# Chapter 5

## Effects of the data sampling strategy on bias and power in epidemiologic studies of low-back pain – a bootstrapping approach

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

Submitted

## ABSTRACT

Studies of work-related low-back pain (LBP) often classify workers into exposure groups for which e.g., lifting or awkward trunk postures are estimated from measurements on a sub-population. The present study investigated combined influences of the sizes of the total study population and the sub-sample on exposure-outcome associations.

At baseline, lifting, trunk flexion, and trunk rotation was observed for 371 of 1131 workers in 19 task groups. Self-reported LBP (dichotomous) was obtained from all workers during three years of follow-up. All three exposures were associated with LBP ($p<0.01$) according to logistic regression.

All possible combinations of $n=10,20,30$ workers per task group and $k=1,2,3,5,10,15,20$ workers being observed were investigated using bootstrapping. The OR and its p-value was determined for each of 10,000 virtual studies at each combination of n and k, and the average OR and the statistical power ($p<0.05$ and $p<0.01$) across the 10,000 studies were assessed.

For lifts and flexed trunk, studies including $n\geq20$ workers in each task group, and $k\geq5$ observed, led to an almost unbiased OR and a power $>0.80$ (p-level 0.05). A similar performance required $n\geq30$ workers for rotated trunk. Small numbers, k, of observed workers resulted in biased OR, while power was, in general, more sensitive to the total number, n, of workers than to the number, k, of observed workers.

In a group-based exposure assessment strategy, statistical performance may be sufficient if the overall size of the groups is reasonably large, even if exposure is estimated of few workers per group.






## INTRODUCTION

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

Several studies on occupational physical risk factors for LBP have adopted a group-based exposure assessment strategy (Ariëns et al., 2001; Burdorf & Jansen, 2006; Hoogendoorn et al., 2000a). Workers are then classified into groups with an expected contrast in exposure, typically based on their job or tasks. The exposure variable(s) of interest is measured only in a sub-sample of workers within each group, and the resulting 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 (i.e., LBP) from all subjects in the study population. This exposure assessment strategy is based on the assumption that workers within the same group have similar exposures, i.e. that the groups are homogeneous with respect to exposure, and that exposure variability between groups is comparatively large, so that the exposure contrast between groups will be substantial (Kromhout & Heederik, 1995; Mathiassen et al., 2005).

The effect of the number and allocation of exposure measurements on the statistical properties of a group mean exposure estimate is relatively well documented (Hoozemans et al., 2001; Liv et al., 2010; Mathiassen et al., 2002; Mathiassen et al., 2003a). However, the influence of measurement strategies on the strength and statistics of exposure-outcome associations in logistic regression has, to the best of our knowledge not been thoroughly investigated. A theoretical framework has been presented on the issue of bias and precision in linear regression of continuous outcomes on (continuous) exposure measured with random uncertainty (Tielemans et al., 1998), and even logistic regression has been discussed in this context (Reeves et al., 1998). However, the case of estimating exposure in group(s) from observations of a sub-population while using personal outcome data has not been addressed in any of these studies. Also empirical data to complement theoretical findings have not been presented. Therefore, the present study aimed to assess the combined effect of the sample size of the total population and that of the sample on which exposure is actually observed on exposure-outcome associations in a study of occupational physical exposures and LBP.

## STUDY POPULATION AND METHODS

*Population*

The present study is based on data from the Study on Musculoskeletal disorders, Absenteeism and Health (SMASH). As described in detail previously (Coenen et al., 2013b; Hoogendoorn et al., 2000a), 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 these workers. These 1802 workers were classified by experts into 23 task groups, based on their expected physical work load. Within each task group, work was recorded on video from a random sample of roughly one fourth of the workers. After excluding workers dropping out after the baseline measurements, the parent data set for the current study included 1131 workers from those 19 task groups that contained more than 5 observed workers. Video based observation data were available from, in total, 371 workers (Table 5.1).

*Exposure and outcome for the parent data set*

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 minutes 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., 2000a); i.e., the number of lifts during an eight hour 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). The mean exposure of the observed workers in each of the 19 task groups was assigned to all workers classified into that group. In order to evaluate the task group classification, between-group contrasts for each of the three exposure risk factors were calculated, using:

$$\text{Contrast} = \frac{MSE_b}{(MSE_b + s_w)} \qquad \text{Equation 5.1}$$

In which $MSE_b$ is the mean squared error between task groups and $s_w$ is the variability between workers within groups (Kromhout & Heederik, 1995; Mathiassen et al., 2005).

Self-reported LBP was assessed for all 1131 workers once a year for three 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 three years of follow-up, regardless of baseline status.

Logistic regression analyses using the three exposure variables 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 were executed. Results showed both the number of lifts (per 100 lifts; OR: 1.06 (95%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), to be significantly associated with LBP in the parent data set.

*Simulated sampling strategies*

For all 21 possible combinations of n=10,20,30 workers in total per task 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 as follows (Efron & Tibshirani, 1986; Hoozemans et al., 2001; Liv et al., 2010; Paquet et al., 2005). Within each task group of the parent data set, workers were identified as "observed" and "non-observed" depending on whether exposure data were available or not. For each combination of n and k, k workers in each task group were drawn with replacement from the group of observed workers, and n workers were drawn with replacement from all workers (observed and non-observed combined) in the same task group. This led to a virtual study including n workers in total and k observed workers from each task group. For each virtual study, the three mean exposures (number of lifts, trunk flexion, and trunk rotation) of the k observed workers within each task group were then assigned to all n workers in that particular task 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 data set. For each of the 21 possible combinations of n and k, 10,000 virtual studies were constructed using this procedure. Four measures for each investigated exposure assessment strategy were obtained on the basis of the 10,000 virtual study results, i.e. 1) a pooled estimate of the standard deviation (SD) of the mean exposure estimate within a task group, obtained by first calculating the mean variance between subjects, $VAR_{BS}$, across the 10,000 replicates of that variance for each specific task group, and then pooling these 19 variances into the average SD of a mean exposure estimate according to the formula:

$$\text{Pooled SD} = \sqrt{\frac{\text{mean}(VAR_{BS})}{k}} \qquad \text{Equation 5.2}$$

2) The SD across the 10,000 studies of the LBP prevalence in the population, 3) the mean OR across the 10,000 studies, and 4) the power in 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 an OR with the mentioned significances. All calculations were performed using customary scripts in Matlab (MATLAB 7.7.0, The MathWorks Inc., Natick, MA, 2000). Logistic regression analyses were implemented using the Matlab statistical toolbox.

**Table 5.1** | Parent data set. In the upper panel, the total number of workers (N) and the number of workers observed (K) are shown for each task group, together with LBP prevalence among all workers, and group mean exposures and standard deviations for the three investigated physical exposures: number of lifts at work per week, percent time spent with the trunk flexed more than 30°, and percent time spent with the trunk rotated more than 30°.

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 all workers (N) and for those observed (K).

| | Workers | LBP | Observed | Lifts | | Flexion | | Rotation | |
|---|---|---|---|---|---|---|---|---|---|
| **Description task groups** | (N) | | (K) | Mean | StDev | Mean | StDev | Mean | StDev |
| Mainly sitting work | | | | | | | | | |
| Sitting with varying postures | 133 | 39% | 61 | 23.2 | 83.2 | 6.3 | 9.0 | 12.3 | 29.2 |
| Sitting with little varying postures (computer work) | 57 | 35% | 16 | 13.4 | 51.9 | 7.4 | 10.5 | 1.1 | 2.2 |
| Sitting with little varying postures, in awkward postures (no computer work) | 31 | 68% | 11 | 1.1 | 3.5 | 3.6 | 8.9 | 2.9 | 2.7 |
| Sitting with little varying postures, with repetitive movements | 95 | 42% | 31 | 334.3 | 933.7 | 2.4 | 3.1 | 2.4 | 3.1 |
| | | | | | | | | | |
| Mainly standing work | | | | | | | | | |
| Standing with varying postures (including walking) without external forces | 26 | 58% | 9 | 8.0 | 18.8 | 4.1 | 3.5 | 2.2 | 2.8 |
| Standing with varying postures and small external forces | 69 | 38% | 23 | 658.9 | 781.8 | 7.2 | 5.4 | 2.4 | 2.9 |
| Standing with varying postures and moderate external forces | 87 | 44% | 28 | 438.1 | 521.5 | 10.0 | 8.9 | 5.6 | 4.4 |
| Standing with varying postures and large external forces | 65 | 40% | 20 | 299.5 | 283.6 | 11.5 | 6.5 | 6.2 | 5.5 |
| Standing with varying, awkward postures and moderate external forces | 66 | 50% | 22 | 544.4 | 620.0 | 13.7 | 9.1 | 6.1 | 4.2 |
| | | | | | | | | | |
| Awkward postures (mainly static exposure) | | | | | | | | | |
| Standing in static awkward posture without external forces | 42 | 48% | 15 | 133.6 | 177.9 | 8.4 | 7.3 | 4.3 | 3.9 |
| Standing in static awkward posture with small external forces | 70 | 39% | 24 | 194.7 | 277.2 | 10.3 | 6.7 | 6.7 | 5.8 |
| Mainly static back exposures by alternating awkward postures | 28 | 61% | 11 | 814.8 | 1167.3 | 37.6 | 30.7 | 12.1 | 6.8 |
| **Continuation of table 5.1** | | | | | | | | | |
| Alternating exposures (standing, walking and/or sitting) | | | | | | | | | |
| Alternating standing, walking and/or sitting without external forces | 167 | 40% | 29 | 6.4 | 32.1 | 5.7 | 6.1 | 2.4 | 3.4 |
| Alternating standing, walking and/or sitting with small external forces | 36 | 50% | 13 | 82.2 | 71.1 | 8.9 | 5.9 | 4.2 | 5.9 |
| Alternating standing, walking and/or sitting with moderate external forces | 52 | 42% | 15 | 312.9 | 179.8 | 22.0 | 12.0 | 4.7 | 5.8 |
| Alternating standing, walking and/or sitting with large external forces | 21 | 86% | 8 | 2904.0 | 1483.9 | 42.5 | 15.7 | 19.8 | 8.1 |
| Alternating standing and walking in static awkward postures, external forces | 27 | 44% | 17 | 379.2 | 433.4 | 19.2 | 11.8 | 6.8 | 5.9 |
| Alternating standing and walking in postures, moderate external forces | 36 | 56% | 9 | 577.2 | 275.8 | 12.8 | 6.7 | 7.4 | 3.9 |
| | | | | | | | | | |
| Combined functions (as a result of changes in tasks) | | | | | | | | | |
| Combined exposures | 23 | 30% | 9 | 252.2 | 297.0 | 8.4 | 7.5 | 2.4 | 2.3 |
| **Total** | 1131 | 44% | 371 | | | | | | |
| | | | | | | | | | |
| **Descriptive variables** | | | | | | | | | |
| Number of workers | 1131 | | 371 | | | | | | |
| Number of males | 699(69%) | | 219(68%) | | | | | | |
| Number of females | 307(31%) | | 104(32%) | | | | | | |
| Stature (cm) | 175.9(9.6) | | 175.7(9.4) | | | | | | |
| Weight (kg) | 75.9(13.6) | | 74.9(12.3) | | | | | | |
| Age (years) | 35.5(8.8) | | 35.7(8.8) | | | | | | |
| Working hours per week | 37.2(6.9) | | 36.7(6.9) | | | | | | |
| Years of employment | 9.7(7.5) | | 9.4(7.3) | | | | | | |
| Baseline LBP (number of workers) | 366(38%) | | 125(40%) | | | | | | |

## RESULTS

Exposure contrasts between groups were 0.55, 0.48 and 0.23 for the number of lifts, time in flexed trunk posture and time in rotated trunk posture, respectively. While task groups did, indeed, differ in mean exposure (Table 5.1), some were very heterogeneous in terms of the workers differing substantially in exposure.

For all three exposure variables, the pooled SD of the group mean exposure decreased as the number of workers, k, for which exposure was actually observed increased (Figure 5.1). This confirmed that more data lead to more precise exposure estimates. Obviously, this effect did not depend on the total number of workers, n, per task group. The SD of the prevalence of LBP in the study population decreased with an increasing total number of workers, n, included in each task group (Figure 5.2), and obviously this effect did not depend on k. The average OR of the association between exposure and LBP increased with larger k (Figure 5.3), while it was affected only little by the total number of workers, n.

Figure 5.4 shows that power increased with both n and k. 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 number of lifts and time with flexed trunk, a power of 0.80 to detect a significant ($p<0.05$) OR was obtained when at least n=20 workers were included per task group, and the number of actually observed workers in each task group (k) was at least 5. For time working with the trunk rotated, at least n=30 workers per task group were needed to obtain the same power. At the more strict requirement of $p<0.01$, a power of 0.80 was obtained only when the population included at least n=30 workers per task group for lifts and flexed trunk, while this level of power could not be reached at all for the risk factor time working in a trunk rotated posture.

## DISCUSSION

The present study dealt with the common group-based assessment strategy in musculoskeletal epidemiology of measuring exposure to risk factors in a sub-population of workers. Mean exposure estimates are then assigned to all workers having similar tasks or jobs, while information on outcomes is available from each individual worker in the total study population. Our study suggests that the probability of finding significant exposure-outcome associations depends more on the total number of workers included in each task group than the number of workers for whom exposure is actually observed. In our setting comprising 19 task groups intended to represent the general working population, studies including at least 30 workers in each group and basing the task group exposures on at least 5 observed workers were sufficient to secure a reasonable power and an almost unbiased estimate of the odds ratio. However, the exact numbers of subject to establish a certain statistical performance differed between the three investigated exposure risk factors (Figures 5.3 and 5.4). Our results may have important implications for future epidemiological studies, since they suggest that a limited research budget would 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 Figure 5.4, a statistical power around 0.80 ($p<0.05$) can be reached either by a study design comprising 20 workers per task group and only one is actually observed and by a study including 10 workers per task group, and 10 need to be observed. Thus, the “large” study requires outcome data to be collected from 380 workers, but exposure only from 19, while the “small” study is based on outcome data from only 190 workers, but exposure data from all 190. While the budgets of these two alternatives depend on the unit cost of obtaining exposure and outcome information, it seems likely that the “large” study is cheaper to realize. Notably, while these two sampling strategies have comparable abilities to detect a significant association between exposure and LBP, the former will, however, result in a more biased OR (Figure 5.3).

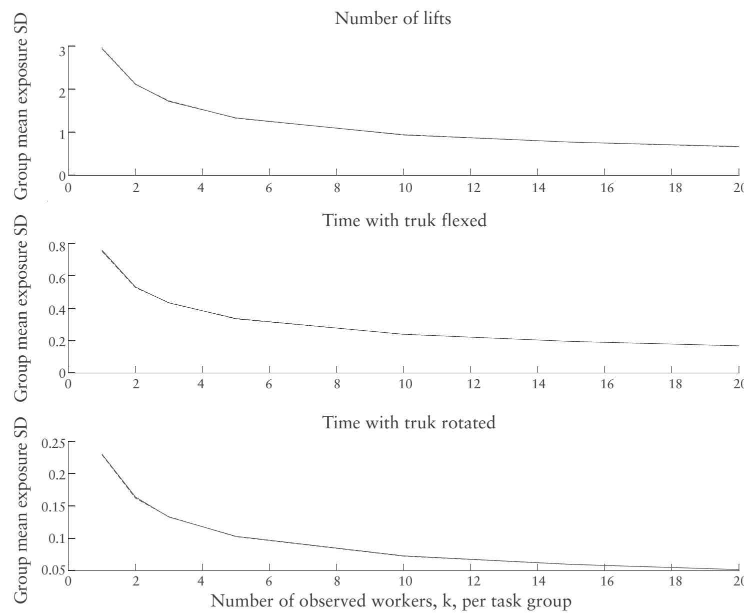


**Figure 5.1** | Pooled estimate of the standard deviation (SD) of the group mean exposure in a task group for each of the 21 combinations of n (different lines) and k (x-axis). SD is presented for the exposure variables: number of lifts (upper panel), time with the trunk flexed (middle panel) and time with a rotated trunk (lower panel). Note that the individual curves for different n-values in each panel overlap completely.

As confirmed by our results, more precise (i.e. more certain) group mean exposure estimates will be obtained when data are collected from more workers. Several studies (Allread et al., 2000; Hoozemans et al., 2001; Mathiassen et al., 2005) have shown that the exposure estimate improves still less when still more workers are included in the estimate. Thus, beyond a certain number of observed workers, it may not be warranted to invest more resources in observing even more workers. Similarly, the estimate of the outcome (i.e., the LBP prevalence) will become more precise when more workers are included in a study, and may reach a sufficient precision at a particular number of workers, beyond which further investments may not be justified.

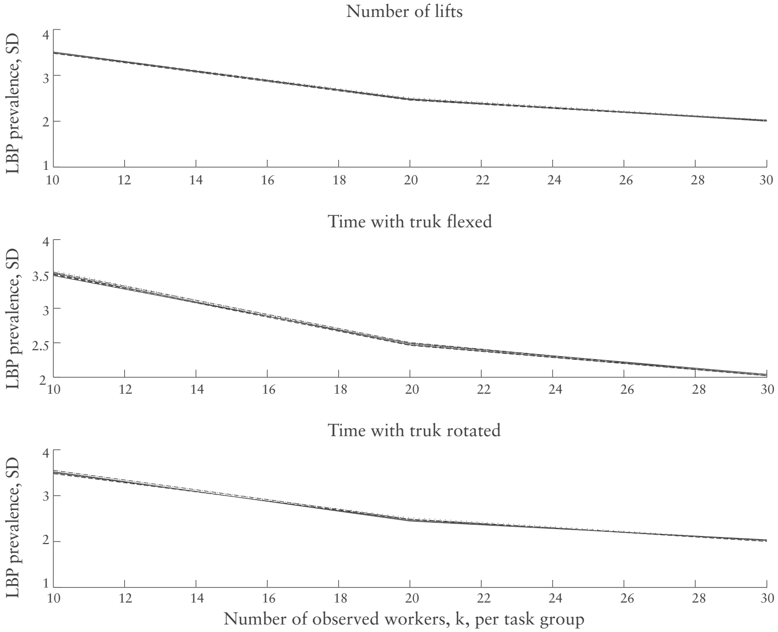


**Figure 5.2** | Standard deviation (SD) of the outcome (i.e. LBP prevalence in the entire data set) across the 10,000 replicates for each of the 21 investigated combinations of n (x-axis) and k (different lines). Standard deviations are presented for the exposure variables: number of lifts (upper panel), 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.





The decrease in average ORs with lower numbers of k, i.e. an attenuation of the OR towards 1, is probably a result of increased uncertainty in the estimate of task group exposures, since the OR was, only weakly influenced by the overall number of workers, n, in each task group. Attenuation of exposure-outcome regression coefficients due to uncertainty in the exposure estimates also occurs in simple linear regression of two continuous variables (Tielemans et al., 1998)Non-U.S. , as well as in logistic regression (Reeves et al., 1998), even though a group-based exposure assessment strategy is generally regarded to be an effective measure to avoid biased regression coefficients, in particular in linear regression (25). Our results showed that the bias was, however, not very strong, and only weakly influenced by the overall number of workers in each task group, n.

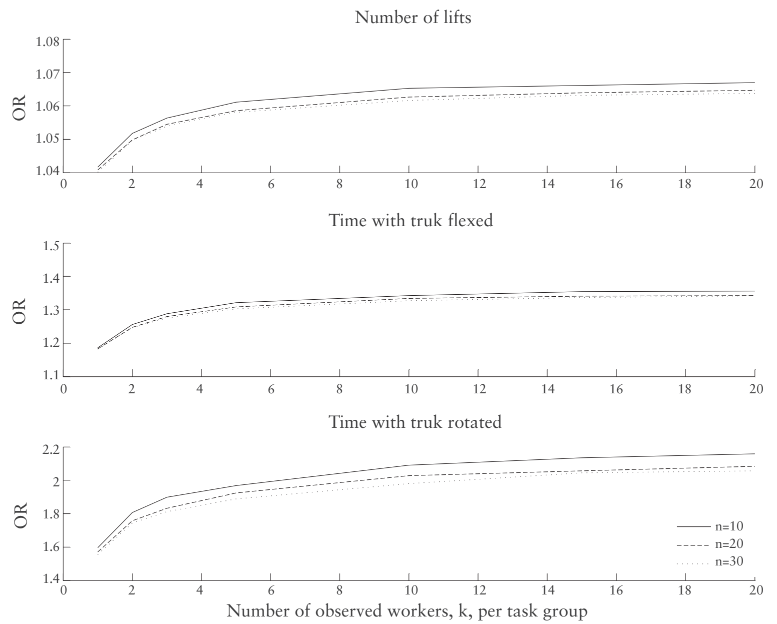


**Figure 5.3** | Average odds ratios (OR) for the association between exposure and LBP across the 10,000 replicates for each of the 21 investigated combinations of n (different lines) and k (x-axis). Average ORs are presented for the exposure variables: number of lifts (upper panel), time with the trunk flexed (middle panel) and time with a rotated trunk (lower panel).

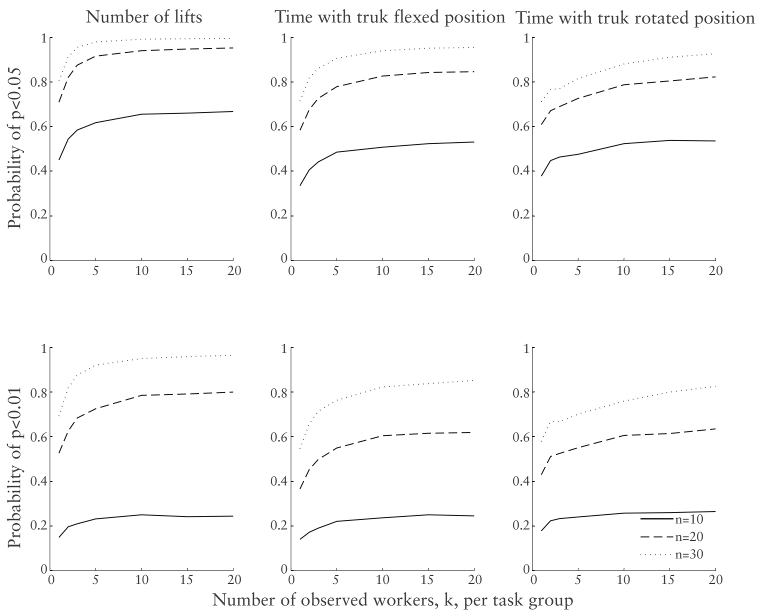


**Figure 5.4** | Statistical power, i.e. the probability of obtaining a significant OR for the association between exposure and outcome, for all 21 investigated combinations of n (different lines in each panel) and k (x-axis). Upper and lower panels: significance levels $p<0.05$ and $p<0.01$, respectively. 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).

Earlier occupational studies as well as statistical textbooks present equations to calculate the power of a study protocol to obtain statistically significant results, as a function of sample sizes and variability (e.g., of exposures) in the study population (Mathiassen et al., 2002; Mathiassen et al., 2003b; Twisk, 2003). While this literature discusses comparatively simple study designs, the present study confirms the general effect of more data improving power for a more complex design. Our study also adds the observation that the size of k does have an effect on power, but that this effect is weaker than that of changing the total number of workers (Figure 5.4).





In the present study, the video recordings of each particular worker were collected at four randomly chosen occasions during the course of one single day. This may be considered a less efficient choice, since distributing these four occasions over several days would likely have resulted in a more certain exposure estimate for that worker, given that exposure probably varied between days within workers (Hoogendoorn et al., 2000a; Kwon et al., 2011; Liv et al., 2010; Paquet et al., 2005; Twisk, 2003). More certain estimates of the exposures of individual workers in a task group would even lead to a more certain mean exposure estimate for the task group as a whole. Thus, collecting exposure data over multiple days per worker could have led to slightly different conclusions. For example, it might have been necessary to observe less workers to obtain the same exposure-outcome associations as what is now obtained with, for instance, k=5 workers in each task group. However, since the uncertainty of the exposure estimates for individual workers is expected to be the result of random statistical processes, the general conclusions of our study would not change.

An expert classification of tasks (jobs) into groups, based on suspected physical workloads, may result in a grouping scheme that does not effectively capture exposure differences between workers in different tasks. Thus, as it appears even in our material, exposure variability between subjects may be considerable within several of the task groups (Table 5.1), and another categorization of some workers might have resulted in more homogeneous task groups. Task groups were carefully set up by the same trained observers who also collected the video recordings, based on their extensive experience of physical work load assessment in occupational settings. According to the exposure contrast values, classification was reasonably successful for the two variables number of lifts and time in flexed postures. For time working in rotated trunk posture, the contrast was lower, mainly due to task group 1 being very heterogeneous (Table 5.1). The latter is a possible explanation that power was generally less for exposure-outcome relationships based on this risk factor (Figure 5.4). Whether a different grouping scheme, with less or more task groups, possibly defined using other criteria, could have been more effective in disclosing exposure-outcome associations for LBP is an open question. Therefore, studies employing other grouping schemes might reach different results as to the statistical performance of sampling strategies than we did. However, we believe that the trade-off between total study size and number of observed workers would be a consistent finding. Moreover, over results suggest that classifications in future studies of tasks and jobs according to expected exposures could benefit from more comprehensive a priori knowledge. As an example, a pilot study in which observational data of a limited amount of workers is collected and analyzed to identify an optimal classification a priori to the full study could probably lead to a more informed and more effective classification.

The present study addressed only three exposure variables (i.e., lifting, trunk flexion and trunk rotation). In our parent data set, these three exposure variables correlated only weakly, with correlation coefficients of 0.34, 0.09 and 0.09, for lifting *vs.* flexion, lifting *vs.* rotation, and flexion *vs.* rotation, respectively. Therefore, it seems reasonable both to assess the effect of these three exposures on LBP independently of each other and to assume that our general results may apply even to other variables describing trunk exposure, i.e. that the results show a fair external validity.

The present simulations were constructed to include the same number of workers from each task group in a balanced study design. This may have affected exposure-outcome associations, as compared to the more usual situation in epidemiologic studies (and in our parent data set) of groups being of different sizes. As a general rule, the statistical power of a balanced study design will be larger than that of an unbalanced design with the same total number of workers, and so the exposure-outcome associations of our simulated study designs are probably stronger and more precise than those in comparable unbalanced designs of the same total magnitude.

In the current bootstrapping procedure, samples of workers were drawn with replacement from each task 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 more than 100% (i.e., sampling at least twice as many workers as available in the parent data) occurred in 4 out of 19 task 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 either in any case of sampling the n workers providing LBP data. We have not been able to identify any discussion in the bootstrapping literature on the acceptability and limits of oversampling, let alone its possible effects on the resulting data distributions (Davison & Hinkley, 1997; Efron & Tibshirani, 1986). However, it is reasonable to assume that effects of oversampling are more prominent if the parent data is small and/or irregularly distributed. We restricted our parent data set to task groups represented by at least 5 observed workers and 21 workers in total (Table 5.1) in order to get a fair representation of workers in the task 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 (Figure 5.4), we believe that oversampling did not have serious effects in our study.

In conclusion, the statistical power of an exposure-outcome study design using group-based exposure estimation depended more on the total number of workers included in the study (with personalized outcome data) than on the size of the population on which exposures were actually determined. When, however, exposure was observed on very few workers, the odds ratio of the exposure-outcome relationship was downward biased irrespective of the total population size. Our findings thus 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 informed use of resources in future epidemiological studies.