



# Statistical Performance of Observational Work Sampling for Assessment of Categorical Exposure Variables: A Simulation Approach Illustrated Using PATH Data

Svend Erik Mathiassen<sup>1\*</sup>, Jennie A. Jackson<sup>1</sup> and Laura Punnett<sup>1,2</sup>

1.Department of Occupational and Public Health Sciences, Centre for Musculoskeletal Research, University of Gävle, Kungsbäcksvägen, SE- 80176 Gävle, Sweden

2.Department of Work Environment, University of Massachusetts Lowell, Lowell, MA 01854, USA

\*Author to whom correspondence should be addressed. Tel: +46-706788158; fax +46-26648686; e-mail: [smn@hig.se](mailto:smn@hig.se)

Submitted 22 February 2013; revised 11 October 2013; revised version accepted 11 October 2013.

## ABSTRACT

**Objectives:** Observational work sampling is often used in occupational studies to assess categorical biomechanical exposures and occurrence of specific work tasks. The statistical performance of data obtained by work sampling is, however, not well understood, impeding informed measurement strategy design. The purpose of this study was to develop a procedure for assessing the statistical properties of work sampling strategies evaluating categorical exposure variables and to illustrate the usefulness of this procedure to examine bias and precision of exposure estimates from samples of different sizes.

**Methods:** From a parent data set of observations on 10 construction workers performing a single operation, the probabilities were determined for each worker of performing four component tasks and working in four mutually exclusive trunk posture categories (neutral, mild flexion, severe flexion, twisted). Using these probabilities, 5000 simulated data sets were created via probability-based resampling for each of six sampling strategies, ranging from 300 to 4500 observations. For each strategy, mean exposure and exposure variability metrics were calculated at both the operation level and task level and for each metric, bias and precision were assessed across the 5000 simulations.

**Results:** Estimates of exposure variability were substantially more uncertain at all sample sizes than estimates of mean exposures and task proportions. Estimates at small sample sizes were also biased. With only 600 samples, proportions of the different tasks and of working with a neutral trunk posture (the most common) were within 10% of the true target value in at least 80% of all the simulated data sets; rarer exposures required at least 1500 samples. For most task-level mean exposure variables and for all operation-level and task-level estimates of exposure variability, performance was low, even with 4500 samples. In general, the precision of mean exposure estimates did not depend on the exposure variability between workers.

**Conclusions:** The suggested probability-based simulation approach proved to be versatile and generally suitable for assessing bias and precision of data collection strategies using work sampling to

estimate categorical data. The approach can be used in both real and hypothetical scenarios, in ergonomics, as well as in other areas of occupational epidemiology and intervention research. The reported statistical properties associated with sample size are likely widely relevant to studies using work sampling to assess categorical variables.

**KEYWORDS:** epidemiology; ergonomics; exposure assessment methodology; precision; statistical efficiency; working postures

## INTRODUCTION

Field assessments to quantify biomechanical exposures (physical loads) at work frequently employ observational methods to determine body postures, type of materials handling, and other ergonomic characteristics relevant for risk of musculoskeletal disorders (Li and Buckle, 1999; Denis *et al.*, 2000; Takala *et al.*, 2010). Many observation methods, such as the Ovako Working posture Assessment System (OWAS) (Karhu *et al.*, 1977), the Task Recording and Analysis on Computer method (TRAC) (van der Beek *et al.*, 1992), the Back-Exposure Sampling Tool (Back-EST) (Village *et al.*, 2009), and several more (e.g. Hoogendoorn *et al.*, 2000; Neumann *et al.*, 2001; Bao *et al.*, 2009), are based on work sampling, with momentary observations, collected at either fixed or random time intervals. The resulting series of individual observations is then typically summarized in terms of proportions of time in predetermined categories, such as posture intervals or specific tasks. Time-based work sampling has been used for decades in industrial engineering (Richardson and Pape, 1982) and ergonomics (Dempsey and Mathiassen, 2006) as a tool to simultaneously assess the occurrence of tasks and work-related risk factors to musculoskeletal disorders. While observational methods based on continuous, event-based observation of biomechanical exposure are available (e.g. Punnett *et al.*, 1991; Christensen *et al.*, 1995; Fransson-Hall *et al.*, 1995; Fallentin *et al.*, 2001; Dartt *et al.*, 2009; Hooftman *et al.*, 2009; Mathiassen and Paquet, 2010), work sampling was recently shown to be the more cost-efficient approach for observing working postures (Rezagholi *et al.*, 2012).

Other types of occupational exposures may also be described with categorical variables, in particular the presence of workers in various chemical or acoustic environments and/or tasks (e.g. Preller *et al.*, 1995; Susi *et al.*, 2000; Neitzel *et al.*, 2011). In both ergonomics and occupational hygiene, operations are often analyzed by task to identify sources of exposure as targets for intervention (Dempsey and Mathiassen, 2006). At the level of individual workers, the proportion of time in tasks can be used together with information on task-specific

exposures to estimate job exposures. Such task-based exposure modeling has been used for both biomechanical (e.g. Burdorf *et al.*, 1997; Chen *et al.*, 2004; Mathiassen *et al.*, 2005; Svendsen *et al.*, 2005; Bovenzi, 2009) and other occupational exposures (e.g. Benke *et al.*, 2000; Harrison *et al.*, 2002; Semple *et al.*, 2003; Neitzel *et al.*, 2011). Correct information on task proportions is a prerequisite for these models to operate as intended, i.e. produce an unbiased estimate of the modeled exposure (Mathiassen *et al.*, 2003a; Burstyn, 2009).

In general, guidance is scarce on how to design an appropriate data collection strategy for estimating exposures for operations, tasks, or jobs of individual workers using observational methods (Takala *et al.*, 2010). This is a serious concern, considering that awareness and proper appreciation of the statistical properties and performance of the data collecting strategy used with a particular observation method is, arguably, at least as important to the interpretation of the resulting exposure data as is the basic validity and reliability of that method (Takala *et al.*, 2010).

For studies designed to assess mean exposures on a continuous scale, the ability of an exposure sampling strategy to produce a correct exposure estimate, i.e. its statistical performance, can be assessed using information on exposure variability in the target population and the size of the exposure sample (Samuels *et al.*, 1985; Mathiassen *et al.*, 2002, 2003a; Jackson *et al.*, 2009; Liv *et al.*, 2011). These algorithms are based on assumptions about the distribution of the underlying exposures that are likely often not met for exposures measured on categorical scales (Mathiassen and Paquet, 2010) and for exposure variables other than the mean, including exposure variability metrics (Liv *et al.*, 2012). Simulation can be a viable alternative for assessing statistical performance in such cases (Liv *et al.*, 2011). Simulations can be based on expected exposure distribution parameters (Semple *et al.*, 2003) or on resampling of empirical data sets, as in non-parametric bootstrapping (Burdorf and van Riel, 1996; Hoozemans *et al.*, 2001; Paquet *et al.*, 2005; Mathiassen and Paquet, 2010; Liv *et al.*, 2011, 2012).

The aim of the present study was to develop a general procedure for assessing the statistical performance of observational work sampling of categorical exposure variables and to use that procedure in a representative occupational scenario to gain a better understanding of the influence of sample size on bias and precision of estimates of variables expressing central tendencies and variability of task occurrence and of working postures at the level of operation and tasks.

## MATERIALS AND METHODS

### Parent data set

Previously collected data using the PATH—Postures, Activities, Tools, and Handling—observation method to assess tunnel and highway construction (Tak *et al.*, 2011) were utilized as the parent data set for this methodological study. PATH (Buchholz *et al.*, 1996) is a work sampling tool to estimate biomechanically relevant exposure variables. PATH has primarily been used to provide exposure estimates at the operation level, although exposures can also be estimated for separate tasks and even individual workers. The PATH method is reproducible, given adequate training of observers (Park *et al.*, 2009), and valid compared with the results of direct technical measurements (Paquet *et al.*, 2001; Tak *et al.*, 2007). In a recent review, PATH was rated as a ‘thoroughly developed’ method with a ‘systematic and well-designed sampling approach’ (Takala *et al.*, 2010). Thus, PATH serves as a suitable model for observational exposure assessment employing a work sampling approach.

From the nine operations represented in the parent data set, we selected ‘jacking pit construction’ by laborers as a model of an operation performed by several workers over an extended period of time. Four component tasks occurred during the days observed: top work, pit wall construction, manual excavation, and other miscellaneous work (Paquet *et al.*, 2005). The PATH observations of jacking pit construction were collected over 12 days spanning one calendar month. Observation periods ranged from 120 to 460 min day<sup>-1</sup>. The same two analysts observed this operation on each day; on one day, a third observer was present.

After excluding one worker with fewer than 40 observations, the resulting data set comprised a total of 3103 observations distributed among 10 workers (Table 1). For the present paper, the primary biomechanical exposure

of interest was trunk posture, which was recorded as a categorical variable with four divisions: neutral (<20° flexion), mild flexion (between 20° and 45°), severe flexion (>45°), and twist (with or without flexion). Trunk posture was selected because it is an important risk factor for back disorders (Punnett *et al.*, 1991) and because non-neutral postures of the trunk were frequently observed in all of the operations represented in the large construction data set (Tak *et al.*, 2011).

### Simulated sampling strategies

Six data collection strategies were simulated to reflect different durations of sampling, ranging from 5 to 75 h, with observations ‘collected’ at 1-min intervals. The six strategies included 300, 600, 900, 1500, 3000, and 4500 observations, roughly corresponding to 1, 2, 3 full days and 1, 2, 3 full weeks of sampling. For each of the six strategies, simulated observations were generated using a probability-based procedure with the following stepwise algorithm:

- i. A worker was randomly selected from the group of all 10 workers, all workers having equal probabilities of being selected.
- ii. A task was randomly determined for that worker based on the probabilities in the parent data set of that worker performing each of the four possible tasks.
- iii. An exposure for that worker performing that task was randomly determined based on the probabilities in the parent data set of that worker experiencing each of the four possible exposure levels when performing that specific task.

This simulation procedure reflects and reproduces the multinomial structure of categorical PATH observations at the three hierarchical levels of subjects (10 categories), tasks within subject (four categories), and exposures within task and subject (four categories). Any individual observation will result in a positive (‘yes’) answer in exactly one of the possible categories at each of these levels, and the probabilities of obtaining a ‘yes’ in any category within the level naturally add up to 100%. Thus, samples of (independent) multiple observations are categorically distributed, with properties determined by the true outcome probabilities in the set of categories within the same level (cf. Appendix).

**Table 1. PATH parent data by individual worker showing the total number of observations ( $n_{\bullet\bullet s}^s$ ), job-level trunk posture ( $\mu_{E\bullet s}$ , % total time)<sup>a</sup>, proportional task occurrence ( $w_{TS}$ , % total time)<sup>b</sup>, and task-level trunk posture ( $\mu_{ETS}$ , % task time)<sup>c</sup> for each worker**

	Variable	Posture	Worker										
			1	2	3	4	5	6	7	8	9	10	
Job level	$n_{\bullet\bullet s}^s$	—	294	68	272	245	273	315	248	289	393	706	
	$\mu_{E\bullet s}$	Neutral	74.8	73.5	83.1	73.9	68.9	60.0	54.4	75.8	77.9	67.1	
		Mild	12.6	13.2	7.4	13.5	14.3	17.8	17.7	13.5	13.7	20.7	
		Severe	10.2	8.8	3.7	10.2	12.1	19.7	24.2	6.6	6.9	11.1	
Top work task	$w_{TS}$	Twisted	2.4	4.4	5.9	2.5	4.8	2.5	3.6	4.2	1.5	1.1	
		—	19.1	8.8	17.7	26.9	75.5	0.0	0.0	32.5	50.4	0.0	
		$\mu_{ETS}$	Neutral	89.3	50.0	72.9	69.7	69.4	—	—	62.8	70.1	—
	$\mu_{ETS}$	Mild	7.1	16.7	12.5	13.6	15.1	—	—	28.7	15.7	—	
Severe		1.8	16.7	12.5	13.6	12.1	—	—	7.5	11.6	—		
Twisted		1.8	16.7	2.1	3.0	3.4	—	—	1.1	2.0	—		
Pit wall construction task	$w_{TS}$	—	30.6	51.5	12.9	16.7	3.7	33.0	21.4	2.8	0.0	22.8	
		$\mu_{ETS}$	Neutral	54.4	71.4	74.3	75.6	50.0	43.3	47.2	62.5	—	54.0
		Mild	17.8	14.3	8.6	12.2	30.0	26.9	24.5	12.5	—	24.8	
	$\mu_{ETS}$	Severe	22.2	11.4	0.0	12.2	20.0	28.9	28.3	12.5	—	20.5	
Twisted		5.6	2.9	17.1	0.0	0.0	1.0	0.0	12.5	—	0.6		
Manual excavation task		$w_{TS}$	—	40.8	22.1	42.7	27.4	0.0	28.6	65.3	55.7	37.4	69.6
	$\mu_{ETS}$		Neutral	84.2	66.7	84.5	67.2	—	37.8	50.6	80.1	84.4	70.3
	Mild		12.5	20.0	7.8	22.4	—	24.4	17.9	6.8	12.9	19.1	
	$\mu_{ETS}$	Severe	2.5	6.7	3.5	6.0	—	32.2	27.2	6.8	2.0	9.2	
Twisted		0.8	6.7	4.3	4.5	—	5.6	4.3	6.2	0.7	1.4		
Miscellaneous work tasks		$w_{TS}$	—	9.5	17.7	26.8	29.0	20.9	38.4	13.3	9.0	12.2	7.7
	$\mu_{ETS}$		Neutral	71.4	100.0	91.8	83.1	70.2	90.9	84.9	100.0	87.5	77.8
	Mild		7.1	0.0	2.7	5.6	8.8	5.0	6.1	0.0	8.3	22.2	
	$\mu_{ETS}$	Severe	21.4	0.0	0.0	9.9	10.5	2.5	3.0	0.0	2.1	0.0	
Twisted		0.0	0.0	5.5	1.4	10.5	1.7	6.1	0.0	2.1	0.0		

<sup>a</sup>Job exposure:  $\mu_{E\bullet s} = \frac{n_{E\bullet s}^s}{n_{\bullet\bullet s}^s}$  (% total time).

<sup>b</sup>Task proportion:  $w_{TS} = \frac{n_{\bullet TS}^s}{n_{\bullet\bullet s}^s}$  (% total time).

<sup>c</sup>Task exposure:  $\mu_{ETS} = \frac{n_{ETS}^s}{n_{\bullet TS}^s}$  (% task time).

$n_{\bullet\bullet s}^s$ , total number of samples from subject  $s$ , across all tasks and exposure categories;  $n_{\bullet TS}^s$ , number of samples from subject  $s$  performing task  $T$ , irrespective of exposure category;  $n_{E\bullet s}^s$ , number of samples from subject  $s$  within exposure category  $E$ , irrespective of task; and  $n_{ETS}^s$ , number of samples from subject  $s$  within exposure category  $E$ , while performing task  $T$ .

Resampling at the level of individual observations as described above was repeated until a complete set of simulated data had been created, as dictated by the number of observations—from 300 to 4500—required for each of the six specific data collection strategies. For each strategy, 5000 such data sets were generated. Simulations were performed

using a custom software program written in MatLab (MathWorks, Natick, MA, USA; code provided as [Supplementary material](#), available at *Annals of Occupational Hygiene* online). Five thousand repeats have been considered a sufficient basis for analyzing distributions in previous simulation studies (e.g. [Semple et al., 2003](#); [Liv et al., 2012](#)).

We selected this probabilistic resampling procedure as opposed to non-parametric resampling with replacement from the parent data set (conventional bootstrapping; Burdorf and van Riel, 1996; Hoozemans *et al.*, 2001; Mathiassen and Paquet, 2010) because it allows the probabilities assigned to the occurrence of workers, tasks, and exposures to be manipulated. This, in turn, allows scenarios differing from the one represented by the parent data set to be investigated, as illustrated by the assignment of equal selection probabilities for all workers in this study (step ‘i’ in the algorithm above) even though they were, in fact, represented to different extents in the parent data (cf. Table 1,  $n_{\bullet\bullet S}^s$ ).

### Task and exposure variables

For each simulated data set, eight summary exposure variables were then calculated. As defined and explained in Table 2, we examined the following:

- Two operation-level variables: average operation exposure in the group of workers ( $\mu_{E\bullet\bullet}$ ) calculated using a mean-of-means approach (Samuels *et al.*, 1985), and variance between subjects in job exposure ( $s_{BS-E\bullet S}^2$ ).
- Four task-level variables (calculated for each of the tasks in the operation): the relative occurrence of the task in the operation ( $w_{T\bullet}$ ), the variance between subjects in task occurrence ( $s_{BS-w_{TS}}^2$ ), the group mean task exposure ( $\mu_{ET\bullet}$ ), and the variance between subjects in task exposure ( $s_{BS-ETS}^2$ ).
- Two variables summarizing differences between the four tasks in the operation: task diversity ( $MS\Delta_E$ ) and task contrast within the operation ( $C_E$ ).

Since a complete set of variables was obtained for each simulated data set, 5000 sets of all eight exposure variables (some of them quadrupled, since the operation contained four tasks) were available for each of the six investigated sampling strategies.

### Sampling performance

The mean and 5th–95th percentile range of the cumulative probability plots of the 5000 simulated values for each task and posture variable, category, and sampling strategy were calculated as summary

measures of statistical performance, reflecting the precision of the resulting exposure estimate. As an additional measure of performance, the proportion of the 5000 values falling between 90 and 110% of the value in the parent set, i.e. a  $\pm 10\%$  level of ‘coverage probability’ (Landon and Singpurwalla, 2008), was determined for all variables, categories, and sampling strategies. This metric reflects the ability of the sampling strategy to produce a result in close proximity to the ‘true’ target value, capturing the combined effects of bias and imprecision. The distributional properties of the cumulative probability plots were also examined visually. A custom software program was written in MatLab to calculate the eight task and posture variables and to assess statistical performance from the cumulative distributions, using the metrics described above.

For the variables  $\mu_{E\bullet\bullet}$ ,  $w_{T\bullet}$ , and  $\mu_{ET\bullet}$ , the coverage probabilities reflect results in each individual category, independent of other categories in the same set (e.g. in each posture category at the operation level). However, they do not capture correlations—negative or positive—between results within a category set, which will be present due to the ‘compositional’ nature of data (Aitchison, 1986; Reimann *et al.*, 2012), i.e. that results inherently add up to 100%. Thus, in order to measure the ability of our sampling strategies to deliver results close to the truth in several categories (tasks or postures) simultaneously, we calculated ‘compositional coverage probabilities’, i.e. the proportions of the 5000 data sets under each strategy that returned a result between 90 and 110% of the parent data set value in 0, 1, 2, 3, or all four categories.

## RESULTS

### Parent data set

Exposures in the parent data set are summarized in the column ‘Target’ in Tables 3a (operation level), Table 4a (task level, miscellaneous work), and Supplementary Table S1a, available at *Annals of Occupational Hygiene* online (task level, other tasks). The 10 jacking pit construction workers spent, on average, 70.9% of their job time in neutral postures, 14.4% in mild flexion, 11.3% in severe flexion, and 3.3% in twisted trunk postures ( $\mu_{E\bullet\bullet}$ ; Table 3a). The four component tasks, i.e. top work, pit

**Table 2. Summary of the eight exposure variables utilized to evaluate the statistical performance of work sampling strategies. In the present study, ‘exposure’ refers specifically to trunk postures**

Variable	Notation	Description	Formula	Rationale
Job exposure, group mean	$\mu_{E\bullet\bullet}$	Percent of total job time spent in each exposure category; mean across workers.	$\mu_{E\bullet\bullet} = \frac{1}{n_{\bullet\bullet}} \sum_s \mu_{E\bullet s}$	Gives a general descriptive measure of exposure in the job.
Job exposure variability between workers	$s_{BS-E\bullet s}^2$	Overall variance between workers in job exposure (including between-subject and within-subject components). One value for each exposure category.	$s_{BS-E\bullet s}^2 = \frac{1}{(n_{\bullet\bullet} - 1)} \sum_s (\mu_{E\bullet s} - \mu_{E\bullet\bullet})^2$	Indicates whether the population is homogeneous in terms of exposure or whether some workers are ‘outliers’. Important information, for instance, with respect to occupational threshold limits and work roles of individual workers.
Task diversity within the job	$MS\Delta_E$	Mean squared deviation of exposures of the tasks occurring in the job, i.e. mean squared difference across tasks between mean exposures $\mu_{ET\bullet}$ in each individual task (see below) and the mean of all task exposures. One value for each exposure category.	$MS\Delta_E = \frac{1}{n_T} \sum_T (\mu_{ET\bullet} - \frac{1}{n_T} \sum_T \mu_{ET\bullet})^2$	Offers a basic index of task exposure diversity (Mathiassen, 2006), indicating whether tasks within the job represent a potential source of exposure variation in that job.
Task contrast within the job	$C_E$	Contrast in exposure between tasks. Calculated as ratio of task diversity within job ( $MS\Delta_E$ ) to the sum of task diversity and the mean variance across tasks between workers in task exposure (mean $s_{BS-ETS}^2$ ). One value for each exposure category.	$C_E = \frac{MS\Delta_E}{(MS\Delta_E + \frac{1}{n_T} \sum_T s_{BS-ETS}^2)}$	Measures task exposure diversity by an intraclass correlation coefficient acknowledging the size and uncertainty of each task exposure (Mathiassen, 2006). A larger value indicates that tasks differ (contrast) in exposure with greater certainty. $C_E$ is a relative metric (value 0–1) and will get larger both if tasks differ more in mean exposure (as measured by the $MS\Delta_E$ ) and if workers have more similar exposures when performing each task (as measured by the average $s_{BS-ETS}^2$ ).

Table 2. Continued

Variable	Notation	Description	Formula	Rationale
Task occurrence, group mean	$w_{T\bullet}$	Percent of total job time spent performing a particular task; mean across workers.	$w_{T\bullet} = \frac{1}{n_{s\bullet}} \sum_s w_{TS}$	Indicates whether a particular task that might, according to its exposure, be a target for intervention occurs in a sufficient proportion for an intervention to have a noticeable effect on overall exposure in the job.
Task occurrence variability between workers	$s_{BS-wTS}^2$	Variance between workers in task occurrence.	$s_{BS-wTS}^2 = \frac{1}{(n_{s\bullet} - 1)} \sum_s (w_{TS} - w_{T\bullet})^2$	Indicates whether the job includes 'subjobs,' with some workers being specialized on doing those (giving a large variance), or whether all workers are 'generalists' with almost equal task distributions. This is important information for proper targeting of interventions to the work organization (e.g. increased job rotation) or to workers or specific tasks.
Task exposure, group mean	$\mu_{ET\bullet}$	Percent time in each exposure category for a specific task; mean across only workers performing that task.	$\mu_{ET\bullet} = \frac{1}{n_{s(T)}} \sum_{s(T)} \mu_{ETS}$	Indicates, for each task, whether that particular task may be a target for intervention due to a critical exposure.
Variance between workers in task exposure	$s_{BS-ETS}^2$	Variance between workers in task exposure. One value for each exposure category.	$s_{BS-ETS}^2 = \frac{1}{(n_{s(T)} - 1)} \sum_{s(T)} (\mu_{ETS} - \mu_{ET\bullet})^2$	Indicates whether each particular task introduces a similar exposure to all workers (indicating that it leaves little autonomy, and that a possible intervention would be most effective if targeted to the task rather than the worker, <a href="#">Mathiassen et al., 2003b</a> ), or whether some workers are particularly exposed, due to unfavorable working techniques and/or because they perform certain subtasks with deviating exposures more than other workers.

$\mu_{E\bullet s}$ , proportion of time spent by an individual worker in exposure category  $E$  for the whole job (cf. [Table 1](#));  $w_{TS}$ , proportion of time spent by an individual worker in exposure category  $E$  while performing task  $T$  (cf. [Table 1](#));  $n_{s\bullet}$ , total number of observed workers;  $n_{s(T)}$ , number of workers observed to perform task  $T$ ;  $\sum_s$ , summation across all observed workers;  $\sum_{s(T)}$ , summation across all tasks in the job;  $\sum_{s(T)}$ , summation across all workers observed to perform task  $T$ .

wall construction, manual excavation, and miscellaneous work, comprised, on average, 23.1, 19.5, 38.9, and 18.4% of the operation, respectively ( $w_{T\bullet}$ ; Table 4a and Supplementary Table S1a, available at *Annals of Occupational Hygiene* online). The mean time across

workers spent within a certain exposure posture category in each task ( $\mu_{ET\bullet}$ ; Table 4a and Supplementary Table S1a, available at *Annals of Occupational Hygiene* online) ranged from 59.2 to 85.8% (neutral), 6.6 to 19.1% (mild flexion), 4.9 to 17.3% (severe flexion),

**Table 3.** (a) Operation-level and job-level exposure variables for all simulation strategies and the parent data set (column ‘Target’). All cells but ‘Target’ show the mean value with 5th–95th percentile values from the simulated distributions located in square brackets below. For explanation of exposure variables see Table 2. (b) Coverage probabilities for operation and job-level exposure variables: percentage of simulations ( $n = 5000$  for each strategy) in which the estimated exposure value was between 90 and 110% of the true target exposure value in the parent data set. Cells with 90% coverage or more highlighted in dark gray; cells with 80–89.9% coverage highlighted in light gray

Variable	Posture	Simulated sampling strategy, number of samples						Target
		300	600	900	1500	3000	4500	
<b>(a)</b>								
$\mu_{E\bullet\bullet}$	Neutral	70.9 [66.6–75.3]	71.0 [67.9–74.1]	71.0 [68.4–73.4]	70.9 [69.0–72.8]	70.9 [69.5–72.3]	70.9 [69.8–72.0]	70.9
	Mild	14.5 [11.2–17.8]	14.4 [12.1–16.8]	14.4 [12.5–16.4]	14.4 [13.0–15.9]	14.4 [13.4–15.5]	14.4 [13.6–15.3]	14.4
	Severe	11.3 [8.4–14.4]	11.3 [9.3–13.5]	11.3 [9.6–13.0]	11.4 [10.0–12.7]	11.4 [10.4–12.3]	11.3 [10.6–12.1]	11.3
	Twisted	3.3 [1.7–5.1]	3.3 [2.1–4.5]	3.3 [2.3–4.3]	3.3 [2.5–4.1]	3.3 [2.8–3.8]	3.3 [2.9–3.8]	3.3
$s_{BS-E\bullet S}^2$	Neutral	140.6 [59.9–243.8]	106.4 [52.7–173.7]	94.6 [51.6–146.5]	86.1 [52.3–124.6]	79.7 [55.5–107.7]	77.7 [58.0–99.1]	73.4
	Mild	55.0 [20.3–102.8]	33.8 [14.1–60.8]	26.9 [11.7–47.1]	21.3 [10.0–35.3]	17.2 [9.6–26.3]	15.8 [9.6–22.9]	13.2
	Severe	70.5 [27.0–131.3]	54.3 [24.9–93.9]	48.9 [24.5–79.7]	44.6 [25.7–67.3]	41.5 [27.7–57.0]	40.5 [29.1–53.1]	38.4
	Twisted	13.0 [3.8–27.5]	7.7 [2.7–15.5]	5.8 [2.2–11.3]	4.5 [1.8–8.2]	3.4 [1.6–5.7]	3.0 [1.6–4.9]	2.3
$MS\Delta_E$	Neutral	124.7 [26.9–255.2]	109.5 [34.4–208.6]	103.7 [40–188.2]	98.1 [50.0–160.2]	94.9 [60.1–134.9]	93.4 [65.7–125.1]	90.9
	Mild	42.9 [6.4–104.6]	33.5 [6.3–78.7]	30.1 [7.4–66.0]	26.7 [9.4–51.8]	24.0 [12.6–39.0]	23.2 [13.7–34.6]	21.7
	Severe	37.4 [5.1–97.8]	29.3 [4.7–76.5]	26.1 [5.8–64.9]	23.1 [7.0–49.9]	21.4 [9.6–37.6]	20.7 [11.3–32.5]	19.3
	Twisted	8.4 [0.7–30.0]	4.9 [0.4–17.7]	3.5 [0.3–11.7]	2.3 [0.2–6.6]	1.4 [0.2–3.7]	1.0 [0.1–2.7]	0.5
$C_E$	Neutral	0.21 [0.05–0.40]	0.24 [0.09–0.42]	0.27 [0.12–0.43]	0.30 [0.17–0.44]	0.34 [0.23–0.45]	0.35 [0.26–0.44]	0.35
	Mild	0.15 [0.02–0.32]	0.18 [0.04–0.34]	0.20 [0.06–0.37]	0.24 [0.09–0.40]	0.28 [0.15–0.42]	0.30 [0.18–0.43]	0.32
	Severe	0.14 [0.02–0.30]	0.15 [0.03–0.30]	0.16 [0.05–0.31]	0.18 [0.07–0.32]	0.21 [0.10–0.32]	0.22 [0.12–0.32]	0.22
	Twisted	0.09 [0.02–0.19]	0.07 [0.01–0.17]	0.07 [0.01–0.15]	0.06 [0.01–0.12]	0.04 [0.01–0.09]	0.04 [0.01–0.08]	0.02



Table 3. Continued

Variable	Posture	Simulated sampling strategy, number of samples						Target
		300	600	900	1500	3000	4500	
(b)								
$\mu_{E\bullet\bullet}$	Neutral	99.4	100.0	100.0	100.0	100.0	100.0	
	Mild	51.2	68.3	77.8	89.0	97.4	99.2	
	Severe	46.8	63.1	72.3	83.1	94.9	98.1	
	Twisted	24.6	33.6	41.5	52.2	68.5	78.6	
$s_{BS-E\bullet S}^2$	Neutral	6.9	12.5	18.4	25.0	35.3	42.4	
	Mild	0.6	3.0	5.4	9.2	18.0	22.7	
	Severe	8.3	15.4	18.0	23.5	33.2	40.3	
	Twisted	0.5	1.9	4.1	7.3	14.3	18.5	
$MS\Delta_E$	Neutral	9.8	14.5	17.2	21.7	31.4	38.7	
	Mild	7.4	10.4	11.3	15.4	22.4	27.3	
	Severe	8.0	10.4	12.3	13.9	20.6	24.4	
	Twisted	0.8	1.6	2.3	3.2	5.5	6.3	
$C_E$	Neutral	9.4	14.4	18.9	26.2	40.4	47.2	
	Mild	5.8	8.4	11.0	16.3	25.1	28.6	
	Severe	10.5	13.5	14.6	18.2	24.6	27.2	
	Twisted	1.7	2.9	3.7	4.5	6.1	6.9	

and 2.7 to 4.4% (twisted). Tasks differed most with respect to neutral postures ( $MS\Delta_E$ ; Table 3a). It should be noted that the values of  $MS\Delta_E$  are somewhat inflated due to exposure variability within each of the four tasks. Task contrasts within the operation corroborated that the tasks differed most consistently in the occurrence of neutral and mild trunk postures ( $C_E$ ; Table 3a).

Individual workers differed considerably in how often they were observed in each of the four posture categories, particularly for neutral and severe trunk flexion ( $s_{BS-E\bullet S}^2$ ; Table 3a), as well as in the relative proportion of time they spent performing each of the four tasks ( $s_{BS-wTS}^2$ ; Table 3a). Even within a task, postures differed considerably between subjects ( $s_{BS-ETS}^2$ ; Table 4a and Supplementary Table S1a, available at *Annals of Occupational Hygiene* online), with the caveat that these variabilities include contributions from within-subject variability (between and within measurement days), which could not be isolated and adjusted for.

#### Statistical performance: operation level

For all sampling strategies, operation exposure,  $\mu_{E\bullet\bullet}$ , was estimated without bias relative to the 'true'

target (Table 3a). This is illustrated by the alignment of the inflection points of the simulated data dispersion curves and the line indicating the target value in Fig. 1a. As expected, the 5th–95th prediction interval decreased with increasing sample size ( $\mu_{E\bullet\bullet}$ ; Table 3a, values in square brackets). This narrowing is illustrated in Fig. 1a by the decreased dispersion of the 5000 simulated data sets at larger sample sizes.

Coverage probability depended on the true occurrence of the postural exposure in the parent data set (Table 3b). For example, for neutral trunk posture, which was the most frequently observed, a high level of coverage probability (99.3%) was shown even for a sampling strategy with only 300 observations. In contrast, for the most rarely observed posture, twisting, even a sampling strategy containing 4500 observations led only to a 78.6% probability of producing an operation exposure estimate within  $\pm 10\%$  of the target value. Compositional coverage probabilities for postures at the operation level increased with sample size (Fig. 2a). With 300 observations, about one of every five data sets contained values deviating more than 10% from the target value in three or all four posture categories. Even with 4500 observations, fewer than 80% of all data sets showed values within  $\pm 10\%$  of the target in all four categories. These compositional

Table 4. (a) Task-level exposure variables for all simulation strategies and the parent data set (column ‘Target’); miscellaneous work. All cells but ‘Target’ show the mean value with 5th–95th percentile values of the simulated distributions located in square brackets below. For explanation of exposure variables, see Table 2. (b) Coverage probabilities of exposure variables for the task miscellaneous work: percentage of simulations ( $n = 5000$  for each strategy) in which the estimated exposure value was between 90 and 110% of the true target exposure value in the parent data set. Cells with 90% coverage or more highlighted in dark gray; cells with 80–89.9% coverage highlighted in light gray. Corresponding data for the other three tasks are shown in Supplementary Tables S1a and S1b (available at *Annals of Occupational Hygiene* online)

Variable	Posture	Simulated sampling strategy, number of samples					Target
		300	600	900	1500	3000	
<b>(a)</b>							
$w_T$	—	18.5 [14.8–22.4]	18.4 [15.8–21.2]	18.4 [16.3–20.5]	18.4 [16.8–20.0]	18.4 [17.3–19.6]	18.4 [17.5–19.4]
$s_{BS-WTS}^2$	—	154.7 [68.6–264.6]	127.8 [68.6–196.6]	119.4 [71.3–173.9]	113.5 [76.9–156.1]	108.7 [82.4–138.3]	107.2 [85.9–131.5]
$\mu_{ET}$	Neutral	85.7 [75.0–94.3]	85.7 [78.6–92.2]	85.7 [80.3–90.8]	85.7 [81.7–89.6]	85.7 [82.8–88.5]	85.7 [83.4–88.0]
	Mild	6.5 [0.6–14.8]	6.6 [1.9–12.4]	6.6 [2.8–10.9]	6.6 [3.5–9.8]	6.6 [4.5–8.8]	6.6 [4.9–8.3]
	Severe	5.0 [0.0–12.0]	5.0 [1.4–9.3]	5.0 [2.0–8.4]	5.0 [2.7–7.5]	5.0 [3.4–6.7]	4.9 [3.7–6.3]
	Twisted	2.8 [0.0–7.2]	2.8 [0.5–5.6]	2.7 [0.9–5.0]	2.7 [1.2–4.4]	2.7 [1.7–3.9]	2.7 [1.9–3.7]
$s_{BS-ETS}^2$	Neutral	629.9 [109.8–1534.3]	289.6 [89.7–732.2]	216.3 [81.6–439.7]	169.6 [75.5–307.7]	139.0 [78.0–219.2]	129.0 [80.1–192.2]
	Mild	227.9 [3.9–988.7]	139.3 [14.1–426.6]	100.8 [16.9–285.0]	73.8 [15.8–186.6]	56.0 [15.8–126.0]	50.5 [16.6–104.6]
	Severe	168.1 [0.0–964.2]	106.4 [11.3–303.8]	87.2 [15.1–244.0]	71.8 [16.4–172.0]	60.2 [21.3–120.3]	55.8 [23.6–102.6]
	Twisted	69.0 [0.0–250.0]	38.8 [2.5–111.9]	29.2 [4.1–76.3]	22.0 [4.8–53.1]	17.2 [5.6–34.8]	15.7 [6.5–29.3]

Table 4. Continued

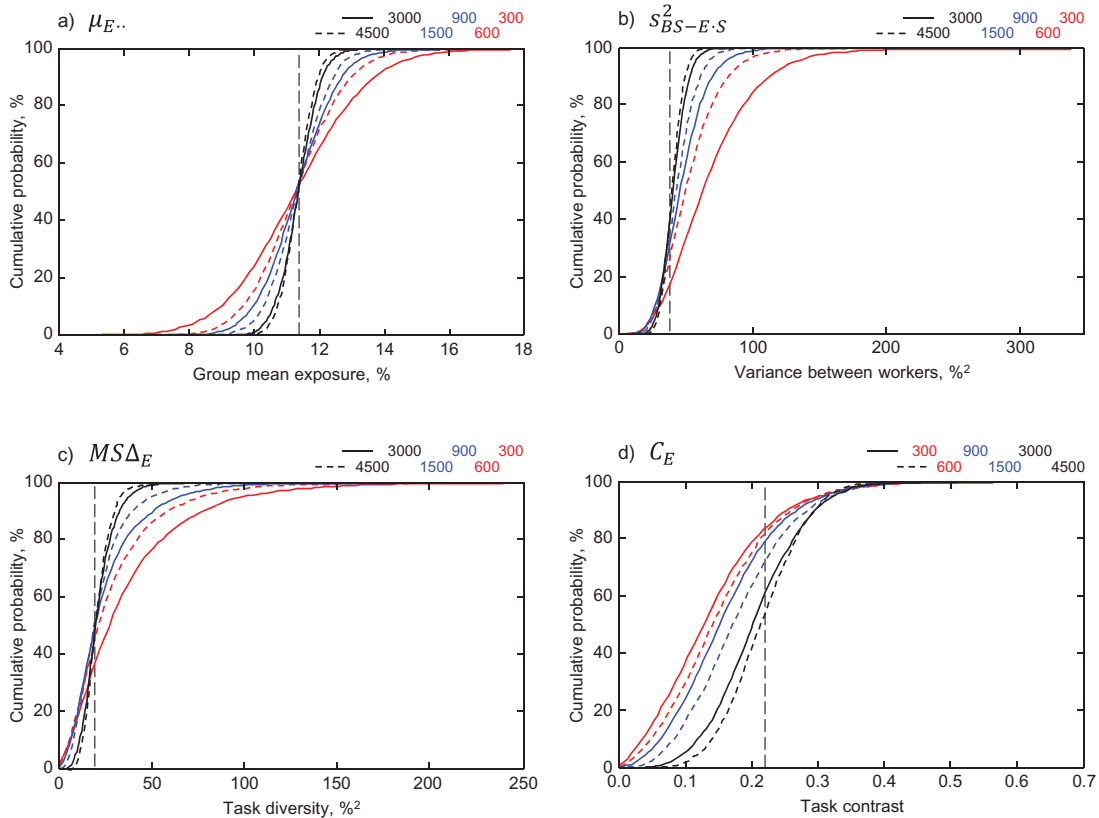
Variable	Posture	Simulated sampling strategy, number of samples					Target
		300	600	900	1500	3000	
(b)							
$w_T$	—	59.1	74.6	84.8	94.2	99.2	100.0
$s_{BS-WTS}^2$	—	12.9	19.6	24.9	33.9	47.3	55.2
$\mu_{ET}$	Neutral	86.2	96.8	99.2	99.9	100.0	100.0
	Mild	10.6	16.4	19.8	27.8	39.3	45.9
	Severe	10.7	16.8	20.7	26.2	37.4	46.2
	Twisted	10.5	12.9	16.7	21.5	31.3	35.7
$s_{BS-ETS}^2$	Neutral	2.1	5.1	8.4	12.9	20.2	27.6
	Mild	4.0	5.2	7.9	9.5	11.3	13.3
	Severe	3.9	7.2	9.3	10.1	14.6	19.3
	Twisted	3.1	5.6	6.0	9.5	12.9	15.6

coverage probabilities deviated by  $\pm 5$  percentage points or less from the probabilities predicted under the assumption that the coverage in each category was independent of that in the other categories in the same set. As an example, the predicted compositional coverage probability under the 300 observation strategy of getting a value close to the truth in all four posture categories at the operation level was (cf. Table 3b)  $0.994 \cdot 0.512 \cdot 0.468 \cdot 0.246 = 100\%$ , i.e. 5.9%, whereas the actual compositional coverage was 5.7% (Fig. 2a).

In contrast to the unbiased mean exposure, estimates of between-subject variability were upwardly biased for all sampling strategies; shorter duration samples were more severely biased and also showed wider prediction intervals ( $s_{BS-E \cdot S}^2$ ; Table 3a). This is illustrated by the horizontal curve shifts for smaller sampling strategies in Fig. 1b. The effect of sample size on bias may result primarily from the fact that the variance of the individual exposure mean values included in the estimate of  $s_{BS-E \cdot S}^2$  is larger for smaller sample sizes and thus inflates the value of  $s_{BS-E \cdot S}^2$  to a larger extent. Coverage probability was much lower for estimates of between-subject variability than that for mean exposures. In the best case, i.e. for neutral trunk postures, coverage probability for  $s_{BS-E \cdot S}^2$  never exceeded 42.4%, even with 4500 samples (Table 3b). In the worst case, that of twisted postures assessed by the 300 observations strategy, only 0.5% of the 5000 simulated data sets fell within  $\pm 10\%$  of the target  $s_{BS-E \cdot S}^2$  value, due to the combined effect of bias and imprecision.

A similar pattern of larger bias and wider prediction intervals at smaller sample sizes was also observed for  $MS\Delta_E$  (Table 3a, Fig. 1c). Again, increased bias was likely due to the effect of within-subject variability on task exposure estimates. Notably, this means that with small sample sizes, task exposures appear to be more different from each other than they really are. Even at the largest sample size (here 4500 observations),  $MS\Delta_E$  did not converge completely to the target value.

By definition (cf. Table 2), the effect of sample size on contrast,  $C_E$ , will be a trade-off between the effects on  $MS\Delta_E$  and on within-task variability,  $s_{BS-ETS}^2$ . In the present material, both decreased with increasing sample size (for  $s_{BS-ETS}^2$ , see below). In most cases, the net effect was that  $C_E$  was underestimated at small sample sizes but increased toward the target value with larger sample



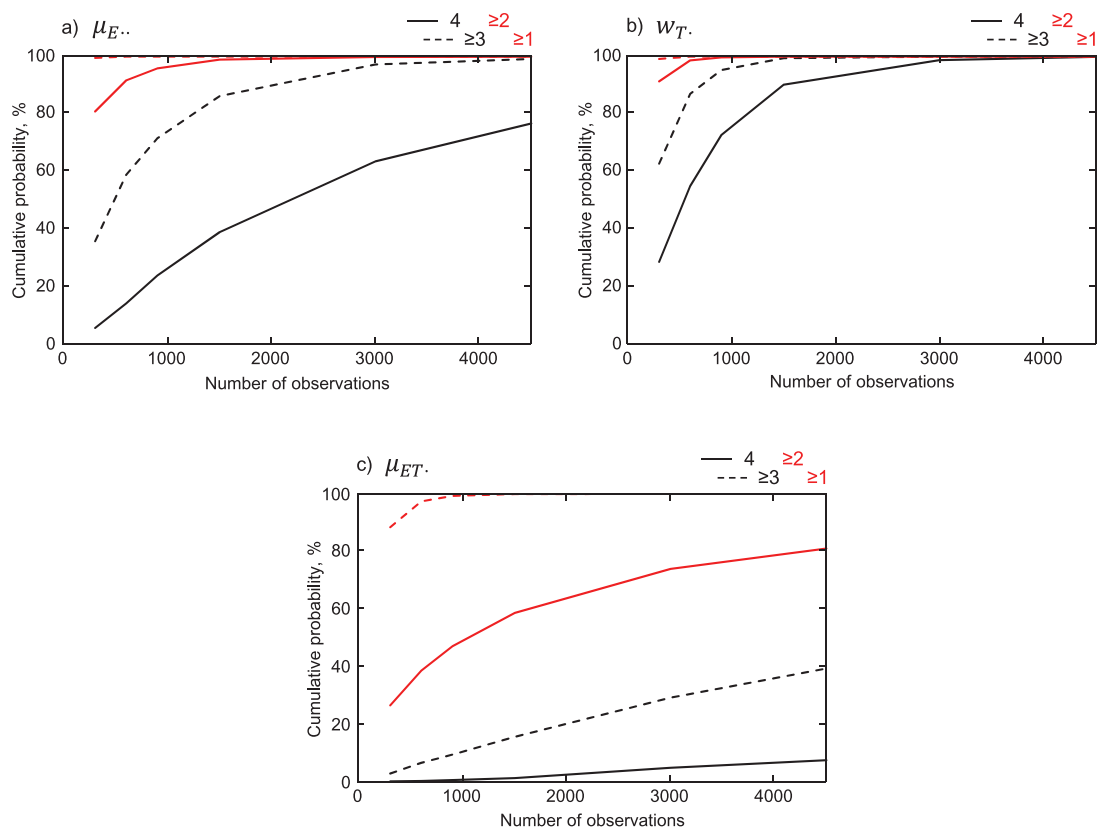
1 Simulated cumulative distributions of variables describing exposure to severe trunk flexion at the operation level: (a) group mean exposure,  $\mu_{E\bullet\bullet}$ ; (b) variance between workers,  $S^2_{BS-E\cdot S}$ ; (c) task diversity (mean squared deviation between task exposures),  $MS\Delta_E$ ; (d) exposure contrast between tasks,  $C_E$ . Each panel shows the distribution of the 5000 simulated results obtained by each of the six investigated observation strategies, from left to right as indicated by the legend in the upper right corner (colored online). Dashed vertical lines indicate the target value read from the parent data set (Table 3a, column ‘Target’).

sizes (Table 3a, Fig. 1d). For twisted postures, however, the opposite was seen:  $C_E$  was overestimated at small sample sizes and decreased toward the target with larger sample sizes. In general, the effect of an increasing sample size on the prediction interval was not as pronounced for  $C_E$  as for the other exposure variables, so coverage probability did not improve as markedly with larger sample sizes. For example, coverage probability for  $C_E$  in severe flexion only increased from 10.5 to 27.2% as the number of samples increased from 300 to 4500 (Table 3b).

#### Statistical performance: task level

The mean proportion of time spent performing each individual task ( $w_{T\bullet}$ ; Table 4a and Supplementary Table S1a, available at *Annals of Occupational Hygiene*

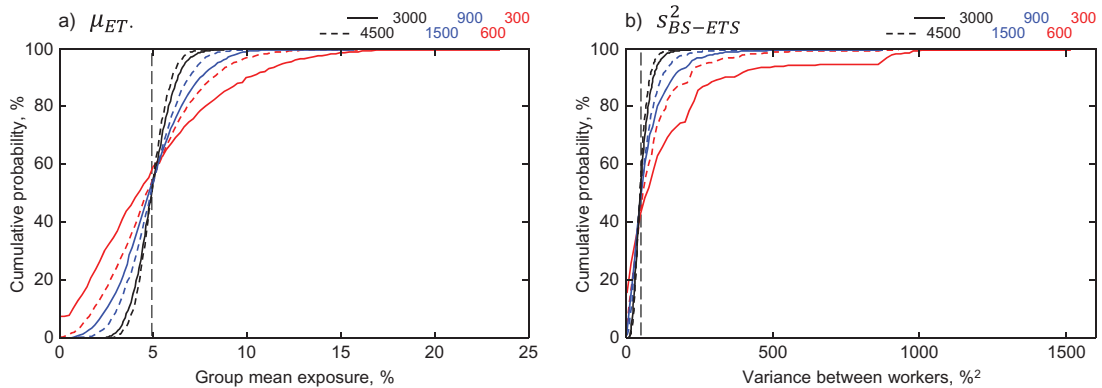
online) was estimated without bias, even when using the smallest sampling strategy. The prediction interval decreased with increased sample size, as expected. Compositional coverage probability was better for the set of task proportions (Fig. 2b) than for postures (Fig. 2a), as might be expected from the larger coverage probabilities for each individual task proportion (Table 4b, Supplementary Table S1b, available at *Annals of Occupational Hygiene* online). Already with only 300 observations, <1 of every 10 data sets contained task proportions deviating more than 10% from the true target value for three or all four tasks (Fig. 2b), and essentially all data sets containing 4500 observations resulted in task proportions within 10% from the target value for all four tasks.



2 Effect of sample size on compositional coverage probabilities for (a) operation-level mean exposure; (b) task proportions; and (c) task-level mean exposure in miscellaneous work. Curves show proportions of the 5000 simulated data sets at each sample size that included values within 10% from the true target in at least 1, at least 2, at least 3, or all 4 categories as indicated by the legend in the upper right corner of each panel (colored online). Posture categories: neutral, mild, severe, twisted; tasks: top work, pit wall construction, manual excavation, miscellaneous work. Diagrams corresponding to panel (c) for the other three tasks are shown in [Supplementary Figure S1](#) (available at *Annals of Occupational Hygiene* online).

In general, even the mean exposure value for individual tasks ( $\mu_{ET}$ ; [Table 4a](#), [Supplementary Table S1a](#), available at *Annals of Occupational Hygiene* online) was unbiased for all six sampling strategies and also demonstrated a decreasing prediction interval with increasing sample size. Coverage probability for task mean exposure estimates,  $\mu_{ET}$  ([Table 4b](#), [Supplementary Table S1b](#), available at *Annals of Occupational Hygiene* online), was, in general, considerably lower than for the corresponding operation level mean exposures,  $\mu_{E..}$  ([Table 3b](#)), mainly because task exposure estimates were based on fewer samples and thus were less precise. This also led to compositional coverage being considerably lower for task mean exposures ([Fig. 2c](#),

[Supplementary Figure S1](#), available at *Annals of Occupational Hygiene* online) than for operation mean exposure ([Fig. 2a](#)). For instance, with 4500 samples, exposures were within  $\pm 10\%$  of the target value in all four posture categories for 76.8% of the assessments of operational exposure ([Fig. 2a](#)), whereas only 7.5% of the compositional exposure estimates in miscellaneous work were within  $\pm 10\%$  ([Fig. 2c](#)). For between-subject variability in task occurrence ( $s_{BS-wTS}^2$ ) and task exposure ( $s_{BS-ETS}^2$ ), pronounced bias and wide prediction intervals were present at smaller sample sizes, in particular for  $s_{BS-ETS}^2$  in the three 'rarer' tasks. Thus, the coverage probability for  $s_{BS-wTS}^2$  and  $s_{BS-ETS}^2$  was low in these three cases ([Table 4b](#), [Supplementary Table](#)



3 Simulated cumulative distributions of variables describing exposure to severe trunk flexion in the task miscellaneous work: (a) group mean exposure,  $\mu_{ET}$ ; (b) variance between workers,  $S_{BS-ETS}^2$ . Each panel shows the distribution of the 5000 simulated results obtained by each of the six investigated observation strategies, from left to right as indicated by the legend in the upper right corner (colored online). Dashed vertical lines indicate the target value read from the parent data set (Table 4a, column ‘Target’).

S1b, available at *Annals of Occupational Hygiene* online).

### Distributional properties of exposure estimates

For some exposure estimates in rare tasks at shorter sampling strategies, the shape of the cumulative distribution across the 5000 simulated values differed from that found for the same exposure variable with larger strategies. For example, in the miscellaneous work task for the 300-sample strategy, the median estimated exposure to severe flexion was smaller than the median obtained with larger sample sizes (Fig. 3a), and severe flexion did not even occur in approximately 8% of the 5000 data sets, as shown by the positive  $y$ -intercept value in Fig. 3a. Under the 300-sample strategy, an average of 55 samples (18.4%;  $w_{T\bullet}$ , Table 4a) are expected to come from miscellaneous work. Since only 4.9% ( $\mu_{ET\bullet}$ , Table 4a) of these 55 samples are expected to show severe trunk flexion, a data set with 300 samples will, occasionally, contain no observations of severe flexion from any of the workers performing the task. In this case, task exposure to severe flexion at the group level will be zero and so contribute to the observed positive  $y$ -intercept.

With the 300-sample strategy, the cumulative distribution of severe trunk flexion in miscellaneous work also appeared slightly jagged (Fig. 3a). This discrete graphical pattern is even more apparent in Fig. 3b, which shows the task-specific, between-worker exposure variability,  $s_{BS-ETS}^2$ ,

for severe trunk flexion during miscellaneous work. Under the 300-sample strategy, five to six simulated observations on average from each individual worker (i.e., 55/10) will be in miscellaneous work. Thus, each observation (sample) will account for ~20% of the time in the task and influence the task exposure of that worker accordingly. The jagged distribution in Fig. 3b indicates that the presence or absence of single samples for each particular worker changes the size of the exposure variability in a stepwise fashion. This pattern is also visible at sample sizes larger than 300, albeit to a lesser extent.

## DISCUSSION

### Statistical performance of work sampling for categorical variables

The primary aim of this study was to develop a procedure for investigating the statistical properties of selected task and exposure variables, estimated from data obtained by different observational work sampling strategies. The proposed probabilistic simulation approach proved useful for disclosing sampling performance when assessing variables that are difficult to access using analytical methods, and resembles in this respect previous simulation studies of precision and bias associated with exposure estimation (e.g. Semple et al., 2003; Liv et al., 2011). Although the empirical illustration based on a large parent data set of PATH observations served primarily to demonstrate the ability of the probabilistic simulation to produce results

of generic relevance to the design of data collection strategies, we also believe the exposure structure of the parent data set to be representative of many categorical data sets of occupational postures obtained using work sampling because all workers performed all tasks and experienced all exposures at least occasionally. Thus, essential results, such as the statistical properties of variability metrics compared with those of mean values, or of variables at the level of tasks compared with at the level of operation, may be generally applicable to most occupational exposure assessment efforts. Therefore, we discuss these results more in detail below, even if numerical values may be specific to the present parent data.

The precision of mean exposure estimates at both the operation and task levels ( $\mu_{E\bullet}$  and  $\mu_{ET\bullet}$ , Tables 3a and 4a) improved with increasing sample sizes from 300 to 4500 observations. This result is consistent with several other studies addressing the relationship between sample size and uncertainty in estimating mean exposures (Burdorf and van Riel, 1996; Hoozemans *et al.*, 2001; Mathiassen *et al.*, 2002; Fethke *et al.*, 2007), as well as a previous analysis of PATH performance in different operations (Paquet *et al.*, 2005). Thus, our study corroborates that highly prevalent exposures may be determined reasonably correctly even with relatively small sample sizes, *in casu*, 300 observations (Table 3a), but that less common exposures, in particular if measured at the task level, may be estimated with considerable uncertainty even with very large total sample sizes, *in casu* 4500 observations (Tables 3a and 4a, and Supplementary Table S1a, available at *Annals of Occupational Hygiene* online). PATH is unlikely to differ much in this respect from other observational methods for assessing categorical variables by work sampling, such as Back-EST (Village *et al.*, 2009), TRAC (van der Beek *et al.*, 1992; Frings-Dresen and Kuijter, 1995; van der Beek *et al.*, 1995), or OWAS (Karhu *et al.*, 1977; Kivi and Mattila, 1991).

The finding that increased sample size improved statistical performance was, of course, expected. The variance of an estimated mean value of a continuous variable will decrease in inverse proportion to the number of samples, provided that the data are randomly distributed (Samuels *et al.*, 1985). In the present case, each basic unit of measurement, i.e. each single PATH observation, is an array of multinomial sets of 'yes' or 'no' answers to whether any particular

worker is observed at that very instant, whether he is performing each specific task, and whether the exposure is within each specific posture category. Thus, data in any particular category are binomially distributed, with the probability of the answer 'yes' being defined by the overall proportion of samples in that category in the parent data set. With large sample sizes, mean values of binomial variables approach normal distributions and so behave statistically as a continuous variable, with properties that can be approximated using analytical methods (cf. Appendix). In the present case, binomial theory led to an expectation that mean values of task proportions and postures would be unbiased and normally distributed with a variance directly computable from the workers' individual exposures and the total number of samples (Appendix).

This theoretical prediction, however, is conditional on collected data sets being large and balanced. In the present case, theoretical expectations, such as the size of prediction intervals (cf. Tables 3a and 4a, and Supplementary Table S1a, available at *Annals of Occupational Hygiene* online), were, indeed, met for most sampling strategies at the operation level and for both task proportions and posture variables. However, at the task level, several deviations from expected performance were observed. For instance, using the 300-sample strategy, the 5th–95th prediction interval for severe flexion in miscellaneous work ( $\mu_{ET\bullet}$ ; Table 4a) was larger (12.0%) than predicted by theory (9.1%) and also somewhat skewed (0.0–12.0%) around the mean value of 5.0%. In this case, only about three observations of severe flexion in miscellaneous work will be available in a complete 300-sample data set (4.9 of 18.4% of 300; cf. Table 4a), and they are not likely to always be equally distributed among different workers. Thus, the data available for estimating exposure to severe flexion in this task are neither 'many' nor balanced. Other examples of irregular distributions were shown in Fig. 2. The occasional discrepancy between theoretical and empirical (simulated) performance illustrates an important use of the probabilistic simulation approach, i.e. to show when assumptions are no longer met in analytical models. In extension, simulations are also very attractive when addressing distributions of variables that cannot be readily addressed by theoretical models, such as exposure variability metrics (Liv *et al.*, 2012).

Both theoretical equations (Appendix) and simulation results draw attention to the fact that the uncertainty of mean exposure estimates did not relate directly to the size of between-subject variability in the exposure. As an illustration, the variability between workers in task proportions ( $s_{BS-wTS}^2$ , Table 4a) was much larger for most tasks than the variability in job exposures ( $s_{BS-E\bullet S}^2$ , Table 3a) and the 5th–95th prediction intervals for task proportions and operation exposures were of the same order of magnitude. This counter-intuitive property of the present work sampling is a result of the sampling being performed in a finite population, which—in large data sets—eliminates the contribution of between-worker variability to the variance of group mean values (cf. Appendix), and of the fact that the average within-subject variability will be less if workers differ substantially in mean exposure than if they are more homogeneous.

Consistent with previous simulations examining continuous exposure variables (Liv *et al.*, 2012), uncertainty (i.e. the 5th–95th prediction intervals) was considerably larger for variables describing variability than for mean values at the same sample size. Prediction intervals for variance estimates were also upwardly skewed with respect to the mean, particularly at small sample sizes (Tables 3a and 4a). As a combined effect of bias and low precision, coverage probabilities for variability metrics were, in general, poor; in some cases with small sample sizes, they were <5%, even at the operation level (Table 3b).

The effect of increased sample size on metrics expressing aspects of exposure variability ( $s_{BS-E\bullet S}^2$ ,  $MS\Delta_E$ ,  $s_{BS-wTS}^2$ ,  $s_{BS-ETS}^2$ ,  $C_E$ ) is more difficult to predict than the effect on mean values ( $\mu_{E\bullet\bullet}$ ,  $\mu_{ET\bullet}$ ,  $w_{T\bullet}$ ), and considerably less literature has been devoted to the statistical performance of such variability metrics (Mathiassen *et al.*, 2002, 2003b; Liv *et al.*, 2012). In the present material, most of these variables ( $s_{BS-E\bullet S}^2$ ,  $MS\Delta_E$ ,  $s_{BS-wTS}^2$ ,  $s_{BS-ETS}^2$ ) were upwardly biased at all sample sizes and particularly at smaller sample sizes (Tables 3a and 4a, and Supplementary Table S1a, available at *Annals of Occupational Hygiene* online). We believe this to result mainly from the larger uncertainty at smaller sample sizes of both within-subject and within-task variability. Within-subject variability was present to different extents for all 10 observed workers, as suggested by their individual

exposure profiles (Table 1), but again within-days and between-days contributions could not be separated.

As noted above, exhaustive sets of mutually exclusive categories lead to compositional results, i.e. constrained data that inevitably add up to a certain number, *in casu*, 100%. This constrained nature of categorical data was reflected and reproduced by the resampling procedure employed for simulating new data sets. Comprehensive use of compositional data, for instance in hypothesis testing or regression analysis, requires specific procedures differing from conventional Euclidian algebra (e.g. Fry *et al.*, 2000; Filzmoser *et al.*, 2009; Filzmoser *et al.*, 2010; Reimann *et al.*, 2012), which fall outside the scope of this study. However, the reported compositional coverage probabilities, measuring the ability of a sample estimate to show values ‘close’ to the truth in an entire category set, revealed that probabilities within a category set, such as the four posture categories in operation exposure, were, indeed, correlated. However, discrepancies were small between empirical compositional coverage probabilities (Fig. 2, Supplementary Figure S1, available at *Annals of Occupational Hygiene* online) and values predicted with the assumption of independence among categories. This suggests that compositional coverage can be fairly well estimated on the basis of binomial theory (cf. Appendix).

#### How many observations are sufficient?

Although we have demonstrated that larger observation samples lead to better statistical performance, if in different ways and at different rates for different posture variables, we have deliberately avoided using the term ‘sufficient’ for any particular level of performance. A number of previous studies have, indeed, identified certain sample sizes as ‘sufficient’, ‘enough’, ‘adequate’, or leading to ‘reliable’ results (Burdorf and van Riel, 1996; Allread *et al.*, 2000; Hoozemans *et al.*, 2001; Paquet *et al.*, 2005; Fethke *et al.*, 2007; Trask *et al.*, 2008). The criterion has, however, often been largely arbitrary and based on the premise that precision will not improve to any notable extent beyond this ‘sufficient’ sample size. Other studies have, on more formal grounds, identified the necessary sample size to obtain a specific precision of a mean exposure estimate (Mathiassen *et al.*, 2003b), the necessary study size to obtain ‘acceptable’ power in studies comparing independent groups (Mathiassen *et al.*, 2002),



or the necessary number of subjects or samples when testing interventions using individuals as their own control (Mathiassen *et al.*, 2003b; Mathiassen and Paquet, 2010).

Design requirements for an exposure data collection will also differ profoundly depending on whether the study is, e.g. devoted to documenting exposures in a specific occupational setting (as the present study), comparing mean exposures between groups or conditions (Mathiassen *et al.*, 2002; Mathiassen and Paquet, 2010), comparing exposures to threshold limit values (Lyles and Kupper, 1996), determining exposure-outcome relationships using either an individual-based or a group-based approach (Burdorf, 1995; Seixas and Sheppard, 1996; Tielemans *et al.*, 1998; Nordander *et al.*, 2004) or estimating sources and sizes of exposure variability (Eliasziv and Donner, 1987; Liv *et al.*, 2012). For each of these study types, the necessary sample size will further differ depending on the choice of summary statistics and the distribution of the selected exposures, as demonstrated in the present study (Tables 3b and 4b) and numerous other studies showing that variability within and between subjects differs among exposure variables (e.g. for working postures: van der Beek *et al.*, 1995; Burdorf and van Riel, 1996; Mathiassen *et al.*, 2003b; Hansson *et al.*, 2006; Bao *et al.*, 2009; Dartt *et al.*, 2009; Wahlström *et al.*, 2010).

Thus, the required statistical performance in any exposure data collection strategy, and hence the necessary sample size, is specific to the purpose, context, variables, and desired sensitivity of that particular study for which the sampling is carried out. In the present study, which focused on a descriptive documentation of exposures in a specific construction operation, 300 PATH observations would be sufficient for obtaining an 80% probability that the resulting estimate of the occurrence of neutral trunk postures is within 10% from the correct value (Table 3b). The same coverage probability for operation exposures to mild and severe trunk flexion would require 1200 and 1500 samples, respectively, while an assessment of twisted postures would not reach 80% coverage probability even with 4500 samples. All variables describing exposure variability between subjects and tasks (Table 3b) showed coverage probabilities below 80% even with 4500 samples, so even larger samples—probably in excess of what can be practiced in many occupational

studies—would be needed to reach a satisfying performance. For all four tasks, proportions could be determined with sufficient coverage probability by 900 samples (Table 4b), whereas only few task exposures and no task exposure variabilities reached sufficient coverage probability even with 4500 samples.

Although tentatively providing these guidelines for sampling in occupational settings and for metrics similar to those represented here, we wish to emphasize that prediction intervals will, for mathematical reasons, reach a width of zero (i.e. perfect precision) only at an infinite number of samples. Thus, 'saturation' or convergence to the 'true' value will never occur in a pure statistical sense. From a practical point of view, however, the return, in terms of improved precision with an increased number of samples, may, at some point, decrease below what is considered reasonable from a resource consumption perspective. A certain level of imprecision may even be deemed acceptable, and additional sampling beyond what is needed to achieve this statistical performance is then of limited value.

#### Understanding categorical work sampling and probabilistic simulation

A number of studies have addressed aspects of sampling performance for exposure variables measured on a continuous scale, using analytical expressions based on variance components (e.g. Mathiassen *et al.*, 2003a, 2003b; Kazmierczak *et al.*, 2006; Lampa *et al.*, 2006; Chen *et al.*, 2009; Jackson *et al.*, 2009). A large majority of these studies have been devoted to understanding the precision of mean exposure estimates. To our knowledge, very few attempts have been made to analyze statistical sampling performance for variables describing exposure variability, let alone task diversity or task contrasts. Standard analytical methods are applicable only for normally distributed and independent data, including data that can, *via* a suitable transformation, reach normality, which is standard practice for most exposure assessments in the field of occupational hygiene (Loomis and Kromhout, 2004). Although some violations of assumptions in standard analysis methods can be handled by modified statistical models, any analytical approach is limited to exposure variables for which sampling performance can be expressed in a closed-form equation, typically expressions of central tendencies (mean values).

We chose to investigate sampling performance for variables expressing exposure variability on the basis of virtual data sets obtained by a simulation procedure, rather than developing an analytical approximation. Our approach of using the probabilities of task and exposure occurrence for each worker as observed in the original PATH observation data set is an example of parametric simulation where data units are assumed to follow a known distribution, and virtual data sets are created by randomly selecting values from that distribution with a predetermined setup of parameters (Semple *et al.*, 2003). In the case of multinomial data, this parametric simulation may be particularly appealing since the distribution is fully characterized by the set of probabilities of ‘positive’ (‘yes’) outcomes in the categories within a set. As an alternative, we could have used non-parametric bootstrap resampling with replacement among the 3103 observations in the parent data set (Efron and Tibshirani, 1986; Burdorf and van Riel, 1996; Mathiassen and Paquet, 2010; Liv *et al.*, 2011). We chose not to do so because of its highly unbalanced structure: the 10 workers were represented to highly different extents, ranging from 68 to 706 data points (Table 1). To mimic a scenario where all workers worked full time and were equally available for observation, we assigned all workers an equal probability of being selected on any single occasion.

This decision also illustrates an attractive property of probabilistic simulation compared with straightforward bootstrap resampling with replacement, namely, that the probabilities assigned to the occurrence of different subjects and exposures can be manipulated so as to explore hypothetical scenarios that differ from the one represented by the parent data set. For instance, in a reorganization of an operation, workers may be assigned new proportions of constituent tasks, even if the overall occurrence of each task in the job does not change. Virtual redistributions of tasks among workers can easily be simulated by changing individual values of task proportions while maintaining the original values of overall task occurrence,  $w_{T\bullet}$ , and the overall mean job and task exposures  $\mu_{E\bullet}$  and  $\mu_{ET\bullet}$ . Additional scenarios accessible to probabilistic simulation are the introduction of new tasks and changes in individual exposures, e.g. following from an ergonomic intervention. Even alternative scenarios referring to the logistics of exposure assessment may be considered by probabilistic simulation.

For instance, some workers may not be accessible for observation during the whole period of data collection because they occasionally work in a location where observations are not feasible. This situation can be simulated by manipulating the probabilities of individual workers being selected for observation. Also the number of accessible workers sharing the tasks in the operation can be changed. We encourage more studies of data collection strategies using probabilistic simulation, especially for categorical data with binomial or multinomial distributions.

When the measurement method itself contributes uncertainty, which is inevitably true in observational studies (Denis *et al.*, 2000; Takala *et al.*, 2010), the variance on the eventual exposure estimate will include a methodological contribution. In our parent data set, variability within and between observers could not be distinguished from other sources of variability, and thus, it was not possible to determine the specific effects of between-observer and within-observer variability on the overall performance of the investigated sampling strategies. Other studies have, however, shown that between-observer reliability is good when PATH observations are made by trained observers (Park *et al.*, 2009), as in the present parent data set. With other observational methods and/or for other variables than those observed in PATH, observer variability has been shown to contribute significantly to the uncertainty of the eventual exposure estimate (Kazmierczak *et al.*, 2006; Dartt *et al.*, 2009; Rezaghali *et al.*, 2012; Mathiassen *et al.*, 2013). We therefore recommend that the effects of variability between and within observers be specifically addressed in future studies of strategies for observing categorical variables.

Finally, a further step in optimizing study designs would be to include considerations to the cost of different data collection strategies that lead to a satisfying statistical performance (Mathiassen and Bolin, 2011; Rezaghali *et al.*, 2012; Mathiassen *et al.*, 2013). Methods for cost-efficiency analyses are, in general, still in their infancy (Rezaghali and Mathiassen, 2010) and have so far been based only on analytical estimations of sampling performance. Analyses of cost-efficient exposure assessment using simulations, including specific investigations of categorical posture data obtained by work sampling, is a challenging issue for further research.

## CONCLUSIONS

The present study proposed a novel probabilistic simulation approach for categorical data and used it to reveal the statistical performance of observations of tasks and trunk postures obtained using a work sampling. Performance improved with increasing numbers of samples from 300 up to 4500. At each particular sample size, mean exposures were, in general, estimated with considerably better precision than variables describing aspects of exposure variability between workers and tasks; estimates of exposure variability were also biased at small sample sizes. Even with 4500 samples, variables describing exposure variability were not estimated with satisfying coverage probability, neither at the level of individual categories nor for compositional sets of categories. The simulation approach thus proved useful for examining the performance of alternative exposure assessment strategies, and we also claim that it can be used to explore hypothetical scenarios of exposure profiles, task occurrences, and access to workers when collecting data.

We believe that these results have a bearing in general on occupational exposure assessment studies where data are recorded in a categorical form. When planning such a study, an analysis—to the extent possible—of the performance of different design options is key to arriving at a data collection that can effectively provide information of the desired quality, as defined by the purpose of the study. The probabilistic simulation approach proposed in this paper is useful in this proactive process.

## SUPPLEMENTARY DATA

Supplementary data can be found at <http://annhyg.oxfordjournals.org/>.

## FUNDING

Collection of the original dataset was funded by the US National Institute for Occupational Safety and Health (NIOSH) through the Center to Protect Workers' Rights (CPWR) (U02/CCU308771, U02/CCU312014, U02/CCU317202). Data processing for the present study and cooperation between the involved research groups was supported by a grant from the Swedish Research Council for Health, Working Life and Welfare (Forte Dnr. 2009-1761).

## ACKNOWLEDGEMENTS

Observations of the jacking pit operation were made by Trevor Schell and Joel Garrett, under the supervision of Bryan Buchholz. We gratefully acknowledge Scott Fulmer, Victor Paquet, and SangWoo Tak for their assistance in documenting the jacking pit operation, and Per Liv for valuable comments on the analyses and interpretations of statistical performance.

## DISCLAIMER

The contents are solely the responsibility of the authors and do not necessarily represent the official views of NIOSH, CPWR, or Forte.

## APPENDIX

Precision of Mean Exposure Estimates Based on Categorical Data

The present calculation of group mean values of task proportions,  $w_{T\bullet}$ , and postures at the operation and task levels,  $\mu_{E\bullet}$  and  $\mu_{ET\bullet}$ , respectively, is based on a mean-of-means approach (Samuels *et al.*, 1985) in which exposure estimates are first obtained for each individual worker, and then averaged across workers to give the group mean. In a balanced data set containing an equal number of measurements for each worker, the variance,  $s_{\mu}^2$ , of a group mean obtained by a mean-of-means approach in a finite (limited) group is

$$s_m^2 = \left(1 - \frac{n_s}{N_s}\right) \frac{s_{BS}^2}{n_s} + \frac{s_{WS}^2}{N^*} \quad (1)$$

where  $n_s$ ,  $N_s$ , and  $N^*$  are the number of workers represented in the data set, the total number of workers in the finite target population, and the total number of samples in the data set, respectively, and  $s_{BS}^2$  and  $s_{WS}^2$  are the between-worker and within-worker variance components. If every worker in the target population delivers data to the mean,  $n_s$  equals  $N_s$ , and the contribution from between-worker variability vanishes in a 'finite population' effect as discussed in previous papers (Mathiassen *et al.*, 2003a; Liv *et al.*, 2011; Mathiassen *et al.*, 2013). With the present simulation procedure, where one worker was randomly selected for each data point from the 10 workers eligible for observation as per the 'standard' PATH protocol (Buchholz *et al.*, 1996), this 'saturation' will

more probably happen if the total number of measurements is 'large'. With  $N_s = 10$ , the probability of any one worker not being represented in the data set is only  $1.9 \cdot 10^{-11}\%$  with a full set of 300 observations, whereas it is approximately 5% with 50 observations, as might occur at the task level under the 300-sample strategy. Thus, all 10 workers were likely represented in all simulated data sets at both the operation and task level, except for some data sets at the task level when using the smallest sampling strategies, in particular if some task or posture category occurred very rarely for some worker(s).

Thus, in 'large' data sets, the group mean variance expressed by equation (1) is determined entirely by the uncertainty associated with estimating the mean exposures of individual workers within the group. Each basic unit of measurement in the present data, i.e., each single PATH observation, is a multinomial set of binary answers 'yes' or 'no' to whether posture and task at that very instant are within a particular category for the selected worker. If the probability of the answer 'yes' in a particular category is  $p$ , the expected variance between individual samples will, according to multinomial theory, be  $p(100 - p)$ , with  $p$  measured in percent, and the variance of a mean exposure estimate based on  $n$  samples is, therefore,  $p(100 - p)/n$ . For a certain sample size  $n$ , this variance is largest when the probability,  $p$ , of a 'yes' is 50% and will decrease symmetrically toward zero at probabilities above and below 50%.

Thus, the variance of a mean exposure estimate in a particular category for one worker is expected to be  $p(100 - p)/n$ , where  $p$  is the exposure probability (in percent) and  $n$  is the number of samples for that worker. When individual workers' mean exposures are averaged to form a group mean, the contribution from within-worker variability to the resulting variance on that mean is therefore

$$\frac{1}{N^* \cdot n_s} \sum_s [\mu_{E \cdot S} (100 - \mu_{E \cdot S})] \quad (2)$$

for operation exposures and similarly

$$\frac{100}{N^* \cdot w_{T \bullet} \cdot n_{s(T)}} \sum_{s(T)} [\mu_{ETS} (100 - \mu_{ETS})] \quad (3)$$

for exposures in task  $T$ , where  $N^*$  is the total number of samples at the operation level, cf. equation (1); notation otherwise as in Tables 1 and 2. Both equations assume that the same number of samples is obtained from all workers, at the operation level, equation (2), as well as the task level, equation (3). For an unbalanced data set, i.e. one with unequal numbers of observations for different workers, the variance is larger, especially if the data are severely unbalanced (Samuels *et al.*, 1985).

The estimates of occurrence for each of the four tasks in the jacking pit construction operation ( $w_{T \bullet}$ ; Table 4a and Supplementary Table S1a, available at *Annals of Occupational Hygiene* online) are equivalent, from a statistical point of view, to the estimates of operation and task postures since all are based on observations having an answer of 'yes' or 'no' to whether that particular unit represents each particular task or posture category. Thus, if all workers are represented in the data set, the variance of the estimated group mean proportion of any one of the four tasks can—for a balanced data set—be expressed as

$$\frac{1}{N^* \cdot n_s} \sum_s [w_{TS} (100 - w_{TS})] \quad (4)$$

If a 'large' number of binomially distributed samples are averaged, the mean value will be normally distributed. Thus, prediction intervals for postures and task proportions can be calculated by using the variances in equations (2–4) in a standard assessment of lower and upper limits associated with a coverage according to choice, *in casu* the 5th and 95th percentile.

## REFERENCES

- Aitchison J. (1986) The statistical analysis of compositional data. London: Chapman & Hall.
- 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.
- Bao S, Howard N, Spielholz P *et al.* (2009) Interrater reliability of posture observations. *Hum Factors*; 51: 292–309.
- Benke G, Sim M, Fritschi L *et al.* (2000) Beyond the Job Exposure Matrix (JEM): the Task Exposure Matrix (TEM). *Ann Occup Hyg*; 44: 475–82.
- Bovenzi M. (2009) Metrics of whole-body vibration and exposure-response relationship for low back pain in professional drivers: a prospective cohort study. *Int Arch Occup Environ Health*; 82: 893–917.

- Buchholz B, Paquet V, Punnett L *et al.* (1996) PATH: a work sampling-based approach to ergonomic job analysis for construction and other non-repetitive work. *Appl Ergon*; 27: 177–87.
- Burdorf A. (1995) Reducing random measurement error in assessing postural load on the back in epidemiological surveys. *Scand J Work Environ Health*; 21: 15–23.
- Burdorf A, van Riel M. (1996) Design of strategies to assess lumbar posture during work. *Int J Ind Ergon*; 18: 239–49.
- Burdorf A, van Riel M, Brand T. (1997) Physical load as risk factor for musculoskeletal complaints among tank terminal workers. *Am Ind Hyg Assoc J*; 58: 489–97.
- Burstyn I. (2009) Measurement error and model specification in determining how duration of tasks affects level of occupational exposure. *Ann Occup Hyg*; 53: 265–70.
- Chen C-C, Chuang C-L, Wu K-Y *et al.* (2009) Sampling strategies for occupational exposure assessment under generalized linear model. *Ann Occup Hyg*; 53: 509–21.
- Chen JC, Chang WR, Shih TS *et al.* (2004) Using exposure prediction rules for exposure assessment: an example on whole-body vibration in taxi drivers. *Epidemiology*; 15: 293–9.
- Christensen H, Pedersen MB, Sjøgaard G. (1995) A national cross-sectional study in the Danish wood and furniture industry on working postures and manual materials handling. *Ergonomics*; 38: 793–805.
- Dartt A, Rosecrance J, Gerr F *et al.* (2009) Reliability of assessing upper limb postures among workers performing manufacturing tasks. *Appl Ergon*; 40: 371–8.
- Dempsey PG, Mathiassen SE. (2006) On the evolution of task-based analysis of manual materials handling, and its applicability in contemporary ergonomics. *Appl Ergon*; 37: 33–43.
- Denis D, Lortie M, Rossignol M. (2000) Observation procedures characterizing occupational physical activities: critical review. *Int J Occup Safety Ergon*; 6: 463–91.
- Efron B, Tibshirani R. (1986) Bootstrap methods for standard errors, confidence intervals, and other measures of statistical accuracy. *Stat Sci*; 1: 54–77.
- Eliaszew M, Donner A. (1987) A cost-function approach to the design of reliability studies. *Stat Med*; 6: 647–55.
- Fallentin N, Juul-Kristensen B, Mikkelsen S *et al.* (2001) Physical exposure assessment in monotonous repetitive work—the PRIM study. *Scand J Work Environ Health*; 27: 21–9.
- Fethke NB, Anton D, Cavanaugh JE *et al.* (2007) Bootstrap exploration of the duration of surface electromyography sampling in relation to the precision of exposure estimation. *Scand J Work Environ Health*; 33: 358–67.
- Filzmoser P, Hron K, Reimann C. (2009) Univariate statistical analysis of environmental (compositional) data: problems and possibilities. *Sci Total Environ*; 407: 6100–8.
- Filzmoser P, Hron K, Reimann C. (2010) The bivariate statistical analysis of environmental (compositional) data. *Sci Total Environ*; 408: 4230–8.
- Fransson-Hall C, Gloria R, Kilbom Å *et al.*, Stockholm Music I Study Group. (1995) A portable ergonomic observation method (PEO) for computerized on-line recording of postures and manual handling. *Appl Ergon*; 26: 93–100.
- Frings-Dresen MHW, Kuijer PPFM. (1995) The TRAC-system: an observation method for analyzing work demands at the workplace. *Safety Sci*; 21: 163–5.
- Fry JM, Fry TRL, McLaren KR. (2000) Compositional data analysis and zeros in micro data. *Appl Econ*; 32: 953–9.
- Hansson G-Å, Arvidsson I, Ohlsson K *et al.* (2006) Precision of measurements of physical workload during standardised manual handling. Part II: Inclination of head, upper back, neck and upper arms. *J Electromyogr Kinesiol*; 16: 125–36.
- Harrison RM, Thornton CA, Lawrence RG *et al.* (2002) Personal exposure monitoring of particulate matter, nitrogen dioxide, and carbon monoxide, including susceptible groups. *Occup Environ Med*; 59: 671–9.
- Hooftman WE, van der Beek AJ, van de Wal BG *et al.* (2009) Equal task, equal exposure? Are men and women with the same tasks equally exposed to awkward working postures? *Ergonomics*; 52: 1079–86.
- 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*; 25: 3087–92.
- Hoozemans MJM, 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.
- Jackson JA, Mathiassen SE, Dempsey PG. (2009) Methodological variance associated with normalization of occupational upper trapezius EMG using submaximal reference contractions. *J Electromyogr Kinesiol*; 19: 416–27.
- Karhu O, Kansi P, Kuorinka I. (1977) Correcting working postures in industry: A practical method for analysis. *Appl Ergon*; 8: 199–201.
- Kazmierczak K, Mathiassen SE, Neumann WP *et al.* (2006) Observer reliability of industrial activity analysis based on video recordings. *Int J Ind Ergon*; 36: 275–82.
- Kivi P, Mattila M. (1991) Analysis and improvement of work postures in the building industry: application of the computerised OWAS method. *Appl Ergon*; 22: 43–8.
- Lampa EG, Nilsson L, Liljelind IE *et al.* (2006) Optimizing occupational exposure measurement strategies when estimating the log-scale arithmetic mean value - an example from the reinforced plastics industry. *Ann Occup Hyg*; 50: 371–7.
- Landon J, Singpurwalla ND. (2008) Choosing a coverage probability for prediction intervals. *Am Stat*; 62: 120–4.
- Li G, Buckle P. (1999) Current techniques for assessing physical exposure to work-related musculoskeletal risks, with emphasis on posture-based methods. *Ergonomics*; 42: 674–95.

- 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.
- Liv P, Mathiassen SE, Svendsen SW. (2012) Accuracy and precision of variance components in occupational posture recordings: a simulation study of different data collection strategies. *BMC Med Res Methodol*; 12: 58.
- Loomis D, Kromhout H. (2004) Exposure variability: concepts and applications in occupational epidemiology. *Am J Ind Med*; 45: 113–22.
- Lyles RH, Kupper LL. (1996) On strategies for comparing occupational exposure data to limits. *Am Ind Hyg Assoc J*; 57: 6–15.
- Mathiassen SE. (2006) Diversity and variation in biomechanical exposure: what is it, and why would we like to know? *Appl Ergon*; 37: 419–27.
- Mathiassen SE, Bolin K. (2011) Optimizing cost-efficiency in mean exposure assessment - cost functions reconsidered. *BMC Med Res Methodol*; 11: 76.
- 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: 27–39.
- Mathiassen SE, Burdorf A, van der Beek AJ *et al.* (2003a) Efficient one-day sampling of mechanical job exposure data—a study based on upper trapezius activity in cleaners and office workers. *Am Ind Hyg Assoc J*; 64: 196–211.
- Mathiassen SE, Liv P, Wahlström J. (2013) Cost-efficient measurement strategies for posture observations based on video recordings. *Appl Ergon*; 44: 609–17.
- Mathiassen SE, Möller T, Forsman M. (2003b) 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.
- Neitzel RL, Daniell WE, Sheppard L *et al.* (2011) Evaluation and comparison of three exposure assessment techniques. *J Occup Environ Hyg*; 8: 310–23.
- Neumann WP, Wells RP, Norman RW *et al.*, OUBPS Working Group. (2001) A posture and load sampling approach to determining low-back pain risk in occupational settings. *Int J Ind Ergon*; 27: 65–77.
- Nordander C, Balogh I, Mathiassen SE *et al.* (2004) Precision of measurements of physical workload during standardised manual handling. Part I: Surface electromyography of *m. trapezius*, *m. infraspinatus* and the forearm extensors. *J Electromyogr Kinesiol*; 14: 443–54.
- Paquet V, Punnett L, Buchholz B. (2001) Validity of fixed-interval observations for postural assessment in construction work. *Appl Ergon*; 32: 215–24.
- 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.
- Park J-K, Boyer J, Tessler J *et al.* (2009) Inter-rater reliability of PATH observations for assessment of ergonomic risk factors in hospital work. *Ergonomics*; 52: 820–9.
- Preller L, Kromhout H, Heederik D *et al.* (1995) Modeling long-term average exposure in occupational exposure-response analysis. *Scand J Work Environ Health*; 21: 504–12.
- Punnett L, Fine L, Keyserling WM *et al.* (1991) Back disorders and non-neutral trunk postures of automobile assembly workers. *Scand J Work Environ Health*; 17: 337–46.
- Reimann C, Filzmoser P, Fabian K *et al.*, The GEMAS Project Team. (2012) The concept of compositional data analysis in practice—total major element concentrations in agricultural and grazing land soils of Europe. *Sci Total Environ*; 426: 196–210.
- Rezagholi M, Mathiassen SE. (2010) Cost-efficient design of occupational exposure assessment strategies—a review. *Ann Occup Hyg*; 54: 858–68.
- Rezagholi M, Mathiassen SE, Liv P. (2012) Cost efficiency comparison of four video-based techniques for assessing upper arm postures. *Ergonomics*; 55: 350–60.
- Richardson W, Pape E (1982). Work sampling. In: Salvendy G, editor. *Handbook for industrial engineering*. New York: Wiley.
- Samuels SJ, Lemasters GK, Carson A. (1985) Statistical methods for describing occupational exposure measurements. *Am Ind Hyg Assoc J*; 46: 427–33.
- Seixas NS, Sheppard L. (1996) Maximizing accuracy and precision using individual and grouped exposure assessments. *Scand J Work Environ Health*; 22: 94–101.
- Semple SE, Proud LA, Cherrie JW. (2003) Use of Monte Carlo simulation to investigate uncertainty in exposure modeling. *Scand J Work Environ Health*; 29: 347–53.
- Susi P, Goldberg M, Barnes P *et al.* (2000) The use of a task-based exposure assessment model (T-BEAM) for assessment of metal fume exposures during welding and thermal cutting. *Appl Occup Environ Hyg*; 15: 26–38.
- 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: 18–26.
- 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.

- Tak S, Punnett L, Paquet V *et al.* (2007) Estimation of compressive forces on lumbar spine from categorical posture data. *Ergonomics*; 50: 2082–94.
- Takala EP, Pehkonen I, Forsman M *et al.* (2010) Systematic evaluation of observational methods assessing biomechanical exposures at work. *Scand J Work Environ Health*; 36: 3–24.
- 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, Teschke K, Morrison J *et al.* (2008) How long is long enough? Evaluating sampling durations for low back EMG assessment. *J Occup Environ Hyg*; 5: 664–70.
- Wahlström J, Mathiassen SE, Liv P *et al.* (2010) Upper arm postures and movements in female hairdressers across four full working days. *Ann Occup Hyg*; 54: 584–94.
- van der Beek AJ, Kuiper JI, Dawson M *et al.* (1995) Sources of variance in exposure to nonneutral trunk postures in varying work situations. *Scand J Work Environ Health*; 21: 215–22.
- van der Beek AJ, van Gaalen LC, Frings-Dresen MHW. (1992) Working postures and activities of lorry drivers: a reliability study of on-site observation and recording on a pocket computer. *Appl Ergon*; 23: 331–6.
- Village J, Trask C, Luong N *et al.* (2009) Development and evaluation of an observational Back-Exposure Sampling Tool (Back-EST) for work-related back injury risk factors. *Appl Ergon*; 40: 538–44.