Int Arch Occup Environ Health (2006) 79: 499–508
DOI 10.1007/s00420-005-0068-1

ORIGINAL ARTICLE

Pierre-R. Somville · An Van Nieuwenhuyse
Laurence Seidel · Raphaël Masschelein · Guido Moens
Philippe Mairiaux · The BelCoBack Study Group

Validation of a self-administered questionnaire for assessing exposure to back pain mechanical risk factors

Received: 21 April 2005 / Accepted: 22 November 2005 / Published online: 31 December 2005

Abstract *Objective*: To validate a self-administered questionnaire assessing exposure to mechanical risk factors, developed for a cohort study aiming at assessing the influence of physical and psycho-social factors on the incidence of low back pain (LBP). *Methods*: The study first involved a criterion validity test. A sample of the cohort workers ($n=152$) was observed at the workplace during four 30 min periods randomly distributed along the shift. At the end of the work shift, the questionnaire was filled in both by the worker and the observer. Agreements were tested between self-reports and observations, and between self-reports and observer opinion. Secondly, a comparison of exposure–effect relationships based on self-reports to those based on observations was carried out on the whole study cohort ($n=716$). Both sets of Relative Risks of being an incident case (LBP lasting at least 7 consecutive days in the follow-up year) were tested for heterogeneity. *Results*: Self-reports agreement levels were better with observer opinion than with observational data and were higher for answers at a dichotomous level. Vehicle driving, manual handling without estimation of weight and frequencies, or trunk bending without rotation showed a fair to good agreement with the external criteria. Limits in the validation procedure did not allow validating the sitting and standing durations. As regards the health outcome comparison, questionnaire and observations led to homogeneous Relative Risks for the variables tested. *Conclusions*: Results show that self-reports provide a limited accuracy to assess actual frequencies and durations of work activities. Using a questionnaire, classifying the workers into exposure categories is rather relative, but questionnaire and observations seem similar in their relationships to outcome.

Keywords Questionnaire · Risk factors · Low back pain · Occupational exposure · Validation study

Pierre-R. Somville (✉) · Ph. Mairiaux
Occupational Health and Health Education unit,
Department of Public Health, University of Liège,
Sart Tilman (B23), 4000 Liege, Belgium
E-mail: pr.somville@ulg.ac.be
Tel.: +32-43-662491
Fax: +32-43-662889

A. V. Nieuwenhuyse · R. Masschelein
Department of Public Health, Section of Occupational,
Environmental and Insurance Medicine,
Katholieke Universiteit Leuven, Kapucijnenvoer 35,
3000 Leuven, Belgium

L. Seidel
Biostatistics unit, Department of Public Health,
University of Liège, Sart Tilman (B23),
4000 Liege, Belgium

G. Moens
IDEWE - Occupational Health Services, Interleuvenlaan 58,
3001 Heverlee, Belgium

Introduction

The present study is part of the BelCoBack study, [Belgian Cohort study on low Back pain (LBP)], a prospective study with a 2-year follow-up, which aims at assessing the influence of individual, physical and psychosocial factors on the incidence of LBP and sick leave.

A critical issue when developing such an epidemiological project is to assess as accurately as possible both exposure and outcome in order to study exposure–effect relationships. For assessing the exposure to biomechanical factors, several methods are available that can be classified into three levels of increasing complexity (Van der Beek and Frings-Dresen 1998): a first level of subjective evaluation (by questionnaires, diaries or interviews), a second one based on observations and a third one, very specialized, needing the subject instrumentation. Both the accuracy of measurements and the costs are increasing when going from the first to the third level and hence, the feasibility decreases within an epidemiological context when complex assessment methods have to be applied to large populations (Van der Beek and Frings-Dresen 1998). In the BelCoBack project, it was decided to combine both self-reports using a questionnaire and direct observations at the workplace to assess mechanical exposure.

A critical analysis of previously published questionnaires (Somville and Mairiaux 2003b) evaluating mechanical risk factors for musculo-skeletal disorders showed that none of those questionnaires was designed so as to assess the whole set of LBP risk factors including postural constraints (trunk bending and/or twisting), manual handling of loads (lifting, carrying, pushing or pulling), whole body vibration (vehicle or engine driving) and, to a lesser degree, static trunk postures (Burdorf and Sorock 1997; Derriennic et al. 2000; Hoogendoorn et al. 1999). To ensure this content, the BelCoBack questionnaire was thus partly derived from existing questionnaires (Campbell et al. 1997; Hollmann et al. 1999; Hoogendoorn et al. 2000; Viikari-Juntura et al. 1996; Wiktorin et al. 1999) and partly developed for the study.

The present paper describes the validation procedure of this study questionnaire and its results. In order to validate the questionnaire, the data collected through the questionnaire were first compared to two external criteria taken as reference: observation of the work activities on the one hand and the observer's own expert opinion on the other hand. If the comparison to observation is frequently used in the literature (Campbell et al. 1997; Viikari-Juntura et al. 1996; Wiktorin et al. 1993), a comparison between worker and observer judgements answering the same questionnaire is less often mentioned (Rossignol and Baetz 1987; Wells et al. 1997). These criterion validity tests have been carried out on a sample of the BelCoBack study population. Secondly, thanks to the prospective design of the BelCoBack study, questionnaire data were compared to observational data in terms of their relationships to the health outcome. This last test was performed on the whole BelCoBack cohort population.

Materials and methods

The BelCoBack study

This study was carried out on a population of young workers (less than 31 years old) working in the distribution and health care sectors, and free from LBP for the previous 12 months. From 1,672 eligible workers, 1,200 (72%) agreed to participate; 159 were excluded because they did not meet the last inclusion criterion leaving a sample of 1,041 workers. The questionnaire at baseline was completed by 972 (93%) of them.

One year later, 800 workers (82%) returned a follow-up questionnaire. As a minimal experience of at least 2 months in the current function at baseline was requested, the cohort was reduced to 716 workers at 1 year follow-up. This sample included 61% women and 39% men; 64% were recruited in the health care sector and 36% in the distribution one. The outcome variable was defined as the occurrence of an episode of LBP lasting at least 7 consecutive days in the follow-up year, as assessed by the self-report (Van Nieuwenhuyse et al. 2005). Exposure was also assessed by direct observation at the workplace on a sample of the study participants ($n=152$). Among those workers, 20 were actually not part of the study cohort for organizational reasons, but were performing the same tasks than the cohort workers. The sample shows a similar gender and sector distribution (62% of women and 65% recruited in the health care sector).

Two outcome models were developed. One based on the questionnaire assessment (Van Nieuwenhuyse et al. 2005) and the other based on observations. Both models included the whole cohort. For the observation-based model, the 716 cohort workers were distributed into exposure group (EG) defined on the basis of the 152 observed workers (see subsequently).

Self-administered questionnaire

In order to ensure the questionnaire content validity, the questionnaire variables were selected according to the biomechanical risk factors for LBP established in the literature (Burdorf and Sorock 1997; Derriennic et al. 2000; Hoogendoorn et al. 1999). The selected questionnaire variables and answer modalities are summed up in Appendices 1 and 2.

The time period on which the subject was asked to evaluate its physical workload consists of a "typical workday", or in case of work activities varying from a day to another, of "the activity or job the most often performed during the last month". Like in other studies exploring the subject perceptions (Duquette et al. 1997; Masset et al. 1998), some questions are designed so as to explore the perceived heaviness of efforts and movements (Q 17 for example). As other questionnaires validated in the literature use pictograms in order to facilitate the question understanding (Hollmann et al. 1999; Pope et al. 1998; Viikari-Juntura et al. 1996; Wiktorin et al. 1996), the two questions (Q 5 and Q 6) assessing the trunk posture are graphically illustrated. Two principles have been taken into account concerning answer modalities (Somville and Mairiaux 2003b): according to the results of other studies (Battie et al. 2002; Campbell et al. 1997; Wiktorin et al. 1993), a dichotomous answer mode is used in most items and, when duration or frequency estimates are deemed necessary, an ordinal and objective scale is used.

Using such answer modalities implies the choice of one (or more) cut-off values. Based on the literature, those values were selected that showed strong associations between the studied factor and the health effects (Somville and Mairiaux 2003b).

Direct observations at the workplace

Based on other observation methods (Buchholtz et al. 1996; Ridd et al. 1989; Van der Beek et al. 1992), real-time observations of the working activities were carried out by a unique trained observer using a discontinuous capture mode with a 15 s time sampling interval. This method consisted of looking at the worker every 15 s and then, to fill in the observation grid, using a Fujitsu stylistic LT C-500* sensitive screen computer. This grid involves three observable data categories (see Appendix 3): the basic motor action, the posture and the load. The posture category is further divided into basic posture, trunk flexion and trunk rotation postures. This grid thus consists of five complementary columns including, in a given column, mutually exclusive items (Somville and Mairiaux 2003a). For each observed worker, four periods of 30 min observation were randomly distributed along the workday.

Validation study procedures

The study first involved a criterion validity test (against direct observation and against observer opinion); both tests were performed on the sample of observed workers. Validation was not performed by gender or by sector for there was a large overlap between these variables: a large majority (about 80%) of women worked in the health care sector, while the same proportion of men worked in the distribution sector.

Secondly, Relative Risks of being an incident case based on self-reports were compared to the ones based on observation data; this comparison was carried out on the basis of the whole BelCoBack study cohort.

Questionnaire validation against direct observation

At the end of the observation shift, each observed worker was invited to answer the self-administered questionnaire having in mind "the present workday if it is a typical workday". Out of the 152 workers, 5 did not return the questionnaire.

As the observational data were of continuous nature and expressed in percentage of encoded events, secondary discrete variables had to be derived taking into account the cut-offs used in the questionnaire, in order to allow a comparison of these two data sets. In order to estimate the frequencies, each encoding was considered as one event; so, each "frequency variable" was expressed in number of times per hour. For estimating durations, each event was assumed to last 15 s; the same events were added and their total duration was translated in hours and minutes; so, "duration variables" were expressed in hour per workday like in the questionnaire. Concerning questions 2 and 3 about sitting and standing, no precise cut-offs were given in the questionnaire and the term "for long periods" had to be interpreted to distribute observed subjects between "exposed" and "non-exposed" workers: a 2 h cut-off was applied based on other validated questionnaires (Campbell et al. 1997; Viikari-Juntura et al. 1996). A similar issue was raised by question 15: "high physical exertion to push/pull a load" was interpreted as a more than 10 kg exertion. Finally, to identify those workers performing manual handling (Q 9) or lifting/carrying loads (Q 10), a limit of >1 kg was chosen for the load estimated weight.

Questionnaire validation against observer opinion

The analyst who observed the worker answered the same questionnaire at the end of the shift and it was thus possible to compare their respective self-reported estimates.

Comparison of Relative Risks based on self-reports or observations

Comparing the outcomes resulting from two different exposure assessments methodologies implies the building of two comparable outcome models. Both models included the 716 cohort workers who self-reported their LBP outcome. The model based on exposure assessment through self-reports followed exactly the BelCoBack model, but only univariate RRs were calculated. The model based on observation data was different from the BelCoBack, for the study subjects had to be dichotomized the same way than in the questionnaire.

The 152 observed workers were sampled within 79 job categories as defined by the management. Similar job categories in terms of exposure level were grouped on the basis of discussions with team managers and according to qualitative observations: 23 function groups (FG) were defined out of the 79 job categories. The exposure level of these FGs consists on the arithmetic mean of the included workers.

These 23 FGs were further grouped using an Euclidian cluster analysis. In the distribution sector, most workers were polyvalent, rotating between two main functions in the same month. So special groups were created for polyvalent workers taking into account the percentage of time spent in each function. In those specific groups, time-weighted averages were used to define the group exposure level. Finally, 28 exposure groups (EG) covering the whole cohort exposure range were developed. The 716 cohort workers were then distributed within those 28 EG and consequently, each worker in a given EG was thus assigned to the same exposure level.

In order to calculate comparable RRs, the continuous observational data were derived into discrete variables using the same cut-offs than in the questionnaire as it has been made for the questionnaire validation against observations.

Statistical treatment

To test for the agreement between the 147 workers' self-reports and the corresponding data derived from the observations, the Cohen's kappa statistic and full agreement percentage were applied in all cases; the Spearman's rank correlation coefficient was calculated for variables having an ordinal answer modality. Moreover, for these ordinal variables, supplementary kappa tests have been performed by grouping the last ranks to reduce the answer scale first to a 3-point and then to a 2-point (dichotomous) scale. Chosen interpretation of kappa values were as follows: <0.40 = poor agreement, 0.40–0.75 = fair to good agreement and >0.75 = excellent agreement (Fleiss 1981). The test concerned 12 from the 18 questionnaire variables because the 6 subjective questions could not be tested by this method.

The statistical comparison between workers questionnaires and the analyst ones made use of the same tests. Contrary to the validation against observation, every item of the questionnaire could be tested.

To compare exposure–effect relationships based on self-reports to those based on observations, both sets of RRs from baseline questionnaire and direct observation were tested for heterogeneity using a Cochrane Q test derived from a meta-analytic procedure (Cucherat 1997). RRs were considered heterogeneous for a P level <0.1.

Results

Concurrent validity

The concurrent validity test has thus been conducted on 147 workers observed at their workplace. Table 1 shows the respective frequencies of these workers' answers to the questionnaire (Self-R) at the end of the observation shift, the corresponding data derived from observational variables (OBS) and the answers based on the observer opinion (Obs-opinion).

From Table 1, it can be seen that the workers, in comparison to observational data and the observer opinion, are overestimating the durations of the periods spent when driving a vehicle, with the trunk bended, and with the trunk in a bended and twisted posture. For frequencies, there is also some overestimation for trunk rotation movements, but not for trunk bending, and a clear overestimation of more than 25 kg loads lifts. For static postures, the workers seem to underestimate the time spent in the sitting position and in the standing position. Worker and observer estimates are in agreement for most subjective items except the posture quality, while lifting/carrying a load (Q 13) and the hindrance to pushing/pulling efforts due to a factor external to the load (Q 17). No tendency to over- or underestimation can be demonstrated for the other items.

For the 11 questionnaire variables with a dichotomic answer modality (yes/no), the results are given in Table 2. Agreement between observational data and self-report data could not be tested for the subjective questions about manual handling (Q 13, 14, 16 and 17) and about the ability to change posture regularly (Q 4). For the other six questions, no kappa values can be judged as excellent. The best values are found for manual handling activities. Postural constraints as well as static postures show poor kappa values.

Table 2 shows that kappa values are always higher when worker self-reports are compared to the observer judgments than when compared to observational data. Manual handling activities exhibit the best agreement values again, whereas agreement can be considered as fair to good for frequent trunk bending but poor for frequent trunk rotation. Prolonged sitting position shows a much better kappa value in this test; possibility to change posture shows a fair to good agreement. The sample size of the four subjective questions about manual handling (Q 13, 14, 16 and 17) was smaller because not every worker reported handling loads at work and even less workers reported having to push/pull loads. Agreement is fair to good for the perceived heaviness associated to the manual handling of loads (Q 17) and the ability to hold the load close to the body (Q 14), but is poor for the other two questions (Q 13 and 16).

Table 3 describes the agreement levels for the eight variables with an ordinal answering mode. As shown by the table, the scale reduction (to 3-point and dichotomous scales) is associated with increasing kappa values.

When self-reports are compared to direct observations, all agreement values are poor and with the exception of driving a vehicle that offers a fair to good to excellent agreement depending on the scale reduction.

When workers self-reports and observer opinions are compared, similar results are found for vehicle driving. Agreement is fair to good at dichotomous level for the trunk bended, but the level of agreement remains poor in all cases for the flexion/rotation combination. Agreement for lifting/carrying more than 10 kg loads becomes fair to good at a 3-point scale; however, levels of agreement are always poor for lifting/carrying more than 25 kg loads. Pushing/pulling a load presents a fair to good agreement at a 3-point scale. Finally, the 10-point Borg scale shows a good Spearman coefficient.

Comparison of Relative Risks based on self-reports or observations

Relative Risks could be compared for 9 questions out of 18. First, the six subjective questions had to be dropped out. Secondly, as regards to observational data, for the question 12, not any cohort worker was exposed and for two other ones (Q 9 and 10), every worker was exposed. Table 4 compares univariate Relative Risks of suffering a LBP episode lasting at least 7 consecutive days during the first year follow-up depending on the way exposure is assessed: by means of questionnaire or through observations. For the nine questions tested, almost every scale had to be reduced to a 3-point scale as no worker was exposed in the highest categories; for the same reason, the work with the trunk bended and twisted (Q6) had to be estimated at a dichotomous level. Heterogeneity Cochrane Q tests on these nine variables demonstrate that the RRs are similar in every case except for the estimation of work duration in a bended posture. For this variable, a

Table 1 Answer frequencies of data from concurrent validity tests (self-report, corresponding variables from direct observations and observer opinion)

	Questionnaire variables	Answer items	Answer frequencies		
			Self-R	OBS	Obs-opinion
1	Driving a vehicle or engine?	No	89	103	93
		< 2 h	11	16	24
		≥2 ≤ 6 h	11	28	17
		> 6 h	36	0	13
2	Work in a sitting position for long periods?	No	115	92	85
		Yes	32	55	62
3	Standing (without walking) for long periods?	No	102	72	45
		Yes	45	75	102
4	Possibility to change posture regularly?	No	13	–	7
		Yes	134	–	140
5	Work with the trunk bended (> 45°) for long periods ?	No	65	30	43
		< 1/2 h	19	103	37
		≥1/2 h, ≤ 1 h	18	14	64
		> 1 h, ≤ 2 h	21	0	4
		> 2 h	24	0	0
6	Work with the trunk bended and twisted for long periods?	No	86	46	70
		< 1/2 h	18	99	69
		≥1/2 h, ≤ 1 h	18	2	7
		> 1 h, ≤ 2 h	13	0	1
		> 2 h	12	0	0
7	Frequent trunk bending (more than 12 times per hour)?	No	43	41	50
		Yes	104	106	97
8	Frequent trunk rotation (more than 12 times per hour)?	No	60	122	104
		Yes	87	25	43
9	Manual handling (lifting. carrying. pushing or pulling a load)?	No	31	14	19
		Yes	116	133	128
10	Lifting or carrying loads?	No	32	16	22
		Yes	115	131	125
11	Lifting or carrying more than 10 kg loads?	No	46	78	41
		< 1/h	39	20	62
		≥1/h, ≤ 12/h	44	41	27
		> 12/h	18	8	17
12	Lifting or carrying more than 25 kg loads?	No	83	144	109
		< 1/h	30	3	30
		≥1/h, ≤ 12/h	29	0	7
		> 12/h	5	0	0
13	Good posture of the back while lifting or carrying a load?	No	51	–	80
		Yes	64	–	35
14	Ability to hold loads close to the body while lifting or carrying?	No	54	–	51
		Yes	61	–	64
15	Important pushing or pulling efforts ?	No	63	103	49
		< 1/h	36	16	66
		≥1/h	48	28	32
16	Pushing or pulling efforts hindered by a factor external to the load?	No	25	–	61
		Yes	51	–	15
17	Are the loads handled too heavy due to their weight or handling frequency?	No	64	–	59
		Yes	52	–	57

dose–response was found only when the RR is estimated on the basis of observation.

Discussion

Criterion validity

The pattern of agreement between the data collected through the questionnaire and the ones drawn from observations fits to the results of other studies. Dichotomous variables showed a better agreement than ordinal variables (Campbell et al. 1997; Pope et al. 1998; Wiktorin et al. 1993). Agreement was stronger for well-defined activities, such as vehicle driving, manual handling without estimation of weight and frequencies (Pope et al. 1998; Wiktorin et al. 1993), or trunk bending compared to the bending/twisting association (Rossignol and Baetz 1987). Concerning the weight estimation of handled loads, kappa values were higher for the intermediate weight category (more than 10 kg) (Campbell et al. 1997; Wiktorin et al. 1993). Results show nevertheless that kappa values were generally lower than in other studies when self-reports are compared to observations (Campbell et al. 1997; Pope et al. 1998; Rossignol and Baetz 1987; Torgen et al. 1999; Wiktorin et al. 1993) and that

Table 2 Agreement tests for dichotomous variables (n subjects, kappa at $P < 0.05$ level and full agreement percentage) between worker self-report data and corresponding variables from direct observations and between worker and observer estimates

	Variables	Self-report compared to observation			Self-report compared to observer opinion		
		n	Kappa (95%CI)	Full %	n	Kappa (95%CI)	Full %
2	Work in a sitting position for long periods?	147	0.29 (0.14–0.44)	69	147	0.52 (0.38–0.67)	78
3	Standing (without walking) for long periods?	147	NS	55	147	0.25 (0.14–0.37)	57
4	Possibility to change posture regularly?	–	–	–	147	0.47 (0.31–0.62)	93
7	Frequent trunk bending (more than 12 times per hour)?	147	0.33 (0.17–0.48)	73	147	0.55 (0.39–0.71)	80
8	Frequent trunk rotation (more than 12 times per hour)?	147	0.15 (0.04–0.26)	52	147	0.34 (0.21–0.48)	65
9	Manual handling (lifting, carrying, pushing or pulling a load) ?	147	0.51 (0.37–0.66)	87	147	0.65 (0.50–0.80)	90
10	Lifting or carrying loads?	147	0.56 (0.41–0.71)	88	147	0.69 (0.53–0.84)	90
13	Good posture of the back while lifting or carrying a load?	–	–	–	115	0.28 (0.13–0.43)	63
14	Ability to hold loads close to the body while lifting or carrying?	–	–	–	115	0.46 (0.29–0.62)	73
16	Pushing or pulling efforts hindered by a factor external to the load?	–	–	–	76	0.17 (0.05–0.29)	50
17	Are the load handled too heavy due to their weight or handling frequency?	–	–	–	117	0.50 (0.34–0.66)	75

the agreement was better when self-reports are compared to the observer opinion. This could be ascribed to some methodological limitations of the present study.

First of all, the comparison between questionnaire and observation data has needed a transformation of the exposure grid items into questionnaire variables.

Secondly, it must be kept in mind that, according to the BelCoBack research protocol, the results are based on four periods of 30 min observation randomly distributed along the workday and this raises a question: as previously discussed by Burdorf and Laan (1991) one may wonder whether these periods were representative enough to assess a whole workday as the worker was invited to do when completing his questionnaire. Such a question was also raised recently by Heinrich et al. (2004) who performed similar criterion validation tests in a population of computer workers and obtained low levels of agreement between self-reports and observations.

Table 3 Agreement tests for ordinal variables (n subjects, Spearman's rank coefficient. kappa at $P < 0.05$ level and full agreement percentage) between worker self-report data and corresponding variables from direct observations and between worker and observer estimates

	Variables	Answer modality	Self-report compared to observation, $n = 147$			Self-report compared to observer opinion, $n = 147$		
			rs	Kappa (95%CI)	Full %	rs	Kappa (95%CI)	Full %
1	Driving a vehicle or engine?	4-Point duration scale	0.93	0.47 (0.38–0.55)	69	0.90	0.61 (0.51–0.71)	78
		3-Point duration scale		0.73 (0.61–0.85)	86		0.72 (0.61–0.84)	85
		Dichotomous scale		0.93 (0.77–1.09)	97		0.94 (0.78–1.10)	97
5	Work with the trunk bended (>45°) for long periods?	5-Point duration scale	0.39	NA	27	0.50	0.14 (0.06–0.21)	33
		3-Point duration scale		NA	31		0.29 (0.18–0.41)	55
		Dichotomous scale		0.23 (0.09–0.37)	64		0.40 (0.25–0.55)	71
6	Work with the trunk bended and twisted for long periods?	5-Point duration scale	NS	NA	31	0.36	0.18 (0.09–0.27)	46
		3-Point duration scale		NA	32		0.22 (0.13–0.32)	50
		Dichotomous scale		0.18 (0.04–0.32)	56		0.35 (0.19–0.51)	67
11	Lifting or carrying more than 10 kg loads?	4-Point frequency scale	0.40	0.21 (0.12–0.31)	44	0.63	0.39 (0.29–0.48)	55
		3-Point frequency scale		0.28 (0.17–0.39)	52		0.41 (0.30–0.52)	60
		Dichotomous scale		0.34 (0.19–0.48)	66		0.63 (0.46–0.79)	84
12	Lifting or carrying more than 25 kg loads ?	4-Point frequency scale	0.16	NA	57	0.36	NA	61
		3-Point frequency scale		NA	57		0.26 (0.15–0.37)	61
		Dichotomous scale		NA	59		0.34 (0.20–0.49)	69
15	Important pushing or pulling efforts?	3-Point frequency scale	0.33	0.21 (0.10–0.32)	52	0.56	0.44 (0.33–0.55)	62
		Dichotomous scale		0.25 (0.12–0.39)	61		0.57 (0.41–0.73)	80
18	Perceived general exertion?	10-Point Borg Scale	–	–	–	0.72	–	–

NA Not analysed due to too few subjects in all categories, *rs* Spearman's rank correlation coefficient

Table 4 Univariate Relative Risks (RR) and 95% confidence interval (95% CI) for a low back pain (LBP) episode lasting 7 days or more during a 1-year follow-up when exposure is assessed by means of a questionnaire (RR quest) or from direct observations (RR obs)

	Variable	Cut-off	Observation assessment				Questionnaire assessment				*Q* value	*P* value
			LBP	No LBP	RR	(95% CI)	LBP	No LBP	RR	(95% CI)		
1	Driving a vehicle or engine?	No	50	401	1.00			46	359	1.00		
		< 2 h	14	78	1.37	(0.79–2.38)	12	66	1.35	(0.75–2.44)	0.01	0.974
		≥2	24	137	1.34	(0.86–2.11)	32	189	1.27	(0.84–1.94)	0.03	0.866
2	Work in a sitting position	No	49	328	1.00		73	497	1.00			
	for long periods?	Yes	39	288	0.92	(0.62–1.36)	16	122	0.91	(0.54–1.50)	0.01	0.967
3	Standing (without walking)	No	37	277	1.00			60	458	1.00		
	for long periods?	Yes	49	339	1.07	(0.72–1.60)	28	154	1.33	(0.88–2.01)	0.532	0.466
5	Work with the trunk bended	No	13	120	1.00			36	289	1.00		
	(> 45°) for long periods?	< 1/2 h	67	473	1.27	(0.72–2.23)	11	32	2.31	(1.27–4.19)	2.16	0.141
		≥1/2 h	8	23	2.64	(1.20–5.81)	43	297	1.14	(0.75–1.73)	3.43	0.064
6	Work with the trunk bended	No	28	257	1.00			46	397	1.00		
	and twisted for long periods?	Yes	60	359	1.46	(0.96–2.22)	41	215	1.54	(1.04–2.28)	0.04	0.847
7	Frequent trunk bending	No	15	123	1.00			30	233	1.00		
	(more than 12 times per hour)?	Yes	73	493	1.19	(0.70–2.00)	60	380	1.20	(0.79–1.80)	0.01	0.982
8	Frequent trunk rotation	No	80	576	1.00			35	278	1.00		
	(more than 12 times per hour)?	Yes	8	40	1.37	(0.70–2.66)	54	337	1.24	(0.83–1.84)	0.07	0.797
11	Lifting or carrying more than	No	11	103	1.00			23	182	1.00		
	10 kg loads?	< 1 t/h	44	276	1.43	(0.76–2.66)	21	169	0.99	(0.56–1.72)	0.75	0.387
		≥1 t/h	33	237	1.27	(0.66–2.42)	44	264	1.27	(0.79–2.04)	0.01	0.990
15	Important pushing or	No	23	196	1.00			31	291	1.00		
	pulling efforts?	< 1 t/h	50	316	1.30	(0.82–2.07)	27	174	1.40	(0.86–2.27)	0.04	0.838
		≥1 t/h	15	104	1.20	(0.65–2.21)	30	153	1.70	(1.07–2.72)	0.79	0.373

Q value = Cochrane *Q* test value

In fact, the four periods sample might not be enough to assess the exposure level of one individual, but it must be kept in mind that this sampling protocol was initially designed for a group approach in which exposure level was assessed by several "four periods" days for a same exposure group (Hoozemans et al. 2001).

Thirdly, when considering the nature of the observed working activities, it must be underlined that, in the health care sector, most job categories or functions involved varying tasks in the same day without any regular cycle time while, in the distribution sector, where most activities were more repetitive or cyclic in nature, workers were rotating between several functions within a given month or a given week. So, "the day of today if it is a typical workday" may have been wrongly interpreted by some cohort workers as an average working day due to the variability of their activities. This hypothesis cannot be ascertained as no detailed task information was collected in our study. A task-based exposure assessment being not considered as an optimal strategy in an epidemiologic context (Svendsen et al. 2005), a randomized time sampling strategy was chosen instead (Somville and Mairiaux 2003a). Using a much shorter reference period to compare self-reports and observations like the 1-h period used in the study of Pope et al. (1998) may lessen strongly the "averaging" cognitive process requested from the subject and therefore results in higher agreements. However, except for highly repetitive work, such a short assessment period does not seem appropriate to assess a whole workday.

Finally, the discontinuous capture mode used in the present study might explain the lower agreement levels as this mode only enables an estimation of durations and frequencies. However, such an approximation was judged reasonable in the Kilbom et al. review (Kilbom 1994). For frequency estimation, the 15 s time sampling interval used in our study allowed estimating up to 4 events a minute, a maximum level actually sufficient for, in the questionnaire, the cut-off frequency was "more than 12 events an hour". While a misclassification in terms of frequency underestimation can be rejected, an overestimation is still possible. Concerning durations, the fact that each encoded event was considered as lasting 15 s may also have led to an overestimation of the actual duration of activities, such as working with the trunk bended and/or twisted. In a pilot study, we have thus compared the discontinuous capture mode to a video-based continuous recording. This study showed a tendency to underestimate the actual frequencies when using the discontinuous capture mode, whereas estimated durations were in agreement with those measured using the video recordings (results not shown). Similar results have been reported by other authors (Fransson-Hall et al. 1995).

In contradiction with the literature (Torgen et al. 1999; Viikari-Juntura et al. 1996), there was in the present study a poor agreement between self-reports and observations for static postures (prolonged sitting posture and work in a standing position). This could be ascribed to the fact that a 2 h cut-off was selected to distribute the observed workers into exposed and non-exposed subjects, while no duration limit was proposed in the questionnaire. No qualitative information is available on how the workers have interpreted the definition of the term "for long periods" used in these two questions (see Appendix 1). Concerning these variables, the observation method seems too inaccurate to be considered as a reference.

Comparison of Relative Risks based on self-reports or observations

While in the questionnaire validation test, the population study was restricted to the observed sample ($n = 147$), in this last analysis, the whole study cohort has been used ($n = 716$). Of course, the issues discussed in the validation tests about deriving observational data, being continuous items, into variables expressed in durations or frequencies with precise cut-offs remain. Yet, in such analysis, the exposure level of each cohort worker does not depend anymore on the four observation periods, but on the exposure level of the "exposure group" the worker belongs to. The exposure level within a given exposure group was actually based on more than four observation periods (on average, 26 ± 14 periods depending on the group size). This fact suggests that exposure assessed on an observation basis was more accurate in this analysis because of the increased number of observation periods.

Only one heterogeneous RR was found for the work in a bended posture. While there was a dose–response effect with the observation derived exposure, the probability to develop a LBP episode was higher for the lesser exposed workers as assessed on the basis of questionnaire. However, if the workers' answers are classified dichotomously, by grouping the two exposed categories, the corresponding RRs become homogeneous with a Q value = 0.024 and a P level = 0.876. This observation underlines the importance to avoid asking too many details in terms of duration in such questions about postural constraints.

Conclusions

The results of the present study show that the criterion validity results fit with the literature data except for the static postures, particularly the sitting position. Agreements were always higher at dichotomous level and were always poor over a 3-point frequency or duration scale. A clear tendency for the workers to overestimate actual durations and frequencies was found in the present study. The overall level of agreement was lower than in other studies, and this might partly be due to the study limitations discussed before. However that may be, the present results remind that self-reports accuracy is limited especially when exposure frequencies and durations are asked in detail.

Nevertheless, when the exposure–effect relationships were analysed on the 716 cohort workers, the results were similar, whatever the exposure assessment method was, either based on questionnaire or on direct observation. So, even if the questionnaire did not allow assessing accurately the actual frequencies or durations of work activities, the questionnaire-based classification of workers into exposure categories seems to present a similar relationship to outcome than the one found with observations. So, the results of the present study suggest that the difference between self-reports and observations concerns more the exposure cut-off accuracy than the exposure–effect relationship itself. Workers classification into exposure categories is rather relative using a questionnaire and the health risk cannot be accurately related to an absolute level of the exposure parameters.

Acknowledgments The BelCoBack study was supported by the Belgian government, SPP-PS research projects PS/93/25, PS/12/26, PS/01/27; the present study was also supported by the SMIL Foundation and a FNRS research grant S 2/5-DD-E42. The authors thank the other co-workers from the BelCoBack study group for their contribution to data collection and data processing: G. Crombez, K. Johannik, A. Leys, L. Moors, E. Persijn, D. Pirenne, G. Verbeke and O. Van den Bergh. The authors also thank F. Richy for his contribution in the statistical treatment.

Appendix 1

Questionnaire variables and answer modalities

	Questionnaire variables	Answer scale
1	Driving a vehicle or engine?	4-Point duration scale
2	Work in a sitting position for long periods?	Dichotomous scale
3	Standing (without walking) for long periods?	Dichotomous scale
4	Possibility to change posture regularly?	Dichotomous scale
5	Work with the trunk bended (more than 45°) for long periods?	5-Point duration scale
6	Work with the trunk bended and twisted for long periods?	5-Point duration scale
7	Frequent trunk bending (more than 12 times per hour)?	Dichotomous scale
8	Frequent trunk rotation (more than 12 times per hour)?	Dichotomous scale
9	Manual handling (lifting. carrying. pushing or pulling a load)?	Dichotomous scale
10	Lifting or carrying loads?	Dichotomous scale
11	Lifting or carrying more than 10 kg loads?	4-Point frequency scale
12	Lifting or carrying more than 25 kg loads?	4-Point frequency scale
13	Good posture of the back while lifting or carrying a load?	Dichotomous scale
14	Ability to hold loads close to the body while lifting or carrying?	Dichotomous scale
15	Important pushing or pulling efforts?	3-Point frequency scale
16	Pushing or pulling efforts hindered by a factor external to the load?	Dichotomous scale
17	Are the load handled too heavy due to their weight or handling frequency?	Dichotomous scale
18	Rating of perceived exertion (Borg category-ratio scale)	10-Point Borg scale

Appendix 2

Answer modalities

Answer scale	Answer item				
	1	2	3	4	5
Dichotomous scale	No	Yes			
3-Point frequency scale	No	Less than once an hour	Once an hour or more		
4-Point frequency scale	No	Less than once an hour	From 1 to 12 times per hour	More than 12 times per hour	
4-Point duration scale	No	Less than 2 h a day on average	From 2 to 6 h a day on average	More than 6 h a day on average	
5-Point duration scale	No	Less than 1/2 h a day	From 1/2 to 1 h a day	From 1 to 2 h a day	More than 2 h a day

Appendix 3

Observation grid

Basic posture	Flexion	Rotation	Basic motor action	Load (kg)
Standing	0–20°	0° Rot	No action/holding a load	Less than 1
Sitting	21–45°	Rot +	Walking/carrying	From 1 to 10
Kneeling/ squatting	> 45°		Driving	From 11 to 25
			Lifting/lowering	> 25
			Pushing/pulling	
			Throwing	
			Other action	

References

Battie MC, Videman T, Gibbons LE, Manninen H, Gill K, Pope M, Kaprio J (2002) Occupational driving and lumbar disc degeneration: a case-control study. Lancet 360(9343): 1369–1374

Buchholtz B, Paquet V, Punnett L, Lee D, Moir S (1996) PATH: a work sampling-based approach to ergonomic job analysis for construction and other non-repetitive work. Appl Ergon 27(3):177–187

Burdorf A, Laan J (1991) Comparison of Methods for the Assessment of Postural Load on the Back. Scand J Work Environ Health 17(6):425–429

Burdorf A, Sorock G (1997) Positive and negative evidence of risk factors for back disorders. Scand J Work Environ Health 23(4):243–256

Campbell L, Pannett B, Egger P, Cooper C, Coggon D (1997) Validity of a questionnaire for assessing occupational activities. Am J Ind Med 31(4):422–426

Cucherat M (1997) Hétérogénéité et analyse en sous groupes. Méta-analyse des essais thérapeutiques, Masson, pp 253–261

Derriennic F, Leclerc A, Mairiaux Ph, Meyer JP, Ozguler A (2000) Lombalgies en milieu professionnel: quels facteurs de risque et quelle prévention? INSERM ed, Paris, p 137

Duquette J, Lortie M, Rossignol M (1997) Perception of difficulties for the back related to assembly work: general findings and impact of back health. Appl Ergon 28(5–6):389–396

Fleiss JL (1981) The measurement of interrater agreement: statisticalmethods for rates and proportions, 2nd edn. Wiley, New York, pp 212–236

Fransson-Hall C, Gloria R, Kilbom A, Winkel J (1995) A portable ergonomic observation method (PEO) for computerized on-line recording of posture and manual handling. Appl Ergon 26(2):93–100

Heinrich J, Blatter BM, Bongers PM, Heinrich J, Blatter BM, Bongers PM (2004) A comparison of methods for the assessment of postural load and duration of computer use. Occup Environ Med 61(12):1027–1031

Hollmann S, Klimmer F, Schmidt KH, Kylian H (1999) Validation of a questionnaire for assessing physical work load. Scand J Work Environ Health 25(2):105–114

Hoogendoorn WE, van Poppel MN, Bongers PM, Koes BW, Bouter LM (1999) Physical load during work and leisure time as risk factors for back pain. Scand J Work Environ Health 25(5):387–403

Hoogendoorn WE, Bongers PM, de Vet HC, Douwes M, Koes BW, Miedema MC, Ariens GA, Bouter LM (2000) Flexion and rotation of the trunk and lifting at work are risk factors for low back pain: results of a prospective cohort study. Spine 25(23):3087–3092

Hoozemans MJ, Burdorf A, Van der Beek AJ, Frings-Dresen MH, Mathiassen SE (2001) Group-based measurement strategies in exposure assessment explored by bootstrapping. Scand J Work Environ Health 27(2):125–132

Kilbom A (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(Spec no):30–45

Masset DF, Piette AG, Malchaire JB (1998) Relation between functional characteristics of the trunk and the occurrence of low back pain. Associated risk factors. Spine 23(3):359–365

Pope DP, Silman AJ, Cherry NM, Pritchard C, Macfarlane GJ (1998) Validity of a self-completed questionnaire measuring the physical demands of work. Scand J Work Environ Health 24(5):376–385

Ridd JE, Nicholson AS, Montan AJ (1989) A portable microcomputer based system for "on-site" activity and posture recording. Taylor and Francis. Ergonomics-designing process. Megaw ED ed. London, pp 366–369

Rossignol M, Baetz J (1987) Task-related risk factors for spinal injury: validation of a self-administered questionnaire on hospital employees. Ergonomics 30(11):1531–1540

Somville PR, Mairiaux Ph (2003a) Using on-site observations for low back related work load assessment. Methodological issues. In: Strasser H, Kluth K, Rausch H, Bubb H (eds) Quality of work and products in entreprises of the future. Ergonomia verlag, Stutgart, pp 219–222

Somville PR, Mairiaux Ph (2003b) Evaluation des risques biomécaniques pour la colonne lombaire sur la base d'un questionnaire. Méd Travail Ergon 40(1):5–17

Svendsen SW, Mathiassen SE, Bonde JP (2005) Task based exposure assessment in ergonomic epidemiology: a study of upper arm elevation in the jobs of machinists, car mechanics, and house painters. Occup Environ Med 62(1):18–27

Torgen M, Winkel J, Alfredsson L, Kilbom A (1999) Evaluation of questionnaire-based information on previous physical work loads. Stockholm MUSIC 1 study group. Musculoskeletal intervention center. Scand J Work Environ Health 25(3):246–254

Van der Beek AJ, Frings-Dresen MH (1998) Assessment of mechanical exposure in ergonomic epidemiology. Occup Environ Med 55(5):291–299

Van der Beek AJ, Van Gaalen LC, Frings-Dresen MH (1992) Working posures and activities of lorry drivers: a reliability study of on-site observation and recording on a pocket computer. Appl Ergon 23(5):331–336

Van Nieuwenhuyse A, Somville PR, Crombez G, Burdorf A, Verbeke G, Johannik K, Van den Bergh O, Masschelein R, Mairiaux Ph, Moens G (2005) The role of physical workload and pain-related fear in the development of low back pain in young workers. Evidence from the belcoback study: results after one year of follow-up. Occup Environ Med (in press)

Viikari-Juntura E, Rauas S, Martikainen R, Kuosma E, Riihimaki H, Takala EP, Saarenmaa K (1996) Validity of self-reported physical work load in epidemiologic studies on musculoskeletal disorders. Scand J Work Environ Health 22(4):251–259

Wells R, Norman R, Neumann P, Andrews D, Frank J, Shannon H, Kerr M (1997) Assessment of physical work load in epidemiologic studies: common measurement metrics for exposure assessment. Ergonomics 40(1):51–61

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

Wiktorin C, Hjelm EW, Winkel J, Koster M (1996) Reproducibility of a questionnaire for assessment of physical load during work and leisure time. Stockholm MUSIC I study group. J Occup Environ Med 38(2):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(5):441–455