Scand J Work Environ Health 2001;27(1):70—75

Comparison of expert-rater methods for assessing psychosocial job strain

by Aleck S Ostry, PhD,[1] Steve A Marion, MD,[1] Paul A Demers, PhD,[1] Ruth Hershler, MSc,[1] Shona Kelly, BSc,[1] Kay Teschke, PhD,[1] Cameron Mustard, SD,[2] Clyde Hertzman, MD [1]

Ostry AS, Marion SA, Demers PA, Hershler R, Kelly S, Teschke K, Mustard C, Hertzman C. Comparison of expert-rater methods for assessing psychosocial job strain *Scand J Work Environ Health* 2001;27(1):70–75.

Objectives This study tested the reliability and validity of industry- and mill-level expert methods for measuring psychosocial work conditions in British Columbia sawmills using the demand-control model.

Methods In the industry-level method 4 sawmill job evaluators estimated psychosocial work conditions at a generic sawmill. In the mill-level method panels of experienced sawmill workers estimated psychosocial work conditions at 3 sawmills. Scores for psychosocial work conditions were developed using both expert methods and applied to job titles in a sawmill worker database containing self-reported health status and heart disease. The interrater reliability and the concurrent and predictive validity of the expert rater methods were assessed.

Results The interrater reliability and concurrent reliability were higher for the mill-level method than for the industry-level method. For all the psychosocial variables the reliability for the mill-level method was greater than 0.90. The predictive validity results were inconclusive.

Conclusions The greater reliability and concurrent validity of the mill-level method indicates that panels of experienced workers should be considered as potential experts in future studies measuring psychosocial work conditions.

Key terms expert assessment, psychosocial work conditions, reliability, sawmills, validity.

Most studies of job strain rely on direct self-reports of psychosocial work conditions (1—5) or self-reports pooled within occupations (6—9). However, associations between self-reports of work conditions and self-reported health status may be biased due to common methods variance, whereby people who report health symptoms have a higher probability of reporting exposures than people who do not (10). In spite of repeated calls for the development of expert methods for the assessment of psychosocial work conditions to remedy this problem (11—13), very few investigations have used expert raters (14—18).

The purpose of this study was to test the reliability and validity of 2 expert methods for measuring psychosocial work conditions in British Columbia sawmills using the demand-control model (19). In the 1st method, 4 sawmill job evaluators estimated psychosocial work conditions at a generic sawmill (industry-level expert method). In the 2nd method (mill-level expert method), panels of experienced sawmill workers were randomly selected at 3 sawmills representing the typical technological range in the industry and asked to estimate the psychosocial work conditions in each mill.

Job-title specific scores for psychosocial work conditions were developed using both expert methods. These scores were then applied to the job titles used in an already gathered sawmill worker data set containing self-reported health status and heart disease outcomes. The reliability and validity of both expert methods were determined.

Subjects and methods

Over the past decade we have gathered a cohort of approximately 28 000 British Columbia sawmill workers who worked in 1 of 14 sawmills between 1950 and 1998 (20). As part of a study on technological change a subcohort of 9806 workers was identified. These workers were working in the industry in 1979. Three thousand

1 Department of Health Care and Epidemiology, Vancouver, Canada.
2 Scientific Research Director, Institute of Work and Health, Toronto, Canada.

Reprint requests to: Aleck Ostry, Department of Health Care and Epidemiology, 5804 Fairview Avenue, Vancouver BC, V6T 1Z3, Canada. [E-mail: ostry@interchange.ubc.ca]

workers randomly selected from this subcohort were approached for an interview between October 1997 and March 1999 (21).

A shortened version of the demand-control instrument was used to obtain self-report scores for control, psychological and physical demand, co-worker social support, and noise for the workers' jobs at the time of the interview (21). Self-reports of current health status (on a 5-point scale from excellent to bad) were obtained, as were self-reports of heart disease (yes or no, during the preceding 6 months). Basic demographic and lifestyle information was also obtained, as was each worker's job title at the time of the interview.

In the data set, 66 unique sawmill job titles were identified among the respondents. In order to obtain stable estimations by pooling self-reports within the job titles, we decided that a minimum of 5 self-report scores were necessary per job title. Thirty-one of the 66 job titles were held by ≤4 respondents and so were not used to calculate pooled self-report scores, the result being 35 job titles for which pooled scores could be calculated. Because 7 of these 35 job titles with pooled scores were not estimated with the expert rating method, a final list of 28 job titles was developed for which self-report, pooled self-report, and mill- and industry-level expert scores were available. All the analyses were based on these 28 sawmill job titles.

Of the 3000 workers sampled, 2156 complete responses were obtained for a response rate of 71.9%. Of these 2156 respondents, 270 had filled out a shortened version of the questionnaire with no psychosocial job strain data, leaving 1886 questionnaires with complete data. Seven hundred and eleven of these workers were employed at a study mill at the time of the interview. Of these, 61 workers had experienced unemployment during the year preceding the interview and were excluded from the analysis. The elimination of the respondents for which pooled self-reports were not available (ie, respondents with uncommon job titles) resulted in a final sample of 408 workers.

Obtaining experienced worker estimates of psychosocial work conditions (mill-level expert method)

Three of the 14 sawmills in the cohort study, representing the current range of sawmilling technology, were selected. A list of 54 basic job titles in the sawmill industry was developed using panels of experienced sawmill workers. Using seniority lists, 15 workers with more than 20 year's experience were randomly selected at all 3 sawmills. A single interviewer conducted face-to-face interviews with these workers during 1997 using the modified demand-control instrument.

The 10 best interviews were selected from each of the 3 study mills. The selection criterion was based on the number of job titles estimated. Any rater who was unable to provide an estimate for ≥10% of the job titles was excluded. In 2 of the 3 sawmills, 10 raters were able to estimate ≥90% of the job titles. In the 3rd mill, 12 raters estimated ≥90% of the job titles. Thus, 2 raters were randomly selected and eliminated from the panel in order to obtain a pool of 10 raters at each sawmill. Estimates for the 30 sawmill workers were pooled for each job title.

Obtaining estimates from job evaluators (industry-level expert method)

A union-management system of job evaluation has been in place in the British Columbia sawmill industry since the late 1960s. This job-evaluation system relies on an instrument developed within the industry to measure psychological and physical demand, control over skill use, control over decision making, and physical and other hazards. The expert evaluators were therefore familiar with the dimensions used in the demand-control model and with applying them to assessments of sawmill job titles.

Since the 1960s, a altogether 6 job evaluators have been employed in the sawmill job-evaluation program. These 6 job evaluators, with over 20 years' experience in sawmill job evaluation in British Columbia, were potential interviewees. Three were currently employed by industry, 1 was currently employed by the union, and 2 were recently retired from the union. All 3 industry raters agreed to participate, as did the currently employed union expert. One of the retired union experts was too ill to participate, and the other refused without giving a reason.

Using face-to-face interviews, these 4 expert raters were asked to rate a generic sawmill for current (1997) psychosocial work conditions using the modified demand-control instrument. A total of 54 job titles was estimated.

Reliability analyses

Two measures were used to assess reliability, the intraclass correlation coefficient (ICC) between individual raters and the ICC for all raters (22, 23). The ICC for individual raters (individual ICC) measures how an individual rater's estimate of exposure compares with that of the other raters. However, the purpose of this study was not to assess individual raters, but rather to determine whether or not the group's mean estimate of exposure was reliable. That is, had we selected different groups of raters, would the group means of their estimates have been reliable? The measure that best estimates this form of reliability is the ICC for all raters (group ICC) (23, p 273).

For the mill-level method, individual and group ICC values were calculated for the entire group of

30 workers across the 3 mills. For the industry-level method, individual and group ICC values were calculated for the 4 job evaluators. As well, the proportion of variance distributed between raters, job titles, and the residual was calculated.

Validity analyses

Pooled self-reports and mill- and industry-level expert scores were imputed for each individual on the basis of the job title so that each of the 408 persons in the study had self-report, pooled self-report, mill-level, and industry-level scores for each of the 5 psychosocial work conditions. As a measure of concurrent validity, the agreement between the 4 methods was assessed by calculating pair-wise Pearson correlation coefficients.

As a measure of predictive validity, a logistic regression was conducted using the 4 different methods with self-report health status and heart disease as the outcome variables. These health outcomes were chosen since other studies have shown that psychosocial work conditions are associated with them. Self-reported heart disease has been associated with control, psychological demand, co-worker social support, physical demand, and noise in several studies (4, 17, 24—25). Self-reported health status has been linked with control (26).

Logistic regression was used to model self-reported health status (dichotomized to 0 = excellent or good; 1 = fair, poor, or bad) and self-reported heart disease (dichotomized as 0 = absent and 1 = present) as the outcome variables. The sociodemographic variables age, education, housing status, income, and country of birth were modeled with each outcome. Controlling for sociodemographic variables, multivariate models were conducted with respect to control, psychological demand, physical demand, noise, and co-worker social support for each of the 4 assessments.

Table 1. Interrater reliability across 28 job titles for 4 industry-level experts. (Ps = psychological demand, Ph = physical demand, Social = co-worker social support, ICC = intraclass correlation coefficient)

Source of variance	Control	Ps	Ph	Social	Noise
Between job titles (%)	65.9	46.8	53.6	67.3	58.9
Between raters (%)	0.9	8.4	1.4	0.2	4.3
Residual (%)	33.2	44.8	45.0	32.5	36.8
Individual ICC	0.55	0.32	0.39	0.57	0.47
Group ICC	0.83	0.68	0.72	0.84	0.79

Table 2. Interrater reliability for 28 job titles with 30 mill-level experts across 3 sawmills. (Ps = psychological demand, Ph = physical demand, Social = co-worker social support, ICC = intraclass correlation coefficient)

Source of variance	Control	Ps	Ph	Social	Noise
Between job titles	20.1	15.4	26.5	33.9	28.2
Between raters	48.8	43.6	23.7	16.7	17.6
Residual	31.1	41.0	49.8	49.4	54.2
Individual ICC	0.42	0.33	0.50	0.60	0.53
Group ICC	0.95	0.91	0.94	0.95	0.93

Table 3. Pairwise pearson correlation coefficients for 4 methods of estimating psychosocial conditions of work for 28 sawmill job titles.[a] (Ps = psychological demand, Ph = physical demand, Social = co-worker social support, vs = versus)

	Control	Ps	Ph	Social	Noise
Industry vs mill level	0.60	0.73	0.56	0.79	0.78
Industry vs pooled	0.57	0.44	0.47	0.59	0.71
Mill-level vs pooled	0.89	0.55	0.84	0.83	0.86
Industry vs self-reports	0.32	0.13	0.23	0.32	0.36
Mill level vs self-reports	0.53	0.14	0.41	0.45	0.42
Pooled vs self-reports	0.58	0.25	0.48	0.54	0.50

[a] For all variables P≤0.01.

Results

Reliability

Table 1 shows the reliability of the 4 job evaluators for the industry-level expert method. The group ICC values were highest for control (0.83) and social support (0.84) and lowest for psychological demand.

Table 2 shows that, when the estimates for 30 experienced workers in 3 sawmills were pooled, the group ICC values for all 5 variables were greater than those obtained with the industry-level method. In particular, the group ICC scores for the demand variables (physical and psychological demand) were increased by approximately 20% with the mill-level method.

Concurrent validity

As a measure of concurrent validity, the agreement between the 4 methods was assessed by calculating pairwise Pearson correlation coefficients (table 3). For all 5 psychosocial work condition variables, the level of agreement between the direct self-reports decreased, moving from pooled to mill-level, and to industry-level estimation methods. Similarly, the level of agreement between the pooled self-reports decreased, moving from mill-level to the industry-level expert method. The greatest agreement between methods was observed for the pooled self-reports and the mill-level method for all the variables except psychological demand. The greatest agreement between the methods was obtained for noise. The agreement of the methods for psychological demand was strikingly lower than for the other psychosocial work condition variables.

Predictive validity

Table 4 presents the odds ratios and confidence intervals for reported heart disease. Self-reported noise and psychological demand were both associated with a

Table 4. Logistic regression between 4 methods of assessing psychosocial work conditions for self-reported heart disease.[a] (OR = odds ratio, 95% CI = 95% confidence interval)

Variable	Direct self-report		Pooled self-report		Mill level		Industry level	
	OR	95% CI	OR	95% CI	OR	95% CI	OR	95% CI
Noise	1.92*	2.55—1.28[b]	1.40	0.31—2.49	1.41	0.21—3.00	1.65	0.87—2.43
Control	1.04	0.94—1.14	0.92	0.75—1.08	0.94	0.71—1.16	1.00	0.88—1.12
Physical demand	1.59	0.98—2.09	1.10	0.16—2.03	1.28	0.41—2.14	1.24	0.43—2.05
Psychological demand	1.16**	1.00—1.31[b]	1.73	0.98—2.18	1.38	0.93—1.75	1.23	0.89—1.57
Co-worker social support	1.25	0.78—1.71	0.67	0.17—1.52	0.86	0.39—1.334	0.98	0.66—1.30

[a] Outcomes were coded so that 1 = presence of condition or service and 0 = absence.
[b] Confidence intervals do not cross 1.00.
*P=0.01, **P=0.05

Table 5. Logistic regression between 4 methods of assessing psychosocial work conditions for self-reported health status.[a]

Health outcome	Direct self-report		Pooled self-report		Mill level		Industry level	
	OR	95% CI	OR	95% CI	OR	95% CI	OR	95% CI
Poor self-rated health	Control 0.91	0.85—0.96[b]	Control 0.91	0.81—1.00[b]	Control 0.89	0.76—1.02	Control 0.98	0.92—1.05

[a] Outcomes were coded so that 1 = presence of condition or service and 0 = absence.
[b] Confidence intervals do not cross 1.00.

higher incidence of heart disease. Increased self-reported physical demand was also associated with a higher incidence of heart disease (P=0.06). The 3 other methods for measuring both noise and psychological demand showed weak associations with heart disease.

With the logistic regression with self-reported health status, control was the only psychosocial work condition variable which showed any association (table 5). Self-reports for control were associated with poor self-rated health in the expected direction (ie, self-rated health improved with increasing control). Pooled self-reports for control showed a weaker association between self-reported control and health status in the expected direction. Finally, the mill-level method for measuring control indicated a similar effect size for control and self-rated health status (P=0.07).

Discussion

When the industry and mill-level methods were compared (by pooling 30 workers' scores in the latter method), reliability scores, as measured by the group ICC were higher for the mill-level method. The reliabilities with the mill-level method were greater than 0.90 for all the psychosocial work conditions, which, according to Nunnally (27), is the minimum standard for reliability. The industry-level method was not able to attain this basic level of reliability.

The concurrent validity results showed correlations of 0.32 between self-reported control and control estimated with the industry-level expert method. In the Whitehall Study of British civil servants, for men and women, correlations between self-reports for control and estimations of control conducted by personnel managers ranged from 0.25 and 0.35 (17). These correlations were similar to those obtained in this investigation using the industry-level expert method.

The correlations between the self-reports of control and the mill-level expert estimation were higher (0.53) than the correlations between the self-reports and the industry-level method, and the correlations found between the self-reports and the expert method used in the Whitehall investigation. The greater reliability and concurrent validity of the mill-level method indicates that panels of experienced workers should be considered as potential experts in future studies measuring psychosocial work conditions

The predictive validity results were difficult to interpret. They showed that the self-reports of psychological demand and noise were associated with reported heart disease in the expected direction. In addition, the self-reports of control were associated with reported health status, all in the expected direction. However, both expert rating methods showed weaker associations with health outcomes than the direct and pooled self-report methods did.

There are other reasons for expecting a weaker association between an expert assessment of psychosocial work conditions and health status. For example, the expert assessment of psychosocial work conditions involves cognitive and emotional processing (on the part

of the experts), which will tend to attenuate associations between assessed work conditions and health outcomes (18). As well, associations for the industry-level method should be weaker than for the mill-level method because estimations were conducted at arm's length and were based on a generic sawmill rather than on actual sawmills. The weaker associations using the expert method in comparison with the self-report method in this study are to be expected due to this bias.

There were several other limitations to this study. Because self-reported heart disease was relatively rare in this sample, the statistical power of the regression analyses may not be sufficient to detect associations when they could be present. In addition, there may be issues of measurement reliability for the 2 outcome measures used in the regression analyses.

In spite of these limitations, this study demonstrated that the mill-level method was reliable. The reliability scores (across all the psychosocial work condition variables) for the mill-level method were approximately 20% greater than for the industry-level method. Improvement in reliability with the mill-level method was particularly noteworthy in the case of psychological and physical demand.

In addition, the mill-level method showed greater concurrent validity than the industry-level method. The much higher correlation in this study, between self-reported control and the mill-level method of measuring control (0.53), compared with the correlations between the self-reported and expert-assessed control in the Whitehall Study (0.32), also point to the utility of using panels of experienced workers to estimate psychosocial work conditions.

Acknowledgments

The British Columbia Health Research Foundation, the Center for Health Services and Policy Research at the University of British Columbia, Forest Renewal, British Columbia, Institute of Work and Health (Toronto), and the Canadian Institute of Advanced Research (Toronto), and the Canadian Institute for Health Research supported this study.

References

1. Haan MN. Job strain and ischaemic heart disease: an epidemiological study of metal workers. Ann Clin Res 1988; 20:143—5.
2. Johnson J, Hall E. Job strain, work place social support, and cardiovascular disease: a cross-sectional study of a random sample of the Swedish working population. Am J Public Health 1988;78:1336—42.
3. Johnson J, Hall E, Theorell T. Combined effects of job strain and social isolation on cardiovascular disease morbidity and mortality in a random sample of the Swedish male working population. Scand J Work Environ Health 1989;15:271—9.
4. Karasek R, Baker D, Marxer F, Ahlbom A, Theorell T. Job decision latitude, job demands, and cardiovascular disease: a prospective study of Swedish men. Am J Public Health 1981;71:694—705.
5. Schnall P, Pieper C, Schwartz J, Karasek R, Schlussel Y, Devereux R, et al. The relationship between job strain, workplace diastolic blood pressure, and left ventricular mass index. J Am Med Assoc 1990;263:1929—35.
6. Karasek R, Theorell T, Schwartz J, Schnall P, Pieper C, Michela J. Job characteristics in relation to the prevalence of myocardial infarction in the US health examination survey (HES) and the health and nutrition examination survey (HANES). Am J Public Health 1988;78:910—8.
7. Pieper C, Lacroix A, Karasek R. The relation of psychosocial dimensions of work with coronary heart disease risk factors: a meta-analysis of 5 United States data bases. Am J Epidemiol 1989;129:483—94.
8. Theorell T, de Faire U, Johnson J, Hall E, Perski A, Stewart W. Job strain and ambulatory blood pressure profiles. Scand J Work Environ Health 1991;17:380—5.
9. Baneryd K. Analysis of the psychological job content in sawmill work. Stockholm: National Board of Occupational Safety and Health, 1974.
10. Kristensen T. Job stress and cardiovascular disease: a theoretic critical review. J Occup Health Psychol 1996;1:246—60.
11. Kasl S. Epidemiological contributions to the study of work stress. In: Cooper C, Payne R, editors. Stress at work. New York (NY): John Wiley & Sons, 1978:3—48.
12. Kasl S. Methodologies in stress and health: past difficulties, present dilemmas, future directions. In: Kasl S, Cooper C, editors. Research methods in stress and health psychology. New York (NY): John Wiley & Sons, 1987.
13. Hertzman C, Teschke K, Dimich-Ward H, Ostry A. Validity and reliability of a method for retrospective evaluation of chlorophenate exposure in the lumber industry. Am J Ind Med 1988;14:703—13.
14. Johannson G, Aronsson G, Lindstrom B. Social psychological and neuroendocrine stress reactions in highly mechanised work. Ergonomics 1978;21:583—99.
15. North F, Syme L, Feeney A, Shipley M, Marmot M. Psychosocial work environment and sickness absence among British civil servants: the Whitehall II study. Am J Public Health 1996;86:332—40.
16. Stansfeld S, Rael E, Head J, Shipley M, Marmot M. Social support and psychiatric sickness absence: a prospective study of British civil servants. Psychol Med 1997;27:35—48.
17. Bosma H, Marmot M, Hemingway H, Nicholson A, Brunner E, Stansfeld S. Low job control and risk of coronary heart disease in Whitehall II (prospective cohort) study. BMJ 1997;314:558—65.
18. Frese M, Zapf D. Methodological issues in the study of work stress: objective vs subjective measurement of work stress and the question of longitudinal studies. In: Cooper C, Payne R, editors. Causes, coping, and consequences of stress at work. New York: John Wiley & Sons Ltd, 1988:375—411.
19. Karasek R, Gordon G, Pietrokovsky C, Frese M, Pieper C, Schwartz J, et al. Job content instrument: questionnaire and users' guide. Lowell (MA): University of Massachusetts,

1985.
20. Hertzman C, Teschke K, Ostry A, Dimich-Ward H, Kelly S, Spinelli J, et al. Cancer incidence and mortality in sawmill workers exposed to chlorophenols. Am J Public Health 1996;87:71—9.
21. Ostry A, Marion S, Green L, Demers P,Teschke K, Hershler R, Kelly S, et al. Downsizing and industrial restucturing in relation to changes in psychosocial conditions of work in British Columbia sawmills. Scand J Work Environ Health 2000;26(3):273—7.
22. Winer B. Statistical principles in experimental design. 2nd ed. Montreal: McGraw-Hill, 1971.
23. Armitage P, Berry G. Statistical methods in medical research. London: Blackwell Scientific Publications, 1994:273.
24. House J, McMichael A, Kaplan B, Wells J, Landerman L. Occupational stress and health among factory workers. J Health Soc Behav 1979;20:139—60.
25. Karasek R, Theorell T. Healthy work. New York (NY): Basic Books, 1990.
26. Spector P, Jex S. Relations of job characteristics from multiple data sources with employee affect, absence, turnover intentions, and health. J Appl Psychol 1991;76:46—53.
27. Nunnally J. Psychometric theory, New York (NY): McGraw-Hill, Inc, 1978.

Received for publication: 3 May 2000