Bias and Power in Group-Based Epidemiologic Studies of Low-Back Pain Exposure and Outcome - Effects of Study Size and Exposure Measurement Efforts.

Coenen, P.; Mathiassen, S.E.; Kingma, I.; Boot, C.R.; Bongers, P.M.; van Dieen, J.H.

published in
Annals of Occupational Hygiene
2015

DOI (link to publisher)
10.1093/annhyg/meu102

Ann. Occup. Hyg., 2015, Vol. 59, No. 4, 439–454
doi:10.1093/annhyg/meu102
Advance Access publication 27 November 2014

The Chartered Society for Worker Health Protection

Bias and Power in Group-Based Epidemiologic Studies of Low-Back Pain Exposure and Outcome – Effects of Study Size and Exposure Measurement Efforts

Pieter Coenen[1–3], Svend Erik Mathiassen[4], Idsart Kingma[2,3], Cécile R. L. Boot[3,5], Paulien M. Bongers[3,6] and Jaap H. van Dieën[2,7*]

1.School of Physiotherapy and Exercise Science, Curtin University, Perth, Kent Street Bentley WA-6102, Western Australia, Australia
2.MOVE Research Institute Amsterdam, Faculty of Human Movement Sciences, VU University Amsterdam, van der Boechorststraat 9, NL-1081 BT Amsterdam, the Netherlands
3.Body@Work, Research Center on Physical Activity, Work and Health, the Netherlands
4.Centre for Musculoskeletal Research, Department of Occupational and Public Health Sciences, University of Gävle, SE-801 76 Gävle, Sweden
5.Department of Public and Occupational Health, EMGO Institute for Health and Care Research, VU University Medical Center, Amsterdam, van der Boechorststraat 7, NL-1081 BT Amsterdam, the Netherlands
6.TNO Healthy Living, Leiden, Schipholweg 77-89 NL-2316 ZL Leiden, the Netherlands
7.King Abdulaziz University, 21589 Jeddah, Saudi Arabia
*Author to whom correspondence should be addressed. Tel: +31-20-5988501; fax: +31-20-5988529; e-mail: j.vandieen@fbw.vu.nl
Submitted 11 July 2014; revised 21 September 2014; revised version accepted 28 October 2014.

ABSTRACT

Objectives: Exposure–outcome studies, for instance on work-related low-back pain (LBP), often classify workers into groups for which exposures are estimated from measurements on a sample of workers within or outside the specific study. The present study investigated the influence on bias and power in exposure–outcome associations of the sizes of the total study population and the sample used to estimate exposures.
Methods: At baseline, lifting, trunk flexion, and trunk rotation were observed for 371 of 1131 workers allocated to 19 ***a-priori*** defined occupational groups. LBP (dichotomous) was reported by all workers during 3 years of follow-up. All three exposures were associated with LBP in this parent study ($P < 0.01$). All 21 combinations of $n = 10, 20, 30$ workers per group with an outcome, and $k = 1, 2, 3, 5, 10, 15, 20$ workers actually being observed were investigated using bootstrapping, repeating each combination 10 000 times. Odds ratios (OR) with P values were determined for each of these virtual studies. Average OR and statistical power ($P < 0.05$ and $P < 0.01$) was determined from the bootstrap distributions at each (n, k) combination.
Results: For lifting and flexed trunk, studies including $n \geq 20$ workers, with $k \geq 5$ observed, led to an almost unbiased OR and a power >0.80 (P level = 0.05). A similar performance required $n \geq 30$ workers for rotated trunk. Small numbers of observed workers (k) resulted in biased OR, while power was, in general, more sensitive to the total number of workers (n).
Conclusions: In epidemiologic studies using a group-based exposure assessment strategy, statistical performance may be sufficient if outcome is obtained from a reasonably large number of workers, even if exposure is estimated from only few workers per group.

KEYWORDS: epidemiology; ergonomics; exposure; exposure assessment; group-based measurement strategy; low-back pain; musculoskeletal injury; precision

INTRODUCTION

In the past decades, numerous epidemiological studies have been conducted on occupational risk factors for low-back pain (LBP). Among other factors, physical exposures such as heavy lifting, trunk flexion, and trunk rotation have been suggested to increase the risk of LBP (Lötters *et al.*, 2003; Griffith *et al.*, 2012). However, the literature on physical (biomechanical) risk factors of LBP is not consistent (Bakker *et al.*, 2009; Kwon *et al.*, 2011), one possible reason being that exposure assessment strategies differ between studies (Punnett and Wegman, 2004; David, 2005).

While it is generally recommended to base exposure estimation on direct measurements or observations rather than on self-reports (Winkel and Mathiassen, 1994; van der Beek and Frings-Dresen, 1998), only few epidemiologic studies of physical exposures have attempted to do so. One major reason is that extensive resources would be required for collecting exposure data from each subject in a study of the size necessary to obtain sufficient power in detecting exposure–outcome associations. As a compromise, some studies have chosen a group-based strategy for exposure assessment, where workers are classified into groups, typically based on their job or tasks (Hoogendoorn *et al.*, 2000; Svendsen *et al.*, 2004; Burdorf and Jansen, 2006). The exposure variable(s) of interest is then obtained from measurements on a sample of workers representing each group, and the average exposure of the measured workers is assigned to all workers in the group. Exposure–outcome relationships are then determined using these exposure estimates together with individual data on health outcomes from all participants in the study. The use of this group-based exposure assessment strategy is based on the assumption that workers within a group have similar exposures, i.e. that the exposure groups are homogeneous, and that exposure variability between groups is comparatively large (Kromhout and Heederik, 1995).

Group-based exposure assessment has been discussed as an effective approach in epidemiologic research for about two decades, in particular in the field of occupational hygiene (Houba *et al.*, 1997; Heederik and Attfield, 2000). A theoretical framework has been developed for assessing bias and precision of exposure–outcome studies using linear regression (Reeves *et al.*, 1998; Tielemans *et al.*, 1998). Linear exposure–outcome relationships were shown to be essentially unbiased with group-based exposures, while individual-based models, where each individual is assigned his/her own exposure under a classical error structure, lead to attenuated slopes unless each individual is measured extensively. This advantage of the group-based strategy comes, however, at the price of an increased uncertainty of the regression coefficient and thus reduced power, i.e. reduced ability of a study design to detect a true effect of exposure on outcome (Armstrong, 1998; Tielemans *et al.*, 1998). Therefore, to avoid both bias and lack of power, hybrid strategies attempting to combine the advantages of individual- and group-based strategies have been proposed (Seixas and Sheppard, 1996; Peretz *et al.*, 2005).

However, these general results do not readily apply to studies using logistic or log-linear regression, in which case bias is still an issue with group-based exposure assessment (Reeves *et al.*, 1998; Steenland *et al.*, 2000; Kim *et al.*, 2012). Furthermore, the results are mainly relevant to studies where personal exposure information is available from all participants, and not only from a subsample, as is the case in the studies of LBP cited above (Hoogendoorn *et al.*, 2000; Svendsen *et al.*, 2004; Burdorf and Jansen, 2006). Thus, the properties of the highly relevant case of using a group-based strategy with exposure data obtained from a limited population sample in a logistic regression of exposure versus outcome is not well understood. Jansen and Burdorf (2003) showed in a study of LBP in a-priori defined occupational groups that odds ratios (ORs) were greatly dependent on whether exposures were assigned using an individual- or group-based approach, and even that the result was sensitive to whether all subjects or only a subsample were included in the group-based exposure estimate. In a recent theoretical study complemented by simulations using hypothetical data, Kim *et al.* (2011) confirmed that group-based strategies can, indeed, lead to bias in logistic regression, and that the magnitude of this bias depends on the overall size of the study, the number of workers on whom exposure estimates are based, and the exposure variance within and between groups. However, this study did not address the extent to which the OR estimate and the statistical power of the study design would be affected by the total study size, i.e. the number of subjects from whom outcome data are available, and the number of subjects involved in exposure measurements. Furthermore, empirical data to explore the practical importance of these possible effects were not reported. Therefore, the present study was designed to determine—on the basis of a large empirical data set on low-back exposures and pain—the influence of the number of subjects from whom outcome data are obtained, and of the size of the sample of workers on whom exposure is actually measured, on bias and power of logistic exposure–outcome associations, using a group-based exposure assessment strategy.

STUDY POPULATION AND METHODS

Population

The present study is based on data from the Study on Musculoskeletal disorders, Absenteeism, and Health. As described in detail previously (Hoogendoorn *et al.,* 2000; Coenen *et al.,* 2013), this prospective cohort study recruited workers from 34 companies in the Netherlands. At baseline, 1989 of 2048 invited workers agreed to participate, and questionnaire data on personal factors and work characteristics were obtained from 1802 (91%) of them. These eligible workers were classified by experts into 23 groups, based on their expected physical work load. Within each group, work was recorded on video from a random sample of roughly one-fourth of the workers. After a 3-year follow-up period, data on LBP were collected in the study population. At that point, 1131 (N) workers belonged to groups containing >5 observed workers; i.e. to 19 of the original 23 groups (Table 1). Information on LBP status was available for all of these 1131 workers, and for a total of 371 (K) workers video-based exposure observations were also available; this formed the parent dataset for the present study. Descriptive statistics of the eligible population as well as the parent data set are provided in Table 1.

Exposure and outcome for the parent dataset

For each of the 371 workers recorded on video, four recordings were obtained at randomly chosen times during the course of a single work day. Recordings lasted 5–15 min each, depending on the variability of the worker's tasks. Recordings were analyzed *post-hoc* using a structured protocol for assessing three physical exposures, which were shown to be significantly associated with LBP in the same population (Hoogendoorn *et al.,* 2000); i.e. the number of lifts during an 8-h work week, the percentage of working time with the trunk flexed (defined as >30° trunk flexion), and the percentage of working time with the trunk rotated (defined as >30° trunk rotation).

Exposure variances between occupational groups and between workers within groups (pooled estimate) were determined from the entire dataset using a standard two-way nested analysis of variance (ANOVA) model (Searle *et al.,* 1992; Kromhout *et al.,* 1995), including random effects of group, worker within group, and measurement within worker (based on the four measurement occasions). On the basis of these variance components, the effectiveness of the group classification to generate groups with sufficient between-group variability relative to the residual within-group variability was assessed for each of the three exposures using a contrast metric:

Equation 1:

$$\text{Contrast} = \frac{\text{MSE}_{\text{BG}}}{(\text{MSE}_{\text{BG}} + s^2_{\text{BS(g)}})} \tag{1}$$

in which MSE_{BG} is the mean squared error between groups (estimated as the between-group variance multiplied by 18/19), and $s^2_{\text{BS(g)}}$ is the variance between workers within groups (Kromhout and Heederik, 1995; Mathiassen *et al.,* 2005). Between- and within-worker variance components were also determined for each of the 19 occupational groups separately using standard one-way random ANOVA models (e.g. Searle *et al.,* 1992; Loomis and Kromhout, 2004). Pairwise correlations between the three exposures (lifting versus flexion, lifting versus rotation, and flexion versus rotation) were assessed by Pearson's correlation coefficients calculated across the 371 subjects with measured exposures.

Table 1. Parent dataset.

Description groups	Workers in total (N)	LBP prevalence (95% CI)	Workers observed (K)	Lifts			Flexion			Rotation		
				Mean	$s_{BS(g)}$	$s_{WS(g)}$	Mean	$s_{BS(g)}$	$s_{WS(g)}$	Mean	$s_{BS(g)}$	$s_{WS(g)}$
Mainly sitting work												
Sitting with varying postures	133	39 (31–47)	61	23.24	71.87	8.17	6.33	7.17	0.98	12.27	29.27	0.52
Sitting with little varying postures (computer work)	57	35 (23–47)	16	13.39	41.85	0.00	7.44	7.45	3.18	1.11	1.88	0.31
Sitting with little varying postures, in awkward postures (no computer work)	31	68 (51–84)	11	1.09	0.00	0.00	3.63	9.03	0.87	2.87	2.40	1.40
Sitting with little varying postures, with repetitive movements	95	42 (32–52)	31	334.29	917.23	74.53	2.40	2.56	0.34	2.39	2.42	0.41
Mainly standing work												
Standing with varying postures (including walking) without external forces	26	58 (39–77)	9	8.00	5.16	5.16	4.05	2.78	1.20	2.23	2.43	0.93
Standing with varying postures and small external forces	69	38 (26–49)	23	658.86	713.69	212.67	7.16	3.24	1.29	2.44	2.20	0.48
Standing with varying postures and moderate external forces	87	44 (33–54)	28	438.05	448.15	90.68	9.96	8.26	0.56	5.59	3.71	0.86
Standing with varying postures and large external forces	65	40 (28–52)	20	299.52	246.44	68.25	11.52	0.95	2.50	6.16	5.01	0.75

Table 1. *Continued*

Description groups	Workers in total (N)	LBP prevalence (95% CI)	Workers observed (K)	Lifts			Flexion			Rotation		
				Mean	$s_{BS(g)}$	$s_{WS(g)}$	Mean	$s_{BS(g)}$	$s_{WS(g)}$	Mean	$s_{BS(g)}$	$s_{WS(g)}$
Standing with varying, awkward postures and moderate external forces	66	50 (38–62)	22	544.36	515.21	214.05	13.69	7.38	0.75	6.14	3.13	0.99
Awkward postures (mainly static exposure)												
Standing in static awkward posture without external forces	42	48 (33–63)	15	133.60	137.30	8.76	8.43	6.59	1.53	4.33	3.23	0.39
Standing in static awkward posture with small external forces	70	39 (27–50)	24	194.71	250.40	51.45	10.31	5.95	1.14	6.67	5.28	0.84
Mainly static back exposures by alternating awkward postures	28	61 (43–79)	11	814.81	977.27	386.99	37.63	30.20	5.21	12.10	6.10	2.49
Alternating exposures (standing, walking, and/or sitting)												
Alternating standing, walking, and/or sitting without external forces	167	40 (33–48)	29	6.36	2.23	2.23	5.70	3.16	0.19	2.40	3.17	0.12
Alternating standing, walking, and/or sitting with small external forces	36	50 (34–66)	13	82.15	39.13	29.00	8.94	0.91	2.60	4.15	5.76	0.87
Alternating standing, walking, and/or sitting with moderate external forces	52	42 (29–56)	15	312.92	88.02	59.01	22.03	9.61	3.41	4.66	5.78	0.62

Table 1. *Continued*

Description groups	Workers in total (N)	LBP prevalence (95% CI)	Workers observed (K)	Lifts Mean	Lifts $s_{BS(g)}$	Lifts $s_{WS(g)}$	Flexion Mean	Flexion $s_{BS(g)}$	Flexion $s_{WS(g)}$	Rotation Mean	Rotation $s_{BS(g)}$	Rotation $s_{WS(g)}$
Alternating standing, walking, and/or sitting with large external forces	21	86 (71–100)	8	2904.00	1210.96	588.54	42.53	12.85	4.22	19.84	6.67	1.67
Alternating standing and walking in static awkward postures, external forces	27	44 (26–63)	17	379.16	295.08	70.70	19.22	8.63	5.46	6.75	5.05	1.59
Alternating standing and walking in postures, moderate external forces	36	56 (39–72)	9	577.21	213.92	181.55	12.79	4.57	2.46	7.37	2.87	1.81
Combined functions (as a result of changes in tasks)												
Combined exposures	23	30 (12–49)	9	252.19	290.93	49.10	8.37	5.95	1.42	2.38	2.19	0.23
Pooled	1131	39 (31–47)	371									
Descriptive variables	***N***		***K***		**Eligible**							
Number of workers	1131		371		1802							
LBP at baseline, number of workers (%)	399 (36%)		128 (35%)		631 (36%)							
Age, years (SD)	35.8 (8.8)		35.2 (8.5)		35.4 (8.8)							
Males, number (%)	800 (70%)		259 (70%)		1246 (70%)							
Females, number (%)	331 (30%)		112 (30%)		543 (30%)							
Stature, cm (SD)	176.1 (9.2)		175.8 (9.6)		176.0 (9.3)							

Table 1. *Continued*

Description groups	Workers in total (N)	LBP prevalence (95% CI)	Workers observed (K)	Lifts			Flexion			Rotation		
				Mean	$s_{BS(g)}$	$s_{WS(g)}$	Mean	$s_{BS(g)}$	$s_{WS(g)}$	Mean	$s_{BS(g)}$	$s_{WS(g)}$
Body weight, kg (SD)	75.8 (13.0)		75.8 (12.9)		75.9 (13.4)							
Working time, hours per week (SD)	9.7 (7.8)		9.6 (7.6)		9.6 (7.7)							
Employment time, years (SD)	38.2 (7.8)		38.1 (7.6)		37.9 (7.7)							

In the upper panel, the total number of workers (N) and the number of workers observed (K) are shown for each group, together with LBP prevalence in percent among all workers (with 95% CI). For each group, mean exposures and standard deviations (between and within workers, $s_{BS(g)}$ and $s_{WS(g)}$, respectively) are shown for the three investigated physical exposures: number of lifts at work per week, percent time spent with the trunk flexed >30°, and percent time spent with the trunk rotated >30°. Also, pooled LBP prevalence and standard deviation (between and within groups) are shown. In the lower part of the table, pooled descriptive statistics (gender, length, weight, age, working hours per week, years of employment at the current job, and proportion of workers with LBP at baseline) are shown for the entire group of workers in the parent dataset of this study (N), the group of observed workers (K), as well as the group of workers that were eligible for this study

Self-reported LBP was assessed for all 1131 workers once a year for 3 years after the baseline measurement using a Dutch version of the Nordic Questionnaire (Kuorinka *et al.*, 1987). A case of LBP was registered when a worker reported regular or prolonged LBP during at least one of the 3 years of follow-up, regardless of baseline status.

For each exposure variable, the mean exposure of the observed workers in each of the 19 groups was assigned to all workers classified into that group. Logistic regression analyses were then performed using these exposures as continuous independent variables (in which the number of lifts was divided by 100 and percentages of time in flexed or rotated postures were divided by 10) and LBP as the dichotomous dependent variable. Results showed the number of lifts [per 100 lifts; OR: 1.06, 95% confidence interval (CI): 1.03–1.09, $P < 0.01$], the time working with the trunk flexed [per 10%; OR: 1.31 (95% CI: 1.12–1.52), $P < 0.01$], and the time working with the trunk rotated [per 10%; OR: 1.43 (95% CI: 1.06–1.93), $P < 0.01$], all to be significantly associated with LBP in the parent dataset. Previous studies on the same population found ORs of 1.72 (95% CI: 1.16–2.57), 1.57 (95% CI: 1.06–2.32), and 1.79 (95% CI: 1.22–2.63) for lifting, trunk flexion, and trunk rotation, respectively. In that study, exposure was, however, classified into categories (Hoogendoorn *et al.*, 2000). We chose to instead analyze exposure as a continuous variable (Loomis, 2012), and express ORs as estimated risks per unit of the three exposure variables (i.e. per 100 lifts, per 10% time trunk flexion, and per 10% time trunk rotation, respectively). Therefore, our results are difficult to compare to the previous analyses of the same material, even though they are consistent with those analyses. Correlation coefficients of 0.34, 0.09, and 0.09 were found for lifting versus flexion, lifting versus rotation, and flexion versus rotation, respectively.

Simulated sampling strategies

For all 21 possible combinations of n = 10, 20, 30 workers in total per group and k = 1, 2, 3, 5, 10, 15, 20 workers being observed, exposure–outcome associations were assessed using a non-parametric bootstrap simulation procedure (Efron and Tibshirani, 1986; Davison and Hinkley, 1997). Bootstrapping is an attractive alternative for examining exposure assessment strategies in cases where analytical procedures are either not available or rely on assumptions that are suspected to be violated (Burdorf and van Riel, 1996; Hoozemans *et al.*, 2001; Paquet *et al.*, 2005; Liv *et al.*, 2011; Mathiassen and Paquet, 2010). Within each group of the parent dataset, workers were identified as 'observed' (K, Table 1) and 'non-observed' (N-K, Table 1) depending on whether exposure data were available or not. For each combination of n and k, n workers in each group were drawn with replacement from all workers (N, i.e. observed and non-observed workers combined) and k workers were drawn with replacement from the group of observed workers (K) in the same group. This led to a virtual study including n workers for whom outcome data were available, and k observed workers in each group providing exposure data. The random selection of n and k in two separate processes implies that the k virtual workers providing exposure data within each particular group are not necessarily a subsample of those n virtual workers in the group from whom outcome data are obtained. The selection process also allows for k being larger than n; for instance k = 20 workers per group may provide exposure data in a virtual study including only n = 10 workers with outcome data in each group. Thus, our simulation includes both the common scenario of k being a subsample of the n workers in the outcome population, and scenarios where exposure data are obtained, e.g. from previous exposure assessment studies in the study population. For each virtual study, the three mean exposures (number of lifts, trunk flexion, and trunk rotation) of the k observed workers within each group were assigned to all n workers in that particular group, while the individual LBP status was used as the outcome for each of the n workers. For each virtual study constructed this way, the ORs (with P levels) for the three associations between each of the exposure variables and LBP were assessed using logistic regression analysis as explained above for the parent dataset, in addition to basic descriptive statistics on exposures and outcomes in the 19 occupational groups. For each of the 21 possible combinations of n and k, 10 000 virtual studies were constructed using this procedure. Four summary performance measures for each investigated exposure assessment strategy were obtained from the 10 000 virtual study results, i.e.

1) a pooled estimate of the standard deviation of the mean exposure estimate within a group, obtained by first calculating, for each group separately, the mean observed gross variance between subjects across the 10 000 replicate estimates, $s^2_{BS(g)}{}^*$, and then pooling these 19 variances to give an average standard deviation of a group mean exposure estimate according to the formula:

Equation 2:

$$\text{Pooled SD}_{\text{mean}} = \sqrt{\frac{\text{mean}\left(s^2_{BS(g)}{}^*\right)}{k}} \qquad (2)$$

2) the standard deviation across the 10 000 studies of LBP prevalence across all 19 groups
3) the mean OR across the 10 000 studies
4) the power of each exposure assessment strategy to detect a significant OR at levels $P < 0.05$ and $P < 0.01$, i.e. the proportions of the 10 000 studies resulting in a significant OR:

Equation 3:

$$\text{power} = \left(\sum_{i=1}^{10000} p_i < p\right) \cdot \frac{1}{10000} \text{ for } i = 1,2,\ldots 10000 \qquad (3)$$

In which p_i is the p level of each of the 10 000 virtual studies and p is set at 0.05 and 0.01 in the two analyses of power at these levels of significance.

All calculations were performed using customary scripts in Matlab (MATLAB 7.7.0; The MathWorks Inc., Natick, MA, USA, 2000). Logistic regression analyses were performed using the Matlab statistical toolbox.

RESULTS

Exposure contrasts between groups were 0.61, 0.70, and 0.19 for the number of lifts, time in flexed trunk posture, and time in rotated trunk posture, respectively. While groups did, indeed, differ in mean exposure (Table 1), some were very heterogeneous in terms of workers differing substantially in exposure (large $s_{BS(g)}$, including an unknown contribution from between-days variability within subjects). Besides, some of the groups even showed considerable exposure variability within days for individual workers (large $s_{WS(g)}$).

As expected, for all three exposure variables, the pooled standard deviation of the group mean exposure decreased as the number of workers, k, for which exposure was actually observed increased (Fig. 1). This effect did not depend on the total number of workers, n, per group. Equally trivial, the standard deviation of the prevalence of LBP in the study population decreased when the total number of workers, n, included in each group increased (Fig. 2). This effect did not depend on k.

The OR for the association between exposure and LBP increased with larger k at all values of n (Fig. 3), and it was affected only little by n, i.e. the total number of workers in each group, all reporting their LBP status. Fig. 4 shows that power increased with both n and k. In general, the effect of the total number of workers, n, on power was stronger than that of the number of observed workers, k. However, the magnitude of these effects differed between risk factors. For the number of lifts, a power of 0.80 to detect a significant OR ($P < 0.05$) was obtained when at least $n = 20$ workers were included per group, and the number of actually observed workers in each group (k) was at least 5. For time with flexed trunk and time working with the trunk rotated, the required number of observed workers at $n = 20$ differed (being ~7 and 14, respectively). At a more strict level of significance ($P < 0.01$), a power of 0.80 was obtained only when the population included at least $n = 30$ workers per group for lifts and flexed trunk. The required number of observed workers at this study size was larger for flexed trunk (~8) than for lifts (~2). For time working in a trunk rotated posture, a power of 0.80 at $P < 0.01$ was not obtained with any of the simulated study designs.

DISCUSSION

The present study dealt with a common exposure assessment strategy in epidemiology: measuring exposure to risk factors in one population of workers and use the resulting data as an exposure estimate for workers within the same occupational group. In the typical case addressed in the present study, measured mean exposures in a number of occupational groups are assigned to all workers having similar tasks or jobs, while information on outcomes is available from each individual

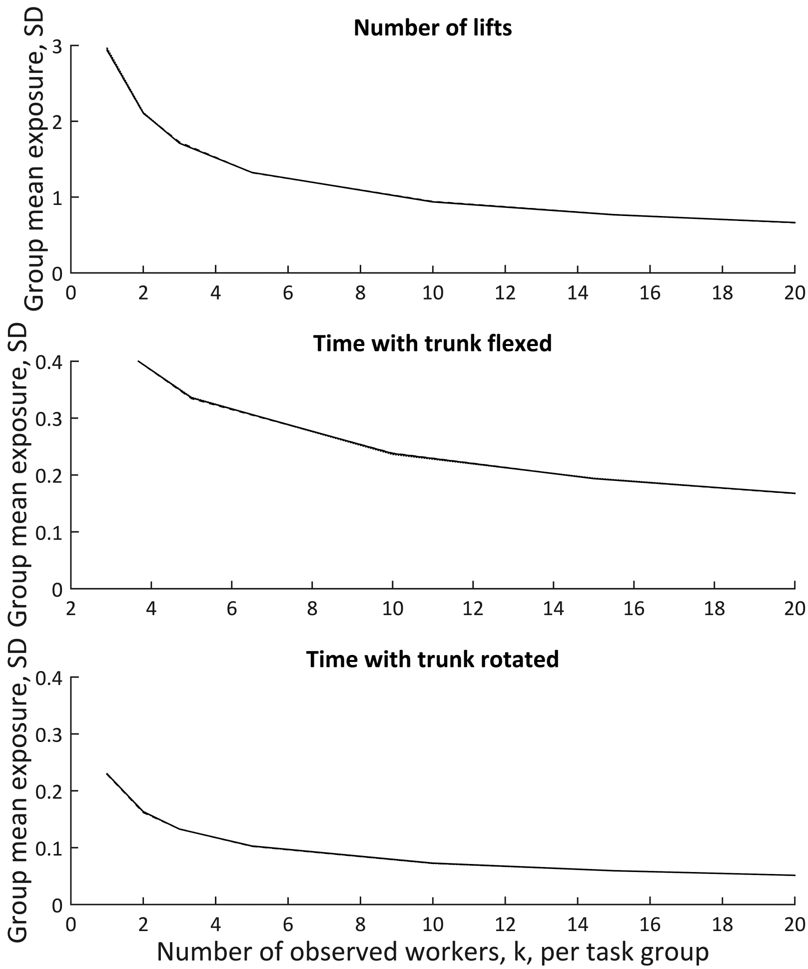

Figure 1 Pooled estimate of the standard deviation of the group mean exposure in a group for each of the 21 combinations of n (different lines) and k (x-axis). Standard deviation is presented for the exposure variables: number of lifts (upper panel), time with the trunk flexed (middle panel), and time with a rotated trunk (lower panel). Note that the individual curves for different n values in each panel overlap completely.

worker in the total study population. Such group-based approaches have been proposed for and are often used in exposure–outcomes studies on chemical agents (Houba *et al.*, 1997), and they have even been applied in musculoskeletal epidemiology (Hoogendoorn *et al.*, 2000; Jansen and Burdorf, 2003; Svendsen *et al.*, 2004). In spite of this, the effects on exposure–outcome bias and study power of using different alternatives for group-based exposure estimation is not well understood.

Interpretation of results

Our study suggests that the probability of finding significant exposure–outcome associations depends more on the total number of workers included in each occupational group than on the number of workers for whom exposure is actually observed. In our setting, comprising 19 groups intended to be representative for the general working population, studies including at least 30 workers in each group with group exposure estimates being based on at least 5 observed workers were sufficient to secure a reasonable power and an almost unbiased estimate of the OR. However, the exact numbers of subjects to obtain a certain statistical performance in terms of OR and statistical power differed between the three investigated exposures (Figs 3 and 4). Our results could have important implications

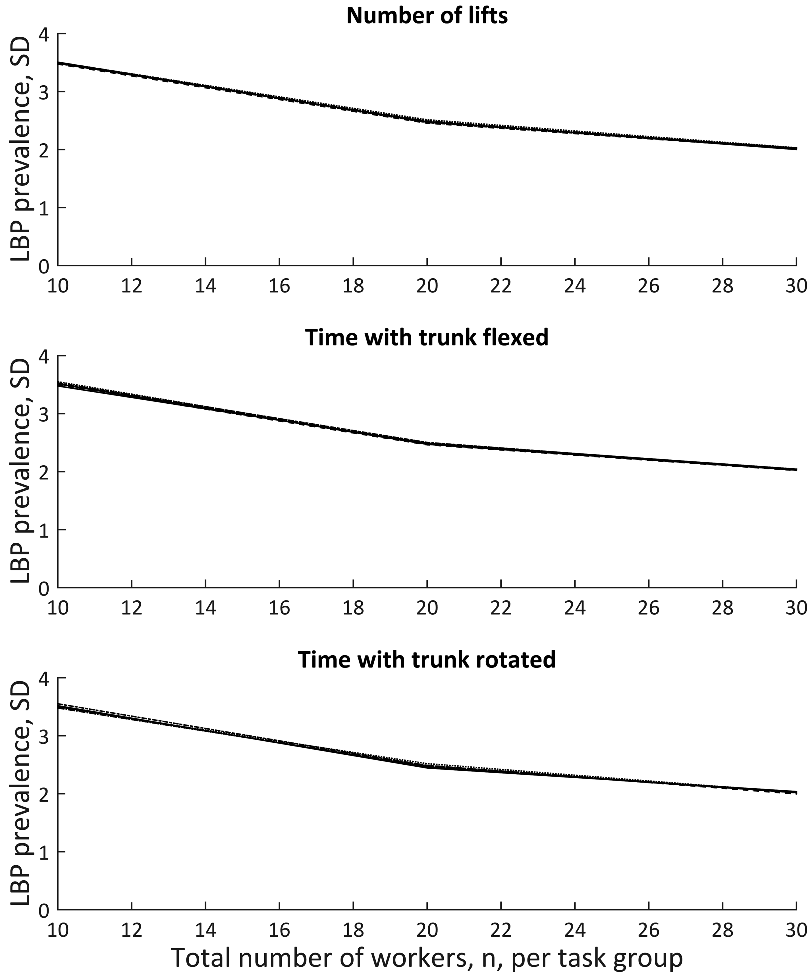

Figure 2 Standard deviation of the outcome (i.e. LBP prevalence in the entire dataset) across the 10 000 replicates for each of the 21 investigated combinations of *n* (*x*-axis) and *k* (different lines). Standard deviations are presented for the exposure variables: number of lifts (upper panel), time with the trunk flexed (middle panel), and time with a rotated trunk (lower panel). Note that the individual curves for different *k* values in each panel overlap completely.

for future epidemiological studies, since they suggest that a research budget could be more efficiently used by collecting outcome data from 'many' subjects than by spending extensive efforts on exposure observations, which are often expensive (Trask *et al.*, 2012). As an illustration, reading from Fig. 4, for the exposure variable 'number of lifts', a statistical power of 0.80 at $P < 0.01$ can be reached both by a study design comprising 30 workers with outcome data per group but only two providing exposure data, and by a study including 20 workers with outcome data per group and 20 being observed for exposures. Thus, the 'large' study requires outcome data to be collected from 570 workers, but exposure only from 38, while the 'small' study is based on outcome data from only 380 workers, but exposure data from all 380. Even if the budgets of these two alternatives depend on the unit cost of obtaining exposure and outcome information, it seems likely that the study with 30 workers per group is cheaper to conduct. Notably, though, while these two sampling strategies have comparable abilities to detect a significant association between exposure and LBP, the former will result in a more biased OR (Fig. 3) as discussed further below.

Standard equations for calculating the power of a study design show that the probability of obtaining

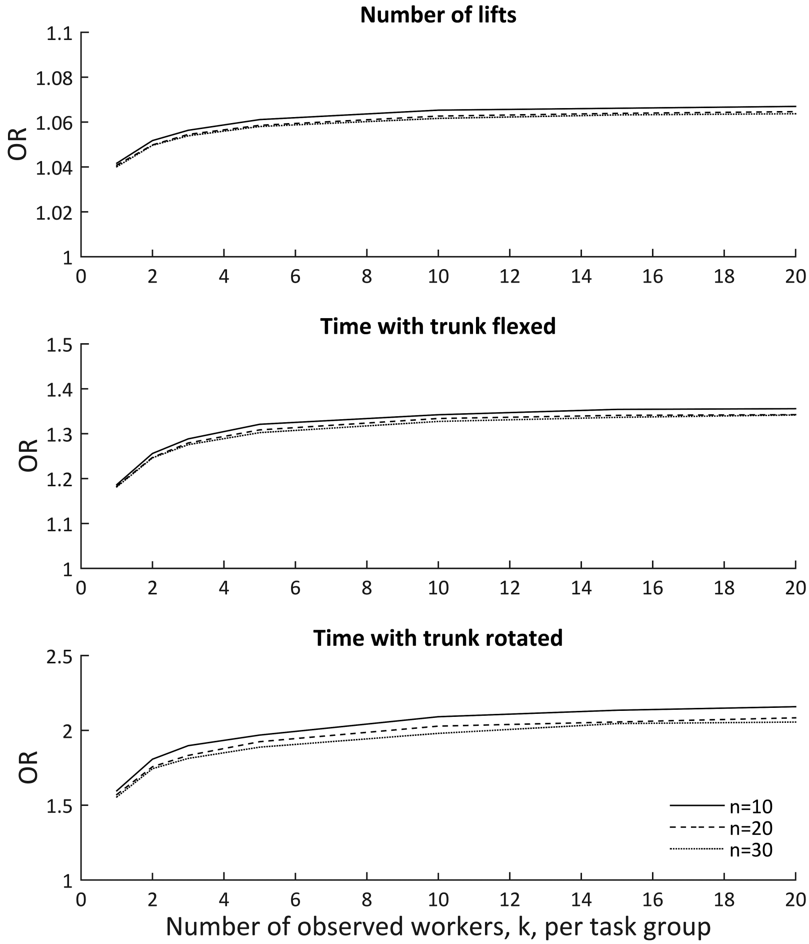

Figure 3 Average OR for the association between exposure and LBP across the 10 000 replicates for each of the 21 investigated combinations of *n* (different lines) and *k* (*x*-axis). Average ORs are presented for the exposure variables: number of lifts (upper panel), time with the trunk flexed (middle panel), and time with a rotated trunk (lower panel).

statistically significant results increases with sample size and decreases with increasing exposure variability (Mathiassen *et al.*, 2002; Mathiassen *et al.*, 2003; Twisk, 2003). Quality guidelines based on these basic statistical facts have been suggested for observational cohort studies of exposure–outcome associations (Vlaanderen *et al.*, 2008). The present study confirms this general effect of more data improving power in a more complex study design than those usually addressed in epidemiologic literature, and illustrates it by quantitative empirical data. More notably, our study adds the novel and important observation that the number, *k*, of measured workers providing the group-based exposure estimates does have an effect on power, but that this effect is weaker than that of changing the number of workers from whom outcome data are available, *n* (Fig. 4).

The decrease in average ORs with smaller numbers of *k*, i.e. a bias of the OR toward 1, is probably a result of increased uncertainty in the estimate of group exposures, since the OR was only weakly influenced by the overall number of workers, *n*, in each group. Our empirical results (Fig. 3) confirm previous theoretical studies showing that the exposure–outcome relationship can, indeed, be biased when a group-based exposure assessment strategy is used in logistic regression (; Reeves *et al.*, 1998; Kim *et al.*, 2011). Furthermore, it emphasizes that the general notion of group-based assessment being an effective measure to eliminate

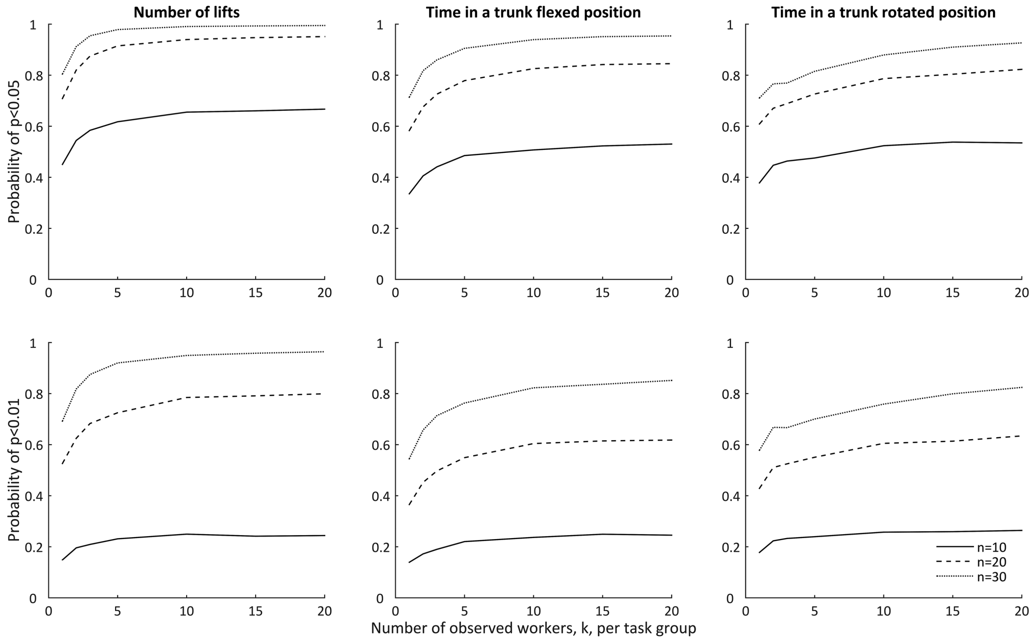

Figure 4 Statistical power, i.e. the probability of obtaining a significant OR for the association between exposure and outcome, for all 21 investigated combinations of n (different lines in each panel) and k (x-axis). Upper and lower panels: significance levels $P < 0.05$ and $P < 0.01$, respectively. Probabilities of obtaining a significant OR are shown for the exposure variables: number of lifts (left panels), time with the trunk flexed posture (middle panels), and time with a rotated trunk (right panels).

the attenuation associated with using individual-based strategies (Reeves *et al.*, 1998; Tielemans *et al.*, 1998) may be valid when linear regression is used to model the exposure–outcome relationship, but less so for logistic regression models. The study by Kim *et al.* (2011) predicts that small group-based studies with large exposure variability within groups and small variability between groups will be particularly prone to show bias. Our results showed that the bias in OR was not very large when exposure was measured on more than five workers per group, and we did not see any obvious difference in the magnitude of bias between the three exposure variables even though one of them (time in rotated trunk posture) exhibited a much smaller exposure contrast between groups than the other two.

Our results confirmed the trivial conclusion reported by many studies that a more precise group mean exposure estimate will be obtained when data are collected from more workers (Allread *et al.*, 2000; Hoozemans *et al.*, 2001; Mathiassen *et al.*, 2005), up to the point where adding more workers will lead to very marginal improvements of precision, and therefore even marginal increases in study power. Our study adds to these well-known effects of increased sample sizes by showing a 'saturation' effect of increasing exposure sample sizes even for bias in OR estimates (Fig. 3).

Similarly, the outcome estimate (*in casu*, LBP prevalence) will get more precise when more workers are included in a study, reaching a satisfying precision at a particular number of workers, beyond which further investments may not be justified. Notably, though, LBP shows a high prevalence (i.e. in our population a 44% LBP prevalence in 3 years of follow-up) compared to many other outcomes in occupational studies. While we do not have reasons to believe that the general structure of the relationship between number of workers and precision of the outcome estimate will depend on the true prevalence of that outcome, we would expect that fewer workers are needed to obtain a satisfying precision if the prevalence is small, given that the occurrence of disorders in a population follows a binomial distribution.

Methodological considerations

One limitation of the present study is that the video recordings of each particular worker in the exposure measurement population were collected during one single day, if at four randomly chosen occasions. Distributing these four recordings across more days would likely have resulted in more certain individual exposure estimates, given that exposure probably varied between days (Liv *et al.*, 2011; Kwon *et al.*, 2011; Trask *et al.*, 2012). Thus, collecting exposure data over multiple days per worker instead of only one could have led to slightly different results in our study. Since the parent dataset did not allow determination of between-days exposure variability, we cannot assess this effect in quantitative terms. However, we find it reasonable to claim that the conclusions of our study concerning the generic effects of increasing the sizes of observed and non-observed study populations would be valid even for data collections distributed across multiple days per worker.

The parent study experienced a substantial loss of participants during the 3-year follow-up period: 1802 workers entered the study at baseline, but only 1131 (63%) were still present at follow-up, and thus available for the present parent dataset. This loss to follow-up could, in theory, have led to selection or attrition bias. However, descriptive statistics (Table 1) showed that the workers in the parent dataset did not differ to any notable extent from the eligible population in terms of gender, age, working hours a week, and percentage of workers with LBP at baseline.

An additional limitation is that our *a-priori* expert classification of jobs into occupational groups may not have been optimal. The grouping scheme may influence the outcomes of a study (van Tongeren *et al.*, 1997; Symanski *et al.*, 2006), for instance in terms of effectiveness in reducing attenuation of an exposure–outcome relationship (Werner and Attfield, 2000). In the present study, jobs were carefully allocated *a priori* to occupational groups on basis of their estimated exposure profiles by the same experienced ergonomists who also collected the video recordings. According to the exposure contrast values, classification was reasonably successful for the variables number of lifts and time in flexed postures, while it was less successful for rotated trunk postures. Thus, the basic purpose of a grouping scheme, i.e. to effectively partition the overall between-subjects exposure variance to occur between groups rather than within groups, was only obtained to some extent, as also illustrated by the substantial between-subject exposure variability in some of the occupational groups (Table 1). Our parent data did, however, not allow for any attempts to group workers according to alternative principles based on occupation (Jansen and Burdorf, 2003; Svendsen *et al.*, 2005), task (Mathiassen *et al.*, 2005; Tak *et al.*, 2011), expected exposure (Hoogendoorn *et al.*, 2000; Ariëns *et al.*, 2001), or even hybrids combining individual exposure information with a classification by group (Seixas and Sheppard, 1996). In extension, we could not examine whether any alternative classification would result in more distinct associations than those found in the present study between the investigated exposures and LBP, and in different numerical effects of changing the sizes of the populations with exposure and outcome data. We do, on the other hand, believe that study size and number of observed workers will, indeed, influence bias and power even in group-based studies on other populations and using different grouping schemes than the one used in our study, if with different numerical results.

The present study addressed three exposure variables, i.e. lifting, trunk flexion, and trunk rotation. In our parent dataset, these three exposure variables correlated only weakly. Therefore, it was justified to assess the effect of these three exposures on LBP independently of each other. While we cannot explicitly extend our results to other exposures, let alone other outcomes, we do claim that the consistency of our results across three weakly correlated exposures suggests that the generic effects of sample sizes on ORs and power may hold a fair external validity. We are further supported in this conviction by noting that our results were consistent across these three exposures even though they differ substantially in both absolute and relative sizes of within- and between-subjects variability (Table 1).

A final limitation is that, in the current bootstrapping procedure, samples of workers were drawn with replacement from each group. Therefore, it was possible to 'oversample' workers (i.e. obtaining a virtual sample of workers that was larger than the number of unique workers available in the group). Oversampling by >100% (i.e. sampling at least twice as many workers as available in the parent data) occurred in 4 out of 19 groups when selecting $k = 20$ workers for the exposure estimates, while it did not occur for values of k between 1 and 15, and not in any case of sampling the n workers providing LBP data. We have not been able to identify any discussions in the bootstrapping literature on the acceptability and limits of oversampling, let alone its possible effects on the resulting data structure. However, it is reasonable to assume that effects of oversampling are more prominent if the parent dataset is small and/or irregularly distributed. We restricted our parent dataset to groups represented by at least 8 observed workers and 21 workers in total (Table 1) to get a fair representation of workers in the group, and thus, among other benefits, reduce the possible effect of oversampling. Since results from the sampling strategies containing oversampled exposure data are in line with results from strategies where no oversampling occurred (Fig. 4), we do not expect serious effects of oversampling in our study.

Conclusion

The statistical power of an exposure–outcome study design using group-based exposure estimation in logistic regression depended more on the total number of workers included in the study (with personalized outcome data) than on the size of the population from which exposure estimates were obtained. When, however, exposure was observed on very few workers, the OR of the exposure–outcome relationship was downward biased irrespective of the total population size. Our findings suggest that (costly) exposure observations are necessary only on few workers, provided that the overall size of the study population is sufficiently large and everybody is followed up with respect to outcome. These results may contribute to a more efficient use of resources in future epidemiological studies on exposure–outcome associations.

FUNDING

The present cooperation was supported by a grant from the Swedish Research Council for Health, Working Life and Welfare (Forte Dnr. 2009-1761).

ACKNOWLEDGEMENTS

The direct and indirect support parties played no role in preparing the research material, writing, reviewing, or approving the submitted manuscript. There were no conflicts of interest.

REFERENCES

Allread WG, Marras WS, Burr DL. (2000) Measuring trunk motions in industry: variability due to task factors, individual differences, and the amount of data collected. *Ergonomics*; 43: 691–701.

Ariëns GA, Bongers PM, Douwes M *et al.* (2001) Are neck flexion, neck rotation, and sitting at work risk factors for neck pain? Results of a prospective cohort study. *Occup Environ Med*; 58: 200–7.

Armstrong BG. (1998) Effect of measurement error on epidemiological studies of environmental and occupational exposures. *Occup Environ Med*; 55: 651–6.

Bakker EW, Verhagen AP, van Trijffel E *et al.* (2009) Spinal mechanical load as a risk factor for low back pain: a systematic review of prospective cohort studies. *Spine (Phila Pa 1976)*; 34: E281–93.

Burdorf A, Jansen JP. (2006) Predicting the long term course of low back pain and its consequences for sickness absence and associated work disability. *Occup Environ Med*; 63: 522–9.

Burdorf A, van Riel M. (1996) Design of strategies to assess lumbar posture during work. *Int J Ind Ergonom*; 18: 239–49.

Coenen P, Kingma I, Boot CR *et al.* (2013) Cumulative low back load at work as a risk factor of low back pain: a prospective cohort study. *J Occup Rehabil*; 23: 11–8.

David GC. (2005) Ergonomic methods for assessing exposure to risk factors for work-related musculoskeletal disorders. *Occup Med (Lond)*; 55: 190–9.

Davison AC, Hinkley DV. (1997) *Bootstrap methods and their applications*. Cambridge: Cambridge University Press.

Efron B, Tibshirani R. (1986) Bootstrap methods for standard errors, confidence intervals, and other measures of statistical accuracy. *Stat Science*; 1: 54–77.

Griffith LE, Shannon HS, Wells RP *et al.* (2012) Individual participant data meta-analysis of mechanical workplace risk factors and low back pain. *Am J Public Health*; 102: 309–18.

Heederik D, Attfield M. (2000) Characterization of dust exposure for the study of chronic occupational lung disease: a comparison of different exposure assessment strategies. *Am J Epidemiol*; 151: 982–90.

Hoogendoorn WE, Bongers PM, de Vet HC *et al.* (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 (Phila Pa 1976)*; 25: 3087–92.

Hoozemans MJ, Burdorf A, van der Beek AJ *et al.* (2001) Group-based measurement strategies in exposure assessment explored by bootstrapping. *Scand J Work Environ Health*; 27: 125–32.

Houba R, Heederik D, Kromhout H. (1997) Grouping strategies for exposure to inhalable dust, wheat allergens and alpha-amylase allergens in bakeries. *Ann Occup Hyg*; 41: 287–96.

Jansen JP, Burdorf A. (2003) Effects of measurement strategy and statistical analysis on dose-response relations between physical workload and low back pain. *Occup Environ Med*; 60: 942–7.

Kim HM, Park CG, van Tongeren M *et al.* (2012) Estimation methods with ordered exposure subject to measurement error and missingness in semi-ecological design. *BMC Med Res Methodol*; 12: 135.

Kim HM, Richardson D, Loomis D *et al.* (2011) Bias in the estimation of exposure effects with individual- or group-based exposure assessment. *J Expo Sci Environ Epidemiol*; 21: 212–21.

Kromhout H, Heederik D. (1995) Occupational epidemiology in the rubber industry: implications of exposure variability. *Am J Ind Med*; 27: 171–85.

Kromhout H, Loomis DP, Mihlan GJ *et al.* (1995) Assessment and grouping of occupational magnetic field exposure in five electric utility companies. *Scand J Work Environ Health*; 21: 43–50.

Kuorinka I, Jonsson B, Kilbom A *et al.* (1987) Standardised Nordic questionnaires for the analysis of musculoskeletal symptoms. *Appl Ergon*; 18: 233–7.

Kwon BK, Roffey DM, Bishop PB *et al.* (2011) Systematic review: occupational physical activity and low back pain. *Occup Med (Lond)*; 61: 541–8.

Liv P, Mathiassen SE, Svendsen SW. (2011) Theoretical and empirical efficiency of sampling strategies for estimating upper arm elevation. *Ann Occup Hyg*; 55: 436–49.

Loomis D. (2012) On the importance of quantification. *Occup Environ Med*; 69: 609.

Loomis D, Kromhout H. (2004) Exposure variability: concepts and applications in occupational epidemiology. *Am J Ind Med*; 45: 113–22.

Lötters F, Burdorf A, Kuiper J *et al.* (2003) Model for the work-relatedness of low-back pain. *Scand J Work Environ Health*; 29: 431–40.

Mathiassen SE, Burdorf A, van der Beek AJ. (2002) Statistical power and measurement allocation in ergonomic intervention studies assessing upper trapezius EMG amplitude. A case study of assembly work. *J Electromyogr Kinesiol*; 12: 45–57.

Mathiassen SE, Möller T, Forsman M. (2003) Variability in mechanical exposure within and between individuals performing a highly constrained industrial work task. *Ergonomics*; 46: 800–24.

Mathiassen SE, Nordander C, Svendsen SW *et al.* (2005) Task-based estimation of mechanical job exposure in occupational groups. *Scand J Work Environ Health*; 31: 138–51.

Mathiassen SE, Paquet V. (2010) The ability of limited exposure sampling to detect effects of interventions that reduce the occurrence of pronounced trunk inclination. *Appl Ergon*; 41: 295–304.

Paquet V, Punnett L, Woskie S *et al.* (2005) Reliable exposure assessment strategies for physical ergonomics stressors in construction and other non-routinized work. *Ergonomics*; 48: 1200–19.

Peretz C, de Pater N, de Monchy J *et al.* (2005) Assessment of exposure to wheat flour and the shape of its relationship with specific sensitization. *Scand J Work Environ Health*; 31: 65–74.

Punnett L, Wegman DH. (2004) Work-related musculoskeletal disorders: the epidemiologic evidence and the debate. *J Electromyogr Kinesiol*; 14: 13–23.

Reeves GK, Cox DR, Darby SC *et al.* (1998) Some aspects of measurement error in explanatory variables for continuous and binary regression models. *Stat Med*; 17: 2157–77.

Searle SR, Casella G, McCulloch CE. (1992) *Variance components*. New York: John Wiley & Sons.

Seixas NS, Sheppard L. (1996) Maximizing accuracy and precision using individual and grouped exposure assessments. *Scand J Work Environ Health*; 22: 94–101.

Steenland K, Deddens JA, Zhao S. (2000) Biases in estimating the effect of cumulative exposure in log-linear models when estimated exposure levels are assigned. *Scand J Work Environ Health*; 26: 37–43.

Svendsen SW, Bonde JP, Mathiassen SE *et al.* (2004) Work related shoulder disorders: quantitative exposure-response relations with reference to arm posture. *Occup Environ Med*; 61: 844–53.

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: 18–27.

Symanski E, Maberti S, Chan W. (2006) A meta-analytic approach for characterizing the within-worker and between-worker sources of variation in occupational exposure. *Ann Occup Hyg*; 50: 343–57.

Tak S, Buchholz B, Punnett L *et al.* (2011) Physical ergonomic hazards in highway tunnel construction: overview from the Construction Occupational Health Program. *Appl Ergon*; 42: 665–71.

Tielemans E, Kupper LL, Kromhout H *et al.* (1998) Individual-based and group-based occupational exposure assessment: some equations to evaluate different strategies. *Ann Occup Hyg*; 42: 115–9.

Trask C, Mathiassen SE, Wahlström J *et al.* (2012) Data collection costs in industrial environments for three occupational posture exposure assessment methods. *BMC Med Res Methodol*; 12: 89.

Twisk JWR. (2003) *Applied longitudinal data analysis for epidemiology: a practical guide*. New York: Cambridge University Press.

van der Beek AJ, Frings-Dresen MH. (1998) Assessment of mechanical exposure in ergonomic epidemiology. *Occup Environ Med*; 55: 291–99.

van Tongeren M, Gardiner K, Calvert I *et al.* (1997) Efficiency of different grouping schemes for dust exposure in the European carbon black respiratory morbidity study. *Occup Environ Med*; 54: 714–19.

Vlaanderen J, Vermeulen R, Heederik D *et al.* (2008) Guidelines to evaluate human observational studies for quantitative risk assessment. *Environ Health Perspect*; 116: 1700–5.

Werner MA, Attfield MD. (2000) Effect of different grouping strategies in developing estimates of personal exposures: specificity versus precision. *Appl Occup Environ Hyg*; 15: 21–5.

Winkel J, Mathiassen SE. (1994) Assessment of physical work load in epidemiologic studies: concepts, issues and operational considerations. *Ergonomics*; 37: 979–88.