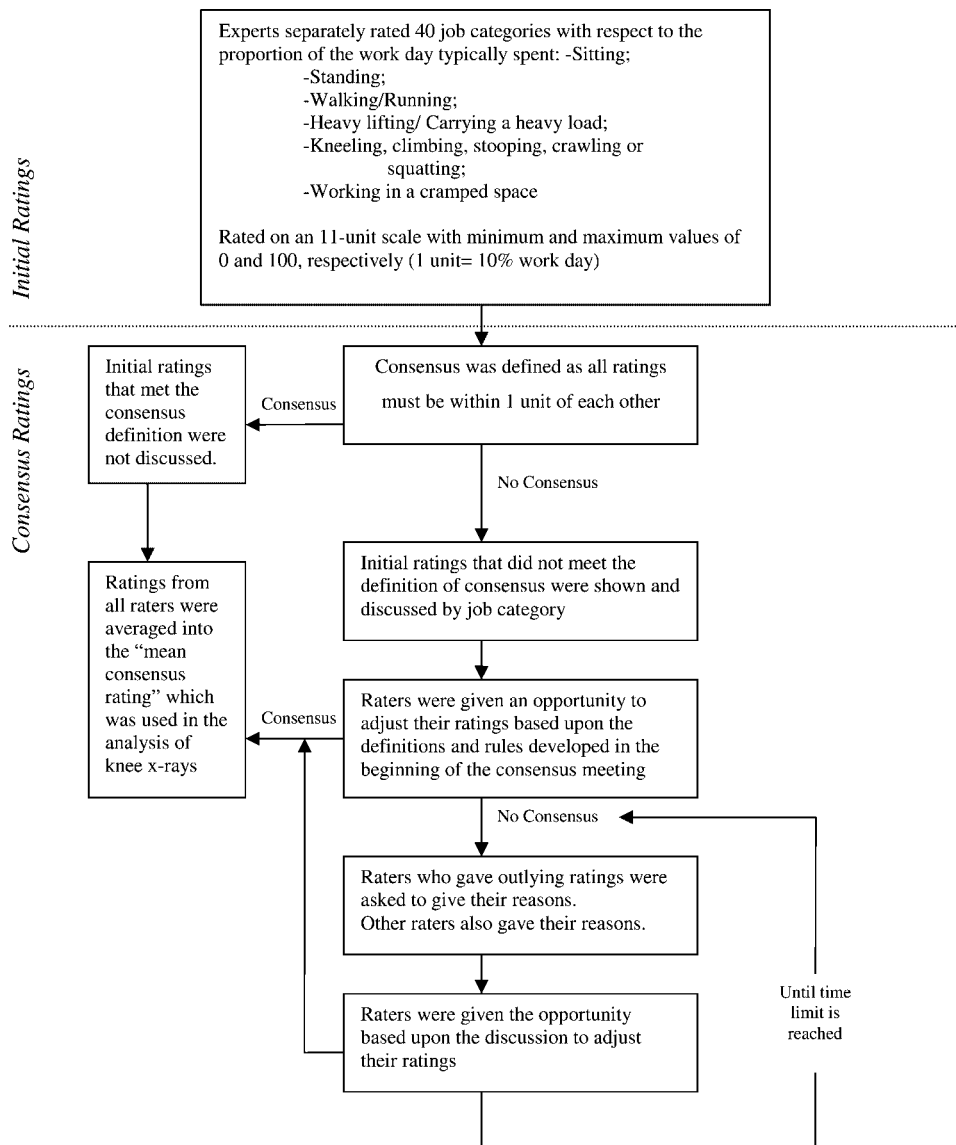


FIGURE 3. Diagram of the process of obtaining expert ratings (of typical exposures).

achieve consensus, assumptions and definitions were discussed, clarified, and revised. Experts developed and used the following definitions and assumptions to establish the final consensus ratings.

Work day: A 1-day shift that could range from 8 to 12 hr (when the worker “clocks-in” and “clocks-out”). This time included lunch breaks and other periodic breaks.

Heavy lifting: When a worker was required to lift a load of 10 kg or more, four or more times per hour. Therefore, a rating of 10% indicated that a worker in the particular job category would lift a load of 10 kg or more, four or more times per hour for approximately 1 hr (i.e., about 10% of the work day).

Carrying a load greater than 10 kg: The actual time a worker spent carrying a load greater than 10 kg.

Kneeling, climbing, stooping, crawling, or squatting (abbreviated as “kneeling”): This occupational activity was intended to capture the amount of time a worker spends with his/her knees flexed, excluding sitting, and walking/running. Therefore, stooping did not refer to the posture of straight legs and a bent back, but must also include flexed knees. In addition, the ratings of this job activity did not distinguish between continuous kneeling position versus periodic bending and straightening of the knees. In other words, a worker who received a rating of 20% may have been in a continuous kneeling position for 20% of the work day or may have been bending and straightening his/her knees many times during the work day, but the accumulated time that his/her knees are “flexed” amounted to 20% of the work day.

Working in a cramped space: This referred to the whole body being cramped and not just arms or hands. This condition was intended to measure the amount of time a worker may spend in a space that did not allow the entire body to stretch or move about “freely.” The condition implied that the worker has flexed knees, but imposed the criteria that the workers’ movements were limited by

space. Therefore, high ratings of “working in a cramped space” imply high ratings of “kneeling,” but not the reverse.

Sitting: This occupational activity referred to the condition where the worker’s weight was supported by a chair or seat.

Standing: This referred to the stationary position where the knees were not flexed. This position did not include or overlap with walking/running.

Walking/Running: This occupational activity referred to the upright standing posture but with “active lower extremities.” Therefore, a worker could not be walking and standing simultaneously.

Kneeling, standing, sitting and walking/running were mutually exclusive. Also, these postures and activities appeared to encompass almost all of the lower extremity positions and activities that a worker may possibly assume or perform. Therefore, the consensus ratings of these four exposures were partly determined based upon the assumption that the sum of the ratings for the four physical exposures would be approximately 100%. Refer to Table I for a partial listing of the consensus ratings for the NHANES III job categories.

Consensus ratings are available upon request from the corresponding author.

Statistics

Agreement between Experts’ Initial Ratings

Weighted kappas (κ_w), which take into account the 11-point ordinal scale, were calculated to evaluate interrater agreement [Fleiss, 1981]. The Cicchetti–Allison weights were used [Cicchetti and Allison, 1971]. Additionally an overall weighted kappa was calculated for each physical activity [Fleiss, 1981].

TABLE I. Expert Consensus Ratings* for Selected NHANES III Job Category

NHANES III Job Category (NHANES III Job Code)	Sitting	Standing	Walking	Kneeling	Heavy lifting	Working in a cramped space
Executive, administrators, and managers (01)	68	16	14	4	6	2
Information clerks (13)	62	28	10	0	4	0
Cleaning and building service occupations (23)	24	26	36	14	14	8
Related agricultural, forestry, and fishing occupations (27)	16	30	36	16	28	6
Construction laborers (37)	10	38	28	24	28	14
Standard deviation	21.01	10.04	10.23	8.44	10.2	7.47

*Ratings represent percent of work day spent in activity.

Intraclass Correlation Coefficients (ICC) were calculated using SPSS v.12.0. These were also used to assess the agreement between raters by treating the ratings on the 11-unit scale as a continuous variable. The ICCs were calculated using a 2-way mixed model (ANOVA), where the experts (raters) were considered fixed effects. Also, the individual ratings were used as the units of analysis (“single measures”).

Spearman Rank Correlation Coefficients were also calculated in order to assess rater’s relative agreement. This method ranks the ratings and measures the correlation among the rankings.

Correlation between expert ratings and external source of ratings (DOL)

Pearson Correlation Coefficients were used to compare expert ratings and DOL ratings. The expert and DOL ratings were not on the same scale, thus only correlations could be examined.

Apart from the ICCs, the analyses were performed using SAS v. 9.1.3.

RESULTS

Inter Rater Agreement

Both the ICCs and overall weighted kappas indicated that agreement among the initial ratings (pre-consensus) was highest for sitting (overall $\kappa_w = 0.56$; ICC = 0.80) and lowest for working in a cramped space (overall $\kappa_w = 0.08$; ICC = 0.40) and standing (overall $\kappa_w = 0.07$; ICC = 0.32). Upon closer investigation of the pairwise interrater weighted kappas for initial ratings of standing, the agreement between raters 1, 2, and 3 were “fair” and neared values of 0.6, which is considered “good agreement” (Fig. 4) [Altman, 1991]. However, the overall poor agreement for initial ratings of standing appeared to have been driven by the low agreement with raters 4 and 5, who tended to give lower ratings relative to the others.

The Spearman Rank Correlation indicated much higher relative agreement, despite the low agreement among the initial ratings. For instance, standing had the lowest overall interrater agreement but the agreement regarding the relative amounts of standing, were much stronger- with many of the Spearman Correlation Coefficients exceeding 0.6 (results can be found in [D’Souza, 2006]). In particular, raters 1 and 4, and 1 and 5, had very low weighted kappas ($\kappa_w = <0.1$), but had Spearman Rank Correlation Coefficients that were slightly greater than 0.6. This was also seen in the initial ratings of working in a cramped space, where the agreement between raters 1 and 4 was low ($\kappa_w = 0.215$) compared to the agreement between raters 1 and 2 ($\kappa_w = 0.581$). But the

Spearman Rank Correlation Coefficients differed by much less (0.534 vs. 0.611).

Consensus Ratings Versus DOL Ratings

The relationship between mean consensus ratings and mean DOL ratings for each job activity was assessed through Pearson Correlation Coefficients. Overall, mean consensus ratings were positively correlated with the mean DOL ratings (Table II). Mean consensus ratings of sitting were the most highly correlated with the DOL ratings (correlation coefficient = 0.892). The mean consensus ratings of standing were the least correlated with DOL ratings (correlation coefficient = 0.597).

In order to partly examine how much the consensus process improved the initial ratings, the Pearson Correlation Coefficients from the following comparisons were evaluated: (1) the correlations between the independent source of ratings (DOL) and the consensus ratings and (2) the correlations between the independent source of ratings (DOL) and each of the raters’ initial ratings (Table II). Overall the correlations between the DOL ratings and the initial ratings are lower than the correlation between the average consensus ratings and the DOL ratings. However, this was not true of ratings of standing where the correlation between some initial ratings and DOL ratings far exceeded the correlation between the consensus ratings and DOL ratings (i.e., raters 1, 2, 3).

DISCUSSION

Summary of Results

In general, experts’ consensus ratings of common lower extremity occupational physical exposures appeared to have good *face* validity, based on the correlations with DOL ratings. The amount of agreement among experts’ initial ratings seemed to be strongly influenced by the degree of similarity between experts’ assumptions and definitions of the physical activities. However, when assumptions and definitions regarding the specific work activities and work day were clearly defined among the experts, the experts were able to reach consensus with relative ease and consensus ratings were often more strongly correlated with DOL ratings than the initial ratings.

Agreement Between Initial Ratings

A characteristic of a useful exposure assessment is its ability to give consistent measures of exposures. In this study, experts were used to give an exposure assessment based upon their professional experience and judgment. Therefore, it was

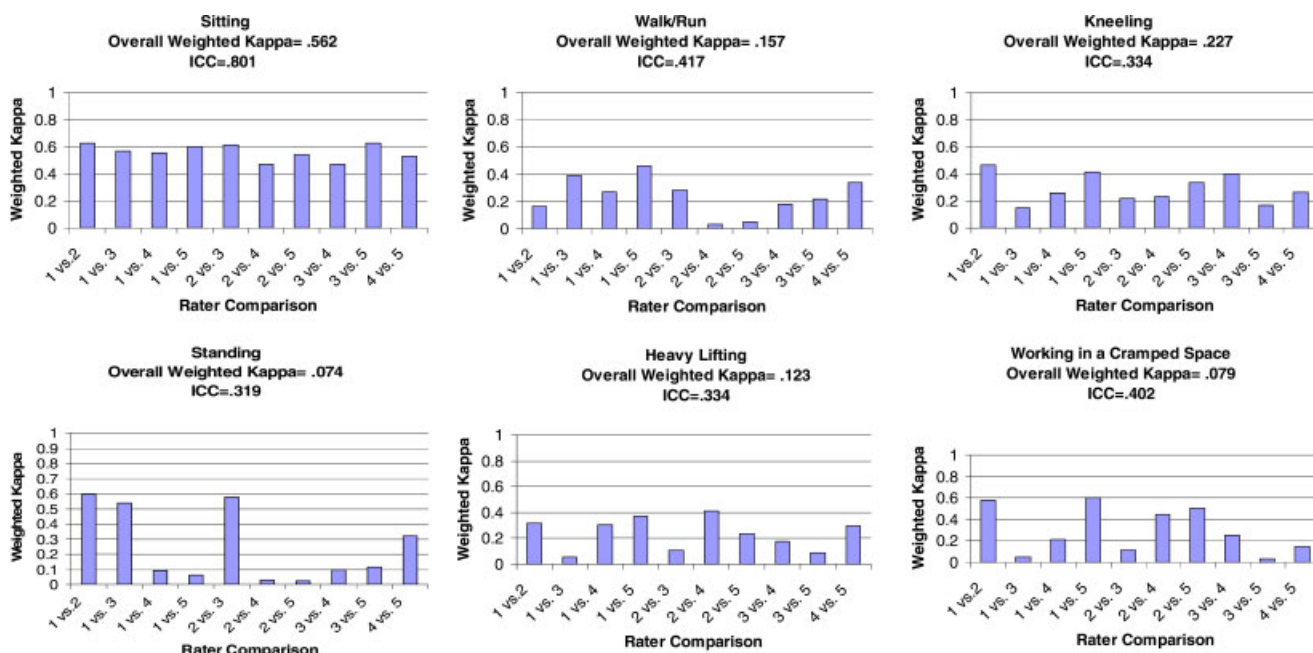


FIGURE 4. Interrater agreement (weighted kappa) by occupational activity.

of interest to examine the agreement among experts’ initial ratings.

Based upon the weighted kappas and ICC, the agreement between the initial ratings of physical activities was poor, with the exception of sitting. Although weighted kappas are suitable for assessing agreement using multiple categories and raters, they are also influenced by the number of rating categories and by the true distribution of the exposures across categories. This study used a relatively large number of rating categories (n = 11). This may have artificially deflated the weighted kappas, since a greater number of rating categories leads to a greater chance of disagreement [Kundel and Polansky, 2003]. Additionally, the true prevalence of the exposures may have also affected the weighted kappa. Extremely high or low prevalence of exposure can lead to lower kappa values [Byrt et al., 1993]. Working in a cramped space was most likely not a highly prevalent job

exposure, and thus may have resulted in a lower overall weighted kappa.

Much of the poor agreement between experts’ initial ratings could also be attributed to experts’ varying definitions of the occupational activities. For example, the lowest amount of agreement was observed for “working in a cramped space” and “standing.” Some raters had defined standing as an upright posture, and thus incorporated walking/running into his/her initial ratings, whereas some of the raters had treated standing as a separate category from walking/running and from sitting. Initially “working in a cramped space” was also defined differently by each expert. Some experts included cramped upper extremity posture thus decreasing the overall amount of agreement among initial ratings.

Also, the high amount of disagreement among the initial ratings could be due to the large amount of exposure

TABLE II. Pearson Correlation Coefficients Between DOL Ratings and Experts’ Average Consensus Ratings and Between DOL Ratings and Experts’ Initial Ratings

	Correlation between DOL and average consensus ratings	Correlation between DOL and initial ratings				
		Rater1	Rater2	Rater3	Rater4	Rater5
Sitting	0.892	0.827	0.862	0.817	0.846	0.809
Standing	0.597	0.817	0.866	0.846	0.580	0.404
Walking	0.745	0.575	0.468	0.768	0.490	0.439
Kneeling	0.858	0.650	0.627	0.829	0.758	0.816
Working in a cramped space	0.710	0.517	0.633	0.693	0.543	0.730

Boldface indicates where correlation between DOL and initial ratings is greater than correlation between DOL and consensus rating.

heterogeneity within a job category. As an extreme example, one job category included writers and athletes. Another job category grouped nursery school teachers with high school teachers. As a result the task of rating the “typical” exposure for such job categories was difficult and may have led to low agreement among the initial ratings.

The variability in the initial ratings was partially due to the lack of agreement on the definitions of specific physical activities, as well as the underlying assumptions of the experts’ ratings. The influence of varying definitions of the work day is illustrated through the high Spearman Rank Correlation Coefficients, where experts were in agreement regarding the relative amounts of exposure among job categories (i.e., job category 5 was exposed to more kneeling activities than job category 20). These results suggest that if the experts had similar definitions of the work day some of the disagreement in the experts’ initial ratings could have been eliminated.

The results show that having clear definitions and assumptions is important in achieving high initial agreement. In this study, the consensus process was necessary in identifying these issues and allowing experts to discuss them. But this could have also been achieved through further pilot testing. Additionally, the results show that the consensus process appeared to improve the correlation between the consensus ratings and the DOL ratings, which was almost always higher than the correlation between the initial ratings and DOL ratings. Based on correlations with the DOL ratings, the consensus process was certainly more valuable for certain activities than others. However, comparisons with an independent source of ratings are limited and do not indicate true validity.

Consensus Versus DOL Ratings

In addition to producing stable and reproducible measurements, exposure assessments should also provide valid exposure measures. Due to the unavailability of specific descriptions of subjects’ longest held jobs and the lack of a “standard,” an assessment of the validity of the experts’ ratings was not possible. Therefore, an attempt was made to compare experts’ ratings with the ratings obtained by the DOL as an alternative to directly validating the exposures. Although it was not known how valid the DOL ratings were, the ratings were of specific job codes (versus broad, heterogeneous job categories) and were also based on the job’s task statements. The DOL ratings were an independent source of ratings, which made it less likely that the observed positive correlations between the DOL ratings and expert ratings were due to chance. These significant correlations provide evidence that the expert ratings can achieve the quality of DOL ratings. This can be useful to occupational epidemiologists whose exposure assessments are often limited to job titles or groups. It is also somewhat remarkable that the experts’ ratings of the most likely exposures in a

broad job category were well correlated with the mean of the individual DOL ratings within the broad job category.

Although the significant positive correlations with the DOL ratings suggest that the expert ratings had a certain level of quality, the DOL rating scale itself was difficult to interpret for ergonomic risk. Since these ratings were developed for career purposes, the ratings were not required to be very exact and were constructed to give only a crude sense of the physical demands of a particular job. Although it is encouraging that the experts’ ratings correlated well with DOL ratings, the comparison with the DOL ratings was not a validation of experts’ ratings and a rigorous validation is clearly necessary.

Limitations

A major limitation of the study was the inability to validate the expert ratings. The large heterogeneity of exposure within the job category was also a complication. Although the expert ratings appeared to have good *face* validity, it was unclear how accurately the ratings reflected the true exposure of workers within each NHANES III job category. DOL ratings were used for comparison and to, in part, provide a separate data source that could provide a reference. As previously mentioned, the fact that both the DOL ratings and expert ratings were well correlated suggested that the expert ratings can achieve the quality of DOL ratings.

The lack of clear and agreed upon definitions of the physical activities was also a limitation in this study. It seems that the amount of variance in the initial ratings was mainly attributed to the differing definitions of the physical variables and the underlying assumptions. For instance rater 5 assessed exposures relative to the ratings given to the other activities. Evaluating exposures with this assumption may have led to overall lower ratings relative to the other raters. These discrepancies were resolved prior to developing the consensus ratings.

Although the strong correlations between experts’ consensus ratings and DOL ratings are promising, the physical exposures in a job could change over time and thus the experts’ ratings may not have reflected the *subject’s* exposures at the time they were employed. Since these ratings will be used as exposures for a currently older population, the experts may not have had professional experience with the subjects’ job tasks and physical exposures, some of which may have occurred 30–40 years prior to the experts’ assessment.

CONCLUSIONS

This study examined a method of obtaining physical job exposure ratings for use in epidemiologic studies. The study used the experience and judgment of expert ergonomists to

rate job categories and to arrive at consensus ratings for the "typical" exposure of a job category.

The analysis of the initial ratings shows that the interrater agreement was highest when the physical variables were well defined beforehand, as was observed for the physical variable sitting. Based on the experts' discussion prior to the consensus ratings, it was important to include in the definition the purpose of capturing the exposure to a certain physical variable. For example, in the physical variable "kneeling, climbing, stooping, crawling, or squatting," experts were not aware that this physical variable was intended to capture "knee-bending." Consequently, when the experts were informed of that fact, they were able to discuss and arrive at consensus ratings more efficiently.

When the experts' consensus ratings were compared with the DOL ratings, the ratings were well correlated. While individual, objective exposure assessment remains the gold standard, as a practical matter, using experts to assess exposures may prove to be valuable in the epidemiologic setting. It is relatively quick and inexpensive and may supplement or potentially provide a better alternative to self-reported exposures and/or dichotomous exposure categories that are commonly used in epidemiologic studies.

REFERENCES

- Altman DG. 1991. *Practical statistics for medical research*. London: Chapman and Hall.
- Anderson JJ, Felson DT. 1988. Factors associated with osteoarthritis of the knee in the first national Health and Nutrition Examination Survey (HANES I). Evidence for an association with overweight, race, and physical demands of work. *Am J Epidemiol* 128:179–189.
- Byrt T, Bishop J, Carlin JB. 1993. Bias, prevalence and kappa. *J Clin Epidemiol* 46:423–429.
- Cicchetti DV, Allison T. 1971. A New Procedure for Assessing Reliability of Scoring EEG Sleep Recordings. *Am J EEG Technol* 11:101–109.
- Coggon D, Croft P, Kellingray S, Barrett D, McLaren M, Cooper C. 2000. Occupational physical activities and osteoarthritis of the knee. *Arthritis Rheum* 43:1443–1449.
- Cooper C, McAlindon T, Coggon D, Egger P, Dieppe P. 1994. Occupational activity and osteoarthritis of the knee. *Ann Rheum Dis* 53:90–93.
- de Zwart BC, Broersen JP, Frings-Dresen MH, van Dijk FJ. 1997. Repeated survey on changes in musculoskeletal complaints relative to age and work demands. *Occup Environ Med* 54:793–799.
- D'Souza JC. 2006. Occupational Activities and Knee Osteoarthritis: An analysis of the Third National Health and Nutrition Examination Survey (NHANES III) and Expert Ratings: The University of Michigan School of Public Health.
- Felson DT, Hannan MT, Naimark A, Berkeley J, Gordon G, Wilson PW, Anderson J. 1991. Occupational physical demands, knee bending, and knee osteoarthritis: Results from the Framingham Study. *J Rheumatol* 18:1587–1592.
- Fleiss J. 1981. *The measurement of interrater agreement. Statistical Methods for Rates and Proportions*. New York: Wiley. p 212–237.
- Kelsey JL, Hochberg MC. 1988. Epidemiology of chronic musculoskeletal disorders. *Annu Rev Public Health* 9:379–401.
- Kundel HL, Polansky M. 2003. Measurement of observer agreement. *Radiology* 228:303–308.
- Maetzel A, Makela M, Hawker G, Bombardier C. 1997. Osteoarthritis of the hip and knee and mechanical occupational exposure--a systematic overview of the evidence. *J Rheumatol* 24:1599–1607.
- Rossignol M, Leclerc A, Hilliquin P, Allaert FA, Rozenberg S, Valat JP, Avouac B, Coste P, Savarieau B, Fautrel B. 2003. Primary osteoarthritis and occupations: A national cross sectional survey of 10 412 symptomatic patients. *Occup Environ Med* 60:882–886.
- Siemiatycki J, Fritschi L, Nadon L, Gerin M. 1997. Reliability of an expert rating procedure for retrospective assessment of occupational exposures in community-based case-control studies. *Am J Ind Med* 31:280–286.
- Tuchsen F, Krause N, Hannerz H, Burr H, Kristensen TS. 2000. Standing at work and varicose veins. *Scand J Work Environ Health* 26:414–420.
- Vingard E, Alfredsson L, Goldie I, Hogstedt C. 1991. Occupation and osteoarthritis of the hip and knee: A register-based cohort study. *Int J Epidemiol* 20:1025–1031.
- Yelin E. 2003. The economics of osteoarthritis. In: Lohmander LS, editor. *Osteoarthritis* 2nd ed. New York: Oxford University Press. p 17–21.
- Yoshimura N, Nishioka S, Kinoshita H, Hori N, Nishioka T, Ryujin M, Mantani Y, Miyake M, Coggon D, Cooper C. 2004. Risk factors for knee osteoarthritis in Japanese women: Heavy weight, previous joint injuries, and occupational activities. *J Rheumatol* 31:157–162.