ELSEVIER

Applied Ergonomics 38 (2007) 167–175

APPLIED ERGONOMICS

www.elsevier.com/locate/apergo

Medium- and long-term reproducibility of self-reported exposure to physical ergonomics factors at work

Angelo d'Errico[a,b,*], Rebecca Gore[a], Judith E. Gold[a,c], Jung-Soon Park[a], Laura Punnett[a]

[a]Department of Work Environment, University of Massachusetts Lowell, Lowell, MA, USA
[b]Epidemiology Unit, Piedmont Region, Grugliasco (TO), Italy
[c]Department SEQ CHAPTER of Public Health, Temple University, Philadelphia, PA, USA

Received 9 November 2004; accepted 17 March 2006

Abstract

Introduction: The literature is sparse on reproducibility of self-reported exposure to physical ergonomics risk factors for musculoskeletal disorders (MSDs). Aims of this study were to evaluate, in a cohort of workers interviewed up to three times: 1-year test–retest reliability; and 5- and 6-year recall of physical exposures. We also examined whether reproducibility was influenced by the presence of UE MSD or by technological changes introduced between the last two surveys.

Methods: A cohort of automobile manufacturing employees was interviewed at baseline, one and six years later about work history, physical and psychosocial exposures at work, upper limb symptoms, injury and medical history, and demographics. Agreement between interviews was evaluated by intraclass correlation and Spearman coefficients. Differences in exposure between 1- and 6-year follow-up were analyzed by Wilcoxon matched-pairs signed-ranks test.

Results: Large and significant decreases in work pace and physical effort were observed from baseline, although an upper extremity composite index was quite stable in the total population. One-year test–retest reliability was fair to good for the composite exposure index ($ICC = 0.58$), whole-body vibration, handling parts, and tool use, but poor for the other variables considered. Long-term reproducibility, from baseline or 1-year follow-up to 6-year follow-up, was poor for the composite index and almost all single items. UE MSD case status influenced 1-year test–retest reliability, with subjects who changed case status from baseline displaying higher reliability, but not reproducibility of recalled exposures. A strong regression to the mean effect was observed on exposures reported at follow-up surveys.

Conclusions: Recalled ergonomics exposures could be employed in retrospective cohort studies as a somewhat reliable and unbiased estimate of the self-reported exposures that would have been obtained up to one year earlier, but not over a longer period (5–6 years). These longer-term results may have been limited by difficulty in matching jobs between interviews; also the regression to the mean effect likely contributed to reduce agreement. Changes in production technology and work organization produced a decrease in physical workload intensity and job pace, but did not have a substantial impact on an exposure index for the upper limb.

Keywords: Reproducibility; Automobile manufacturing; Risk factors; Exposure recall

1. Introduction

There is substantial epidemiologic evidence of associations between physical ergonomics exposures at the workplace, such as lifting, constrained postures, repetitive movements, fast work pace, heavy material manual handling, forceful exertions and vibration, and the occurrence of upper extremity musculoskeletal disorders (UE MSD) (for reviews see e.g. Bernard, 1997; Grieco et al., 1998; Hagberg et al., 1995; National Research Council and Institute of Medicine, 2001; van der Windt et al., 2000). Nevertheless, these findings are still criticized, in part because many of these studies relied upon assessment of exposure to physical loads through worker self-report. One particular concern is whether there is

*Corresponding author. Epidemiology Unit, Piedmont Region, Via Sabaudia 164, Grugliasco (TO) 10095, Italy. Tel.: +39 011 40188335; fax: +39 011 40188201.
E-mail address: angelo.derrico@epi.piemonte.it (A. d'Errico).

0003-6870/$ - see front matter
doi:10.1016/j.apergo.2006.03.002

differential misclassification of self-reported exposure between MSD cases and non-cases.

However, more objective methods, such as observations or direct measurements, are often too time-consuming and expensive for use in epidemiologic studies (Kilbom, 1994). Moreover, observations or direct measurements of occupational exposures can be employed only in cross-sectional or prospective studies, whereas workers often represent the only available source of retrospective information, except for indirect assessment through tasks and job titles, whose validity is supposed to be too low for quantitative assessment (Balogh et al., 2001; Burdorf, 1992; Winkel and Mathiassen, 1994).

It is also worth noting that observations and direct measurements can be regarded as true gold standards of exposure to physical stressors only if the observation or measurement period is representative of the actual average of workers' exposure (Kilbom, 1994). A long observation time may be necessary for workers in non-routine jobs who have many variable tasks, and thus need to be observed or measured for days to obtain a valid exposure assessment. Thus, self-reports often remain a component of these analyses, as well, at the least for preliminary inventory of tasks and their exposure profiles.

There has been some examination of the reliability or reproducibility of self-reported exposure to physical ergonomics exposures (Ainsworth et al., 1993; Balogh et al., 2001; Gamberale, 1988; Halpern et al., 2001; Koster et al., 1999; Leijon et al., 2002; Spielholz et al., 1999; Torgén et al., 1997, 1999; Wiktorin et al., 1996a, b, 1999; Yeung et al., 2002). However, most of these studies were small and most involved short-term recall, i.e., one year or much less. In general, self-reports on different occasions have shown moderate to good agreement over the short term for factors related to whole body activity, such as sitting, standing, walking, and lifting (Stock et al., 2005). Fewer investigations have focused on exposures at specific UE regions, and these have produced varying results, so it is difficult still to draw overall conclusions from the extant literature.

This study takes advantage of a cohort of automobile manufacturing workers who rated their ergonomics exposures at work on three different occasions (1992, 1993 and 1998) through interviewer-administered questionnaires. The selected physical job features, such as work pace, postures, and forceful exertions, were chosen because they were under study as risk factors for UE MSDs in this cohort. Exposure ratings were obtained in each interview for the contemporary job. In addition, in the last interview, physical exposure ratings were also obtained for all other jobs held back to the year of the baseline interview.

Between the second and third interviews, substantial engineering and organizational changes took place in the factories under study, and many subjects had voluntary or involuntary changes in job title. Thus, it was of interest to evaluate potential changes in physical exposures within the cohort, resulting both from the evolution of production technology and from job transfers by individual workers. For this reason, the objectives of the study were as following:

- to describe, at both the group and individual level, changes across surveys in self-reported exposure to physical ergonomics exposures (under study here as risk factors for MSDs);
- to evaluate 1-year test–retest reliability of reporting, in subjects holding the same job title as at baseline;
- to assess long-term reproducibility (5 and 6 years) of recalled exposures to physical risk factors; and
- to evaluate the influence of UE MSD symptoms and other covariates on test–retest reliability and reproducibility of exposure ratings.

2. Materials and methods

The present study is part of a prospective epidemiologic study on work-related UE MSDs among automobile manufacturing workers in a stamping and an engine plant. Workers were interviewed at baseline ("T0"), 1-year ("T1") and 6-year follow-up ("T2") (Gold et al., 2006; Punnett, 1998; Punnett et al., 2004). Important engineering changes occurred in these two factories during the follow-up period, particularly in the stamping plant. In particular, production technology became more automated from 1993 to 1998, although the overall work flow in each plant remained organized around a moving production line. For example, manual part transfer between pairs of stamping presses was replaced by automated mechanical part transfer. There was more job rotation, but no formal teamwork or manufacturing cells were introduced. Thus, we anticipated reductions in physical exposures such as whole body effort, lifting and grip forces, but not in work pace.

2.1. Data collection

The baseline survey (1992–1993) covered 1283 subjects employed in an engine plant or a stamping plant of the same corporation. Interviews of approximately 45 min were conducted by a trained interviewer away from the production line. The standardized questionnaire, based on previously published instruments to the extent possible (Punnett, 1998; Punnett et al., 2004), covered work history, current exposure to physical ergonomics exposures at work, presence of upper extremity MSD symptoms, previous injuries, lifestyle habits (alcohol and tobacco consumption, leisure activity), relevant medical conditions and demographic data.

UE MSD cases were defined by symptoms and by physical examination findings plus symptoms at each of four body areas: neck, shoulder, elbow, and wrist/hand. Participants with prevalent UE MSD symptoms were also asked about pain severity in that region over three time periods (summed to create a pain severity index) and about difficulty performing activities of daily living requiring upper extremity function (10 items summed to create a functional impairment index).

For the physical exposure ratings, subjects were asked to rate their usual job on a Borg scale from 0 to 10 (Borg, 1990) for awkward postures of back, neck, shoulder, elbow and wrist, whole-body and segmental vibration, work pace, manual force exerted in handling parts and tools, and mechanical pressure from hand held tools. A composite exposure index was created for each survey by summing scores for job pace, features of parts handled and tools used, pressure from tool handle, whole-body and segmental vibration (Punnett, 1998; Punnett and van der Beek, 2000).

All workers enrolled in the study also underwent physical examination to assess the presence of UE MSD.

One year later, during 1993–1994, 790 workers still employed in the same company were re-examined. The interview covered most of the same items asked in the baseline questionnaire, with some changes in the questions on physical exposures. In the second follow-up interview (1998) on 519 members of the cohort, the physical exposure ratings were worded identically to those in the baseline questionnaire. On this occasion ratings were obtained both for the current job and for all jobs held since 1992. In the 1- and 6-year follow-up interviews, information on psychosocial exposures at the workplace was collected through the Job Content Questionnaire (Karasek, 1985).

At the 6-year follow-up survey, shop-floor observations and discussions with supervisors and workers were used to categorize each job as "high routinization," if it consisted entirely of one or more fixed-cycle tasks, or "low routinization" otherwise (Gold et al., 2006). The same criteria were applied retrospectively, to the extent possible, to the jobs as they existed at baseline.

2.2. Statistical analysis

Changes in ergonomics exposures were evaluated at the group level, by comparing mean UE exposure index at baseline, T1 and T2, and testing differences in exposure between each survey by Mann–Withney statistics. Since a substantial proportion of the baseline cohort was not interviewed at T1 or T2, workers lost at follow-up were compared for the composite exposure index with those who were subsequently interviewed. Differences in exposure between T1 and T2, i.e., when most workplace transformations occurred, were also analyzed at the individual level by Wilcoxon matched-pairs signed-ranks test. This analysis was stratified by job change, to evaluate the variation in exposure separately in workers who remained in the same job at T2, therefore mainly undergoing technological changes, and in those who changed job, who endured also changes in work tasks and organization.

Since variations in self-reported physical exposures could theoretically be due to changes in workers' attitudes, caused by changes in work organization and work climate, we also examined the correlation between the average difference of the UE composite exposure index between T1 and T2 and the self-reported level of psychosocial strain at T2, computed by standard JCQ algorithms (Karasek, 1985).

Test–retest reliability of exposure over the medium-term (0.5–1 year) was estimated by comparing workers' ratings of physical exposures at 1-year follow-up with those reported by the same subject at baseline, restricting the analysis to subjects who did not report a change in job title or job content between the first two surveys (507 workers).

Long-term reproducibility was examined only in subjects who changed job title between T0 or T1 and T2, since they were asked to rate exposures regarding the job that they were performing at the time of a previous interview. Job history reported at T0 or T1 was reviewed manually to examine if jobs recalled at T2 matched with those previously reported. After exclusion of subjects with inconsistent information on job history or with missing values in most of the items on physical exposure, 163 workers were available for analysis of reproducibility between the T0–T2 surveys and 162 subjects for T1–T2.

Test–retest reliability and reproducibility were evaluated by means of the intraclass correlation coefficient (ICC), computed using a random-effect one-way analysis of variance (Armstrong et al., 1992). Since ICC assumes normality of distribution and equal variances of the groups examined, and for many of these exposures such criteria were not satisfied, we also computed Spearman's ρ coefficient, in order to evaluate possible distortions on ICC estimates.

Agreement between categorical variables (handling parts and use of tools) was assessed by simple kappa statistic, with confidence intervals computed according to Fleiss (1981).

Agreement of the composite exposure index between interviews was also computed by UE MSD case status, defined by symptoms or by physical examination plus symptoms. One-year test–retest reliability analysis was stratified by change in UE MSD status, assuming that those with no change in case status would have a higher reliability of reporting. In the long-term reproducibility analysis the case status at the most recent interview was taken into account, in order to assess if differential recall of past exposures was present between cases and non-cases. To test for significance of the difference between ICCs in each category, a t-test for groups with unequal variance was employed, with degrees of freedom computed according to Satterthwaite's formula (Satterthwaite, 1946).

Analyses were performed using SAS (version 8) and Stata (version 8).

3. Results

Of the original 1283 study participants, about 30% had retired or otherwise left employment by the time of the second follow-up interview. In general, the characteristics

Table 1
Means of UE composite exposure index and proportion of subjects in highly routinized jobs at baseline and follow-ups in the whole cohort, by plant, department and routinization category

GROUP	Baseline				1-year follow-up			6-year follow-up			
	n	Mean	sd	High routinization (%)	*n*	Mean	sd	*n*	Mean	sd	High routinization (%)
Overall	1151[a]	6.9	3.0	89.4	760	6.8	2.7	519	6.3	2.9	38.9
Plant											
Stamping plant	593	7.9	3.0	99.3	413	7.3	2.8	299	6.6	2.8	43.1
Engine plant I	558	5.8	2.6	78.9	347	6.2	2.5	200	5.8	3.0	28.5
Engine plant II	—	—	—	—	—	—	—	20	6.5	2.5	80.0
Department											
Nonroutine (EP[b])	20	4.4	3.5	0.0	34	6.9	2.2	116	5.9	3.0	1.7
Machining (EP)	260	5.6	2.6	61.9	180	6.2	2.5	135	6.0	3.0	41.5
Assembly/Subassy (EP)	281	6.1	2.6	99.3	155	6.2	2.7	34	6.2	2.3	50.0
Large Press (SP[c])	198	8.0	2.5	99.5	108	7.5	2.2	72	7.0	2.6	58.3
Small Press (SP)	156	9.9	3.0	100.0	79	9.1	3.1	20	7.8	3.0	70.0
Assembly/Welding (SP)	236	6.5	2.6	100.0	204	6.5	2.7	142	6.3	2.9	50.0
Job routinization											
High	1029	7.1	3.0		679	6.9	2.7	202	6.3	2.8	
Low	122	5.1	2.7		81	6.1	2.6	317	6.3	3.0	

[a]Subjects with missing values in most physical exposures were excluded ($n = 132$).
[b]Engine plant.
[c]Stamping plant.

of the study population were fairly constant across the three interviews; at each survey the participants were mainly male, older than 40 years (mean age: 46.5 years at baseline; 47.3 years at 1-year; 50.2 years at 6-year follow-up), and African-American. The prevalence of upper extremity MSDs in the cohort, defined on the basis of symptoms, was above 30% in the baseline and 1-year interviews, decreasing to about 17% at 6-year follow-up. The proportion of cases based on physical examination decreased less markedly (from 28% at baseline to around 20% at 6-year follow-up).

From baseline to 6-year follow-up the proportion of workers employed in jobs with high level of routinization decreased from 90% to approximately 40% (Table 1). Such a decrease was evident in both plants and all departments except the "non-routine" group: tool room, forklift drivers, etc.

In Table 1 are shown the average values of the composite UE exposure index at baseline (T0), 1-year (T1) and 6-year follow-up (T2), for the whole cohort, and by plant, department group, and job routinization category. The overall means indicated a progressive decrease of the index from baseline to 6-year follow-up, which occurred entirely in the large and small press departments of the stamping plant. In the non-routine departments, a large increase in exposure was seen between T0 and T1, followed by a smaller decrease at T2. Similarly, the decrease in exposure concerned only subjects employed in jobs with a high level of routinization, while those in low-routinization jobs reported an increase at follow-up.

Analyzing differences at the individual level, in workers in the same job title at T0 and T1 the UE composite exposure index showed only a small increase at T1 ($p = 0.11$), corresponding to approximately 2% of the average value of the index at baseline. This observation would demonstrate that, as expected, exposure to ergonomics factors was quite stable in these workers. Most specific exposure items showed small variations in both directions, too, generally smaller than 10% of the average score at T0, except for a few (whole-body and segmental vibration, balance of tools, mechanical pressure through tool handle).

Differences in the mean exposure index between T1 and T2 among workers present at both interviews were also quite small and non-significant: a slight decrease in the mean exposure index was observed, equal to a 2.6% average decrease, compared to T1.

Among workers in the same job, the mean exposure index increased very slightly (approximately 1% of the index at T1), whereas larger and significant decreases, ranging from 13% to 33%, were noted for job pace, physical effort, pinch force, balance of tools, whole-body vibration, and weight of typical and heaviest parts handled; in this group of workers the only significant increases in exposure concerned segmental vibration and mechanical pressure through the tool handle (Table 4). In the subgroup of workers who changed job between surveys the decrease of the exposure index was larger, although non-significant; similarly, stronger reductions of job pace, physical effort and pinch force were observed, compared to subjects still in the same job title.

Exposure to high psychosocial strain at T2 was positively correlated with the average difference of the exposure index between surveys ($r = 0.10$, $p = 0.05$). Among subjects in the same job at T1 and T2 the correlation was higher ($r = 0.15$, $p = 0.04$), whereas in workers who changed job there was no such association ($r = 0.04$, $p = 0.58$).

In all comparisons, individual differences in the exposure index were strongly inversely correlated with the corresponding index at the previous survey (T0–T1: $r = -0.54$; T0–T2: $r = -0.74$; T1–T2: $r = -0.66$), regardless of the occurrence of a job change, suggesting the presence of a regression to the mean effect. Differences in exposure were also analyzed by quartile of the UE exposure index at a previous interview: workers previously in the lowest exposure quartile reported, on average, a strong increase at follow-up and those in the highest quartile a comparable decrease, whereas differences were smaller among workers in the middle quartiles of exposure (data not shown).

The average exposure index at baseline was not significantly different between workers who were followed-up or not at T1 (13.2 vs. 12.9, respectively). Similarly, no significant difference was found in the mean index for subjects present at T1, but not at T2, compared to those participating in both surveys (13.2 vs. 12.8, respectively).

3.1. Medium-term test–retest reliability of self-reported exposures

The composite exposure index demonstrated fair test–retest reliability between baseline and 1-year follow-up (ICC = 0.58) (Table 2). Among the specific physical stressor ratings, only whole-body vibration intensity showed moderate agreement, whereas reliability of all other self-reported exposures was poor except for the two dichotomous exposures: handling parts ($k = 0.50$, 95% CI: 0.37–0.64) and use of tools ($k = 0.60$, 95% CI: 0.53–0.67).

The analysis stratified by change in UE MSD status (Table 3) showed higher reliability of the composite exposure index among subjects who became cases

Table 2
Intraclass correlation and Spearman ρ coefficients of self-reported physical exposures at baseline and at 1-year follow-up: workers reporting no change in job assignment or job content (507 subjects)

Exposure	ICC	LCL 95%	UCL 95%	Spearman ρ coeff.	Mean diff-erence[a]
Composite UE exposure index	0.58	0.52	0.63	0.58	0.13
Pace of the job	0.41	0.33	0.48	0.45	0.43
Physical effort	0.40	0.33	0.48	0.40	0.36
Weight of typical part handled	0.47	0.40	0.55	0.49	0.43
Weight of heaviest part handled	0.48	0.41	0.55	0.47	−0.09
Vibration through floor	0.56	0.50	0.62	0.59	0.50
Weight of tool	0.47	0.36	0.58	0.38	0.32
Balance of tool	0.30	0.17	0.43	0.30	0.57
Grip force to use tool	0.40	0.28	0.52	0.39	−0.22
Size of tool handle	0.13	0.00	0.28	0.18	0.32
Pressure of tool handle	0.41	0.29	0.53	0.44	−0.98
Vibration through tool handle	0.47	0.36	0.58	0.41	−0.35

[a]Computed as the average of individual differences between continuous exposures.

Table 3
ICC between UE composite exposure index at different times, stratified by change in UE MSD case status

Interviews compared	Job change	Stratification variable	Category (n)	ICC	LCL 95%	UCL 95%	p-value[a]
Baseline vs. 1-year follow-up	No	Change in UE MSD symptoms case status	Yes (116)	0.69	0.49	0.78	0.01
			No (385)	0.54	0.47	0.61	
		Change in UE MSD physical examination + symptoms case status	Yes (95)	0.71	0.61	0.81	0.005
			No (401)	0.54	0.47	0.61	
Baseline vs. 6-year follow-up	Yes	UE MSD symptoms case status at 6-year follow-up	Yes (53)	0.49	0.18	0.80	0.51
			No (110)	0.37	0.23	0.52	
		UE MSD physical examination + symptoms case status at 6-year follow-up	Yes (45)	0.28	0.00	0.58	0.22
			No (110)	0.48	0.35	0.62	
1-year vs. 6-year follow-up	Yes	UE MSD symptoms case status at 6-year follow-up	Yes (49)	0.38	0.05	0.71	0.91
			No (112)	0.36	0.21	0.50	
		UE MSD physical examination + symptoms case status at 6-year follow-up	Yes (34)	0.33	0.03	0.64	0.64
			No (125)	0.41	0.27	0.56	

[a]computed by t-test for groups with unequal variance.

Table 4
Mean differences of exposure scores between 1- and 6-year follow-up (378 subjects present at both interviews), by job change

Exposure	No job change ($n = 190$)		Job change ($n = 188$)	
	Mean difference[a]	p-value[b]	Mean difference[a]	p-value[b]
Composite UE exposure index	0.08	0.77	−0.33	0.25
Pace of the job	−1.25	<0.0001	−2.08	<0.0001
Physical effort	−1.52	<0.0001	−1.99	<0.0001
Vibration through floor	−0.84	0.03	0.05	0.85
Weight of typical part handled	−0.85	0.01	0.69	0.09
Weight of heaviest part handled	−0.80	0.02	0.30	0.54
Weight of tool	−0.05	0.50	0.20	0.98
Balance of tool	−1.14	0.009	−0.01	0.61
Grip force to handle and use tool	0.23	0.27	−0.37	0.49
Pressure of tool handle	1.27	0.02	1.24	0.01
Vibration through tool handle	1.06	0.13	0.93	0.51
Pinch force	−0.66	<0.0001	−1.03	<0.0001
Knee bending	0.04	0.64	0.29	0.01
Kneeling or squatting	0.18	0.15	0.16	0.36
Lifting objects ⩾10 pounds	−0.12	0.20	−0.11	0.27

[a]Computed as the average of individual differences between continuous exposures.
[b]Wilcoxon matched-pairs signed-ranks test.

Table 5
Intraclass correlation and Spearman ρ coefficients of physical exposures reported at T0 or at T1 and recalled at T2 by automobile manufacturing workers

Exposure	T0–T2 ($n = 163$)				T1–T2 ($n = 162$)			
	ICC	LCL 95%	UCL 95%	Spear-man ρ coeff.	ICC	LCL 95%	UCL 95%	Spear-man ρ coeff.
Composite UE exposure index	0.40	0.27	0.53	0.42	0.37	0.24	0.50	0.40
Pace of the job	0.11	0.00	0.26	0.15	0.32	0.18	0.46	0.36
Awkward back postures	0.25	0.10	0.39	0.24				
Awkward neck postures	0.27	0.13	0.41	0.29				
Awkward arm postures	0.29	0.14	0.43	0.30				
Awkward wrist postures	0.28	0.14	0.42	0.27				
Physical effort	0.22	0.07	0.36	0.24	0.17	0.03	0.32	0.20
Vibrations through floor	0.33	0.15	0.51	0.37	0.25	0.05	0.46	0.28
Weight of typical part	0.14	0.00	0.30	0.13	0.23	0.07	0.39	0.26
Weight of heaviest part	0.17	0.01	0.34	0.16	0.23	0.07	0.39	0.21
Weight of tool	0.00	0.00	0.34	0.34	0.20	0.00	0.49	0.37
Balance of tool	0.04	0.00	0.37	0.13	0.18	0.00	0.47	0.16
Grip force to handle and use tool	0.00	0.00	0.34	−0.06	0.34	0.09	0.60	0.35
Pressure of tool handle	0.26	0.03	0.50	0.22	0.24	0.00	0.48	0.37
Vibrations through tool handle	0.19	0.00	0.49	0.19	0.34	0.05	0.63	0.41
Pinch force					0.14	0.00	0.29	0.22
Knee bending					0.25	0.11	0.40	0.27
Kneeling or squatting					0.21	0.06	0.36	0.26
Lifting objects ⩾10 pounds					0.37	0.24	0.51	0.37

(ICC = 0.67 for symptoms and 0.61 for physical examination plus symptoms) or recovered (ICC = 0.71 for symptoms and 0.74 for physical examination plus symptoms), compared to those who did not change UE MSD status between T0 and T1 (ICC = 0.54 for both case definitions) (Table 4).

3.2. Long-term reproducibility of self-reported exposures

Agreement between physical exposure ratings at baseline or at T1 and recalled at T2 (Table 5) was poor for all the items investigated. ICCs of the composite exposure index were similar in both comparisons (ICC = 0.40 in T0–T2 and 0.37 in T1–T2 analysis). This was not true for reproducibility of single items, which showed a high variability between T0–T2 and T1–T2 analyses. In both analyses reproducibility of recall was significantly higher among workers who remained in the same job title at T0 and T1 (ICC = 0.51 for both), compared to those who changed job between the first two surveys (ICC = 0.30 in T0–T2 and 0.03 in T1–T2 analysis). Differences in the exposure index between T0, or T1, and 6-year follow-up were both positive, suggesting that in general these workers tended to overestimate past physical exposures.

Higher reproducibility was found, in the T0–T2 analysis, among UE MSD cases based on symptoms only, compared to non-cases, but lower when the case status was based on symptoms plus physical examination, none of them being statistically significant. In the T1–T2 analysis stratification by UE MSD case status showed only small differences in reproducibility between cases and non-cases (Table 3).

4. Discussion

In this cohort of manufacturing workers, a series of self-reported exposure ratings provided data for several analyses relating to stability and recall of those assessments. One-year test–retest reliability of an exposure index for upper extremity physical hazards could be judged acceptable. Such an index could be employed in retrospective cohort studies to represent with a certain degree of reliability the self-reported exposures that would have been obtained within the past year. Among specific exposure items, however, 1-year reproducibility appeared good only for a few variables: use of tools and parts handling (both dichotomized), and exposure to whole-body vibration (continuous).

Others have reported the reproducibility of self-reported physical activity at work to be higher than that observed in this study (Ainsworth et al., 1993; Leijon et al., 2002; Sandmark et al., 1999; Spieholz et al., 1999; Torgén et al., 1997; Wiktorin et al., 1996b, 1999), although most of them evaluated test–retest reliability over a much shorter time period (i.e., 1 or 2 weeks). Also, the majority of the results from these studies indicated a higher reproducibility for whole-body exposures, such as sitting, perceived exertion and whole-body vibration. Exposure ratings concerning specific body regions, as for hands above shoulders, repetitive movements and hand vibration, are more heterogeneous in the literature but they would seem to have poor or moderate reliability (Leijon et al., 2002; Sandmark et al., 1999; Torgén et al., 1997; Wiktorin et al., 1996b).

Torgén et al. (1997), who analyzed 1-year reproducibility in a Swedish population-based study, observed higher agreement than we found for most of the exposures assessed by both studies, such as physical effort, lifting, whole-body and hand vibration. Besides possible cultural differences in reproducibility, a possible explanation for this discrepancy is that their scores were based on proportion of time spent in each posture (except physical exertion, which was expressed on a semi-continuous scale of intensity), whereas in the present study they were based on intensity of the physical exposures.

Another important difference is that we examined 1-year test–retest reliability, not reproducibility. In our first follow-up survey, workers were asked only to rate their current exposures, not those at baseline. Thus, we could only examine stability of these exposure reports between T0 and T1 among subjects in the same job and with the same work content (by self-assessment). Since technological changes were introduced mainly between the second and the third interviews, exposure to physical factors associated with a specific job was assumed to have remained relatively constant between baseline and 1-year follow-up. Imperfect accuracy of job history could have led to misclassification of this subset and lowered reliability of the physical exposures. Indirect evidence for this is suggested by the agreement observed for the questions on "use of tools" ($k = 0.60$) and "handling parts" ($k = 0.50$), which would have been expected higher if workers' job would have not changed at all.

Reproducibility over the longer follow-up periods of 5 and 6 years was consistently low, with very few items showing acceptable agreement. In contrast, Koster et al. (1999) reported moderate long-term reproducibility (24 years) of self-reported ergonomics exposures, but their questions concerned only the presence/absence of whole-body exposures. Torgén et al. (1999) found good or moderate 6-year reproducibility for proportion of time exposed to a number of factors, including perceived exertion, kneeling and squatting, hand above shoulders and manual handling. Both of these analyses used a sample of the general population, with a likely bigger range in exposures than ours, and thus less sensitivity to small differences in recall.

Since we assessed agreement among respondents in the same jobs, according to self-reported work history without external verification, imperfect accuracy in matching could have decreased estimates of reproducibility. In fact, when the analysis was restricted to workers who were in the same job at T0 and T1, reproducibility of the exposure index improved to moderate in both T0–T2 and T1–T2 analyses. This finding could actually indicate that job matching was worse among subjects who changed job title between T0 and T1 than in workers who did not, although all records underwent a thorough revision to examine job agreement. Nonetheless, it cannot be excluded that the higher reliability resulted from subjects who held their jobs for longer time periods, being better able to remember working conditions related to those jobs.

The strong within-person regression to the mean effect observed for the UE exposure index is expected to have had the consequence of artificially reducing its reproducibility. Its presence also suggests that the within-subject variability of physical exposures was probably quite high in this cohort (Barnett et al., 2005). Since the baseline exposure had substantial predictive value for new cases at 1-year follow-up (Punnett et al., 2004), we do not believe that such variability represents only random error. However, it may be that unavoidable content overlap among exposure rating items leads individual workers to take different strategies for partitioning their experience among domains.

The availability of information on exposure to physical job features before and after changes in production technology provided an opportunity to evaluate if these modifications affected ratings of physical exposures. The changes between the 1- and 6-year surveys varied according to the exposure item examined. As expected, there was a large decrease in physical effort, regardless of whether or not the worker was in the same job as at T1, consistent with the fact that new production technology decreased the intensity of physical workload in this (Gold et al., 2006) and in other vehicle manufacturing companies (Fredriksson et al., 2001; Frieling et al., 1997). Contrary to our expectations, no significant differences were found for grip force and lifting, which showed only small changes. A strong decrease was observed for job pace. New systems of work organization, such as lean production and total quality management, have been reported in general to increase work pace (Landsbergis et al., 1999). However, the changes implemented in these plants did not generally involve such forms of work re-organization, but more the introduction of automation, product and model changes.

The exposure index remained relatively stable between surveys, since large decreases in items as job pace and physical effort were compensated by increases in other physical exposures included in the index, in particular segmental vibration and mechanical pressure through the tool handle.

Exposure to high strain at T2 was only mildly correlated with the difference of the exposure index between surveys, suggesting that physical exposures reporting is not strongly influenced by workplace stress exposure. However, rather than to the mediator effect of psychosocial conditions, the correlation between physical and psychosocial exposures is likely explained by their common origin from a same work organization and technology (MacDonald et al., 2001; Punnett and Wegman, 2004). In addition, some overlap among the items cannot be avoided; for example, assessment of posture intensity probably also involves the experience of posture frequency, duration, and load held in that posture, leading to potential overlap with work pace and effort ratings.

The similarity of ICC and Spearman's ρ observed for most exposures suggests that the distortion of ICC due to lack of normality of the scores' distribution is rather small, demonstrating the robustness of ANOVA techniques when a violation of normality or equal variance assumptions occurs.

Medium-term reliability of the composite exposure index was influenced by change in UE MSD case status from baseline, being significantly higher in those who changed case status. This finding was in the opposite direction of what had been hypothesized and it appears difficult to explain.

Long-term reproducibility was not associated with UE MSD case status at 6-year follow-up, which is consistent with the results of other studies (Halpern et al., 2001; Torgén et al., 1997, 1999; Wiktorin et al., 1996b). Such a finding would permit one to rule out important distortions of the association between self-reported physical exposures and UE MSD occurrence in case-control, retrospective cohort studies and cross-sectional studies. However, the results of some validity studies comparing self-reported physical exposures with observations or direct measurements suggest that subjects with musculoskeletal complaints have a tendency to report higher exposure than that found by objective methods (Balogh et al., 2004; Leijon et al., 2002; Wiktorin et al., 1993), which would produce an overestimate of the relative risks in cross-sectional studies based on self-reports.

5. Conclusions

The present study found acceptable test–retest of self-reported exposure to ergonomics exposures at work over 1 year for a composite exposure index and for specific items on use of tools, manual handling, and exposure to whole-body vibration. Low reproducibility of recall was observed in the long-term (5–6 years) for practically all the exposures investigated, although some of our results would indicate that imperfect job matching, together with the strong regression to the mean effect observed, contributed to reduce agreement artificially. The introduction of technological changes produced on average a decrease in physical workload intensity, as well in job pace, but very small variation in the overall exposure index for the upper limb.

These results are only partially consistent with those reported in the literature, possibly because of methodological differences among studies as well as real changes in the workplaces studied here.

A higher reliability of reporting was noted in the medium-term for subjects who changed UE MSD case status from baseline, contrary to our expectation. No consistent differences in reproducibility were observed by categories of upper limb MSD case status, suggesting that differential misclassification of exposure to physical hazards is unlikely.

Acknowledgements

This research was supported by a contract from the National Joint Committee on Health and Safety of the United Automobile Workers and Daimler-Chrysler Automotive Corporation, and a grant from the U.S. National Institute for Occupational Safety and Health (R01-OH03514). Jamie Tessler, Jenny Ro, and Gary Davis participated in field observations. The authors also thank Robert Karasek, Rebecca Moreland, and David Wegman for advice and assistance.

References

Ainsworth, B.E., Jacobs Jr., D.R., Leon, A.S., Richardson, M.T., Montoye, H.J., 1993. Assessment of the accuracy of physical activity questionnaire occupational data. J. Occup. Med. 35, 1017–1027.

Armstrong, B.K., White, E., Saracci, R., 1992. Principles of Exposure Measurement in Epidemiology. Oxford University Press, Oxford, New York.

Balogh, I., Orbaek, P., Winkel, J., Nordander, C., Ohlsson, K., Ektor-Andersen, J., 2001. Questionnaire-based mechanical exposure indices for large population studies–reliability, internal consistency and predictive validity. Scand. J. Work Environ. Health 27, 41–48.

Balogh, I., Orbaek, P., Ohlsson, K., Nordander, C., Unge, J., Winkel, J., Hansson, G.A., Malmo Shoulder/Neck Study Group, 2004. Self-assessed and directly measured occupational physical activities—influence of musculoskeletal complaints, age and gender. Appl. Ergon. 35 (1), 49–56.

Barnett, A.G., van der Plos, J.C., Dobson, A.J., 2005. Regression to the mean: what it is and how to deal with it. Int. J. Epidemiol. 34 (1), 215–220.

Bernard, B.P., 1997. Musculoskeletal Disorders and Workplace Factors: A Critical Review of Epidemiologic Evidence for Work-related Musculoskeletal Disorders of the Neck, Upper Extremity, and Low Back. National Institute for Occupational Safety and Health, Cincinnati, OH.

Borg, G., 1990. Psychophysical scaling with applications in physical work and the perception of exertion. Scand. J. Work Environ. Health 16 (S1), 55–58.

Burdorf, A., 1992. Exposure assessment of risk factors for disorders of the back in occupational epidemiology. Scand. J. Work Environ. Health 18, 1–9.

Fleiss, J.L., 1981. Statistical Methods for Rates and Proportions. Wiley, New York.

Fredriksson, K., Bildt, C., Hägg, G., Kilbom, Å., 2001. The impact on musculoskeletal disorders of changing physical and psychosocial work environment conditions in the automobile industry. Int. J. Ind. Ergonomics 28 (1), 31–45.

Frieling, E., Freiboth, M., Henniges, D., Saager, C., 1997. Effects of teamwork on the working conditions of short cycled track work: a case study from the European automobile industry. Int. J. Ind. Ergonomics 20 (5), 371–388.

Gamberale, F., 1988. Maximum acceptable work loads for repetitive lifting tasks. An experimental evaluation of psychophysical criteria. Scand. J. Work Environ. Health 14 (S1), 85–87.

Gold, J.E., Park, J.-S., Punnett, L., 2006. Work routinization and implications for ergonomic exposure assessment. Ergonomics 49, 12–27.

Grieco, A., Molteni, G., De Vito, G., Sias, N., 1998. Epidemiology of musculoskeletal disorders due to biomechanical overload. Ergonomics 41, 1253–1260.

Hagberg, M., Silverstein, B., Wells, R., Smith, M.J., Hendrick, H.W., Carayon, P., Perusse, M., 1995. In: Kuorinka, I., Forcier, L. (Eds.), Work related musculoskeletal disorders (WMSDs): a reference book for prevention. Taylor & Francis, London, Bristol.

Halpern, M., Hiebert, R., Nordin, M., Goldsheyder, D., Crane, M., 2001. The test–retest reliability of a new occupational risk factor questionnaire for outcome studies of low back pain. Appl. Ergon. 32, 39–46.

Karasek, R.A., 1985. Job Content Questionnaire and User's Guide. University of Southern California/University of Massachusetts Lowell, Los Angeles/Lowell, MA.

Kilbom, Å., 1994. Assessment of physical exposure in relation to work-related musculoskeletal disorders—what information can be obtained from systematic observations? Scand. J. Work Environ. Health 20 (special issue), 30–45.

Koster, M., Alfredsson, L., Michelsen, H., Vingard, E., Kilbom, A., 1999. Retrospective versus original information on physical and psychosocial exposure at work. Scand. J. Work Environ. Health 25 (5), 410–414.

Landsbergis, P.A., Cahill, J., Schnall, P., 1999. The impact of lean production and related new systems of work organization on worker health. J. Occup. Health Psychol. 4 (2), 108–130.

Leijon, O., Wiktorin, C., Harenstam, A., Karlqvist, L., 2002. Validity of a self-administered questionnaire for assessing physical workloads in a general population. J. Occup. Environ. Med. 44, 724–735.

MacDonald, L.A., Karasek, R.A., Punnett, L., Scharf, T., 2001. Covariation between workplace physical and psychosocial stressors: evidence and implications for occupational health research and prevention. Ergonomics 44 (7), 696–718.

National Research Council, Institute of Medicine, 2001. Musculoskeletal Disorders and the Workplace: Low Back and Upper Extremities. National Academy Press, Washington, D.C.

Punnett, L., 1998. Ergonomic stressors and upper extremity disorders in vehicle manufacturing: cross sectional exposure-response trends. Occup. Environ. Med. 55, 414–420.

Punnett, L., van der Beek, A.J., 2000. A comparison of approaches to modeling the relationship between ergonomic exposures and upper extremity disorders. Am. J Industr. Med. 37, 645–655.

Punnett, L., Wegman, D.H., 2004. Work-related musculoskeletal disorders: the epidemiologic evidence and the debate. J. Electromyogr. Kinesiol. 14, 13–23.

Punnett, L., Gold, J.E., Katz, J.N., Gore, R., Wegman, D.H., 2004. Ergonomic stressors and upper extremity disorders in automotive manufacturing: a one-year follow-up study. Occup. Environ. Med. 61, 668–674.

Sandmark, H., Wiktorin, C., Hogstedt, C., Klenell-Hatschek, E.K., Vingard, E., 1999. Physical workload in physical education teachers. Appl. Ergon. 30, 435–442.

Satterthwaite, F.E., 1946. An approximate distribution of estimates of variance components. Biometric Bull. 2, 110–114.

Spielholz, P., Silverstein, B., Stuart, M., 1999. Reproducibility of a self-report questionnaire for upper extremity musculoskeletal disorder risk factors. Appl. Ergon. 30, 429–433.

Stock, S.R., Fernandes, R., Delisle, A., Vezina, N., 2005. Reproducibility and validity of workers' self-reports of physical work demands. Scand. J. Work Environ. Health 31 (6), 409–437.

Torgén, M., Alfredsson, L., Koster, M., Wiktorin, C., Smith, K.F., Kilbom, A., 1997. Reproducibility of a questionnaire for assessment of present and past physical activities. Int. Arch. Occup. Environ. Health 70, 107–118.

Torgén, M., Winkel, J., Alfredsson, L., Kilbom, A., 1999. Evaluation of questionnaire-based information on previous physical workloads. Stockholm MUSIC 1 Study Group. Musculoskeletal Intervention Center. Scand. J. Work Environ. Health 25, 246–254.

van der Windt, D.A., Thomas, E., Pope, D.P., de Winter, A.F., Macfarlane, G.J., Bouter, L.M., Silman, A.J., 2000. Occupational risk factors for shoulder pain: a systematic review. Occup. Environ. Med. 57, 433–442.

Wiktorin, C., Karlqvist, L., Winkel, J., Stockholm MUSIC I Study Group, 1993. Validity of self-reported exposures to work postures and manual materials handling. Scand. J. Work Environ. Health 19 (3), 208–214.

Wiktorin, C., Selin, K., Ekenvall, L., Alfredsson, L., 1996a. An interview technique for recording work postures in epidemiological studies. Int. J. Epidemiol. 25, 171–180.

Wiktorin, C., Hjelm, E.W., Winkel, J., Koster, M., 1996b. Reproducibility of a questionnaire for assessment of physical load during work and leisure time. J. Occup. Environ. Med. 38, 190–201.

Wiktorin, C., Vingard, E., Mortimer, M., Pernold, G., Wigaeus-Hjelm, E., Kilbom, A., Alfredsson, L., 1999. Interview versus questionnaire for assessing physical loads in the population-based MUSIC-Norrtalje Study. Am. J. Ind. Med. 35, 441–455.

Winkel, J., Mathiassen, S.E., 1994. Assessment of physical work load in epidemiologic studies: concepts, issues and operational considerations. Ergonomics 37, 979–988.

Yeung, S.S., Genaidy, A.M., Karwowski, W., Leung, P.C., 2002. Reliability and validity of self-reported assessment of exposure and outcome variables for manual lifting tasks: a preliminary investigation. Appl. Ergon. 33, 463–469.