Ann. Occup. Hyg., Vol. 53, No. 3, pp. 201–214, 2009
Published by Oxford University Press
on behalf of the British Occupational Hygiene Society 2009
doi:10.1093/annhyg/mep010

Agreement between Task-Based Estimates of the Full-Shift Noise Exposure and the Full-Shift Noise Dosimetry

M. ABBAS VIRJI[1]*, SUSAN R. WOSKIE[2], MARTHA WATERS[3], SCOTT BRUECK[4], DANIEL STANCESCU[5], REBECCA GORE[2], CHERYL ESTILL[3] and MARY PRINCE[3]

[1]Field Studies Branch, Division of Respiratory Disease Studies, National Institute for Occupational Safety and Health, MS 2800, 1095 Willowdale Road, Morgantown, WV 26505, USA; [2]Department of Work Environment, University of Massachusetts Lowell, Lowell, MA 01854, USA; [3]Industry Wide Studies Branch, Division of Surveillance, Hazard Evaluation and Field Studies, National Institute for Occupational Safety and Health, Cincinnati, OH 45226, USA; [4]Hazard Evaluation and Technical Assistance Branch, Division of Surveillance, Hazard Evaluation and Field Studies, National Institute for Occupational Safety and Health, Cincinnati, OH 45226, USA; [5]Office of Compensation Analysis and Support, National Institute for Occupational Safety and Health, Cincinnati, OH 45226, USA

Received 24 July 2008; in final form 26 January 2009; published online 12 March 2009

Noise assessments have been conducted using full-shift dosimetry and short-term task-based measurements. Advantages of the task-based method include the opportunity to directly identify high-noise exposure tasks and to target control measures, as well as obtain estimates of task-based full-shift exposures; however, there is little empirical evidence comparing the two methods. National Institute for Occupational Safety and Health assessed noise exposures at three industrial facilities using dosimetry and task-based methods with the objective of comparing the two strategies and assessing the degree of agreement and causes of disagreement. Eight indices of task-based full-shift exposures were created from task-based sampling using three methods to assess time-at-task (direct observation by industrial hygienist, end-of-shift worker estimates and supervisor estimates) and three methods to assign noise levels to tasks [direct measurement, arithmetic mean (AM) and geometric mean (GM)]. We assessed aspects of agreement (precision, bias and absolute agreement) using Bland–Altman plots and concordance correlation coefficient (CCC). Overall, the task-based methods worked fairly well, with mean biases less than ±2.8 dBA and precision ranges of 3.3–4.4 dBA. By all measures, task-based full-shift estimates based on supervisor assessment of time-at-task agreed most poorly with the dosimetry data. The task-based full-shift estimates based on worker estimates of time-at-task generally agreed as well as those based on direct observation. For task noise level, task-based full-shift estimates based on directly measured task agreed the best with dosimetry data, while agreement for task-based indices based on task AM or GM was variable. Overall, the task-based full-shift estimates based on direct observation task and direct measured task noise level achieved the best agreement with the dosimetry data (CCC 0.84) with 95% of their differences being within 7.4 dBA and 56% of the differences <3 dBA. For this index, a high degree of accuracy was observed (accuracy coefficient = 0.96) with major cause of disagreement arising from a lack of precision (precision coefficient = 0.88). When the measurements were classified by job characteristics, significant improvements in the degree of agreement were observed in the low job mobility, low job complexity and low job variability categories. Our data suggest that a high degree of absolute agreement can be achieved between the task-based and dosimetry-based estimates of full-shift exposures. The task-based approach that uses worker reports combined with task AM or GM levels is similar to the more time-intensive direct observation method to estimate full-shift exposures.

Keywords: accuracy; agreement; bias; concordance correlation coefficient; noise exposure; precision; task-based exposure assessment

---

*Author to whom correspondence should be addressed.
Tel: +1-304-285-5797; fax: +1-304-285-5820;
e-mail: mvirji@cdc.gov

INTRODUCTION

The full-shift time-weighted average (TWA) noise exposures measured via dosimetry is the most common noise exposure metric used in the USA. However, the task-based noise exposure assessment method offers several advantages over the traditional full-shift method, including opportunities to directly identify high exposure tasks for targeted controls (Hager, 1998). Moreover, the task-based method may yield more precise estimates of the mean exposures of occupational groups than estimates based on full-shift measurements, especially when task means and the time-at-task are highly variable (Nicas and Spear, 1993a,b; Benke *et al.*, 2000). In addition, task-based measurements permit a fuller characterization of exposure and the creation of alternative exposure metrics that incorporate time-varying measures including peak exposure for epidemiologic investigations (Seixas *et al.*, 2005). Recently, several studies have reported the use of task-based exposure assessment for noise, especially in the construction and forestry sectors (Neitzel *et al.*, 1999; Seixas *et al.*, 2001, 2003; Kerr *et al.*, 2002; Neitzel and Yost, 2002; Humann *et al.*, 2005; Vipperman *et al.*, 2007). Task-based exposure assessment strategy has also been used successfully in a variety of industries for a range of airborne and skin exposures (Goldberg *et al.*, 1997; Methner *et al.*, 2000; Susi *et al.*, 2000; Verma *et al.*, 2003, 2004; Kromhout *et al.*, 2004; Pronk *et al.*, 2006; Woskie *et al.*, 2008; Virji *et al.*, 2009).

While many potential strengths of the task-based method are well described (Susi and Schneider, 1995), there are a number of challenges associated with this approach including defining tasks, collecting multiple consecutive short-duration task samples (when continuous monitoring is unavailable) and accounting for time-at-task (Goldberg *et al.*, 1997; Seixas *et al.*, 2003). The definition of task remains a critical aspect of a task-based strategy, and lack of consistency in task definition makes it difficult to compare results among studies. A task may be defined with a high degree of specificity or more broadly depending on the sources and degree of task exposure variability. The assessment of time-at-task depends on the specificity of task definitions and may require direct observation for very specific task definitions or worker recording for broader definitions of task. Task exposure levels (Nieuwenhuijsen *et al.*, 1995; Goldberg *et al.*, 1997; Neitzel *et al.*, 1999; Meijster *et al.*, 2008), time-at-task or the frequency of tasks (Preller *et al.*, 1995; Hansen and Whitehead, 1988; Kalil *et al.*, 2004; Ross *et al.*, 2004) are important determinants of full-shift exposures and can influence the accuracy of task-based estimates of full-shift exposures. Unfortunately, little published empirical evidence exists comparing the two methods for estimating full-shift exposures. Seixas *et al.* (2003) and Reeb-Whitaker *et al.* (2004) have reported poor to moderate agreement between the task-based estimates and full-shift noise exposure among construction workers.

A number of statistical methods have been used to evaluate agreement between measurements made on a continuous scale including: Pearson correlation coefficient, paired *t*-test, linear regression, Bland–Altman plots and limits, intraclass correlation coefficients (ICCs) and the concordance correlation coefficients (CCCs) (White and van den Broek, 2004). Since these approaches measure different components of agreement, a combination is needed to fully assess the desired agreement characteristics. Absolute agreement (aggregate measure) incorporates a measure of precision (linearity or variation) and a measure of bias/accuracy (difference or distance from the unity line) (Carrasco and Jover, 2003; Haber and Barnhart, 2006). Neither the Pearson correlation coefficient nor the paired *t*-test adequately measure absolute agreement; the former only measures the degree of association (precision), while the latter only measures the average difference (bias) between two continuous measurements (Lin, 1989; Bedard *et al.*, 2000; Barnhart *et al.*, 2002; White and van den Broek, 2004). In industrial hygiene literature, the ordinary least squares (OLS) regression approach is often used to compare two measurement methods; however, because both methods are often measured with error, OLS regression provides biased estimates for the intercept (biased high) and the slope (biased low) (Ludbrook, 2002; Bland and Altman, 2003). Alternatives to OLS regression include error-in-variables models such as Deming or orthogonal regression (Linnet, 1998) or structural equation models (see for example, Middendorf *et al.*, 1999). The Bland–Altman plot of differences is commonly employed in clinical sciences to compare two measurement methods. The degree of agreement is assessed by evaluating: (i) the pattern of the plot of the difference between the two paired measurements against their means and (ii) the slope and intercept of a regression line through the difference points (Bland and Altman, 1986; Bland and Altman, 2003). The Bland–Altman plots provide valuable graphical representation of the association between two measurement methods.

The ICC and the CCC are commonly used aggregate measures of absolute agreement to compare two measurement methods (Nickerson, 1997; Carrasco and Jover, 2003; Haber and Barnhart, 2006). The ICC (special case 3A[A,1]) is calculated using variance components from two-way mixed models with the random effect of subject and fixed effect of method as described by Mcgraw and Wong (1996). The CCC is based on the mean of the squared difference between two measurement methods, which is then transformed into a correlation coefficient (the CCC) and has two components: (i) accuracy—the deviation of the fitted line from the concordance line (i.e. the line of perfect agreement) and (ii) precision—the deviation of each pair of observations from the fitted line (Lin, 1989; Lin, 2000; Lin *et al.*, 2002). The CCC and the ICC (when specified appropriately) both provide identical measures of absolute agreement and are one and the same (Nickerson, 1997; Carrasco and Jover, 2003; Haber and Barnhart, 2006). The appeal of the CCC arises from having a summary aggregate measure of agreement and the ability to disaggregate it into its components which allows for the evaluation of the sources of disagreement.

In this study, we use noise exposure data collected from three industrial facilities to compare task-based estimates of full-shift exposure to full-shift noise dosimetry measurements. Specifically, our objectives were to evaluate the degree of agreement between the two methods and to provide guidance on when it may be useful to consider a task-based approach. We used several statistical methods to evaluate the degree of agreement between the two methods and to investigate the potential sources of disagreement.

METHODS

The National Institute for Occupational Safety and Health conducted noise exposure assessment surveys at three industrial facilities in Quebec, Canada. Three noise surveys were conducted at each facility between June 2003 and January 2004 which included: (i) a polystyrene food container manufacturing plant, (ii) an aluminum can and bottle cap manufacturing plant and (iii) a heavy equipment repair facility and represented a range of workplaces, processes, job characteristics and noise exposure profiles. Brief descriptions of the facilities, processes and noise sampling strategy are provided below, details of which are documented elsewhere (Brueck *et al.*, 2006).

Workplace descriptions

At the polystyrene food container manufacturing facility, the production processes include melting plastic pellets, extruding plastic sheeting, molding container forms, thermoforming final product containers and recycling scrap plastic material. There are six operational departments with different job characteristics and noise exposures, generally ranging from 80 to 90 dBA. In some departments such as thermoforming and extrusion, there is a constant background noise levels from continuous operations, which occasionally exceeded 90 dBA during specific activities, and employees are cross-trained to work at various positions. In other departments such as molding, noise exposures are intermittent, characterized by the variety of equipment and tools used by employees, for example, milling machines, lathes, drill presses, hammers and drills. The scrap plastic grinding rooms are characterized by high average noise levels (98 dBA) with workers typically working 1.5–2 m from the grinder, while the maintenance workers are exposed to variable background noise ranging from >90 dBA during repairs when they are adjacent to the equipment to low noise levels in non-production areas such as the maintenance shop and warehouses.

At the aluminum can and bottle cap manufacturing facility, the production processes include cutting, pressing, extruding, printing, forming and stamping operations conducted on rolls and sheets of aluminum. There are seven operational departments with similar job characteristics and high continuous background average noise levels generally between 91 and 99 dBA for all the job titles, including intermittent impact noise. In the canning, assembly and lithography departments, high continuous noise levels are due to the type of processes, equipment and production rate. The employees in these departments are cross-trained to work at various positions and spend some time in close proximity to the line during setup and maintenance. A larger proportion of their time is spent at their workstations located several meters away. Operators of presses and machines in the assembly department are also exposed to high continuous background noise levels. These workers move around the area but also spend time in close proximity to the presses. Fork truck operators and maintenance workers are exposed to a wide variety of background noise levels.

The heavy equipment repair facility services a variety of heavy equipment including tractors, earthmovers, skid-steer loaders, forklifts, backhoes, trenching machinery, paving equipment, hydraulic excavators, generators and ore-carrying trucks. Repairs range from maintenance and replacement of worn parts to complete rebuilding of engines, transmissions, wheel tracks, buckets and blades. Mechanics and machinists use a variety of power tools, impact wrenches, hammers, welding equipment, cutting torches and specialty machining equipment for repair work. Much of the noise exposure in the facility is from impact noise, generated primarily in the equipment repair areas. Impact noises generated in the repair areas of the facility tend to be random and unpredictable because of the specific nature of repairs required. The work tasks conducted by employees in this facility depend on the nature of repair work required. The workers do not always have the same repair assignments from day-to-day or repair the exact same equipment every day. However, workers do perform some similar tasks, such as using pneumatic wrenches to disassemble or reassemble parts.

Sampling strategy

A key objective of this study was to compare full-shift noise dosimetry measurements to the task-based estimates of full-shift noise exposures based on: (i) three methods of assessing time-at-task including direct observation by trained technicians, daily worker estimates and supervisor estimate and (ii) three methods for estimating short-term task levels (STTLs), including direct monitoring of STTLs during field sampling, calculation of the overall arithmetic mean (AM) STTLs and calculation of the overall geometric mean (GM) STTLs. Based on pilot studies conducted at the three facilities, job groups with potential noise exposure >85 dBA were identified for monitoring. The job groups were based on similarities in job functions, work tasks, mobility characteristics and exposure sources. Seven volunteers from each of these job groups were selected to participate in the study. Full-shift dosimetry measurements and the worker diaries of time-at-task were collected from participating workers during each of the three noise surveys. The measurement of STTLs and the direct observation of time-at-task were conducted for all workers during the first survey and a subset of workers during the second and third surveys. The supervisor time-at-task estimate was completed once during the study period for each job group.

Definition and identification of tasks

Considerable effort was devoted toward identifying and defining tasks prior to the start of the study. Tasks were defined as either a single activity or step in the production, repair or maintenance process or an activity that used a similar type of tool, equipment or machinery. In practice, the final list of tasks was achieved after a series of steps including reviewing job descriptions, observing work activities during a pilot study, discussing with plant personnel and developing a preliminary task list, conducting more detailed observations of the jobs and modifying the task list and preparing the final task list for data analysis by consolidating similar tasks. The details of these steps are provided elsewhere (Brueck *et al.*, 2006).

Noise instrumentation

Noise measurements were collected using two different models of Larson Davis (Provo, UT, USA) Spark Series Type-2 noise dosimeters with the same dynamic response characteristics and performance capabilities (Brueck *et al.*, 2006). The model 705P was used for personal noise dosimetry while the model 706RC was used for measuring the STTLs. The instruments were set to 'A' frequency weighting, 0 dB threshold, 3 dB exchange rate, slow meter response, 85 dB criterion level and 1-s averaging time. Calibrations were performed before and after each use with a Larson Davis Model CAL200 noise calibrator.

Field sampling

Full-shift noise dosimetry, task-based noise measurements (STTLs) and the direct observation and worker reporting of time-at-task were collected simultaneously from participating workers. For full-shift dosimetry, microphones were attached in an upright position at the center of workers' shoulder on the side of their dominant hand. The STTLs were collected during a typical and representative 30 s to 2 min portion of task based on the researcher's judgment. Dosimeters were held in the workers' hearing zone on the side of their dominant hand during at least one occurrence of each representative task. The name of task, task location, STTLs, time of measurement and duration of measurement were recorded on sampling forms.

For the direct observation of time-at-task, a technician (trained by an industrial hygienist on identifying tasks) observed a single worker throughout the work shift and documented the tasks conducted, start and stop times and auxiliary information about the job characteristics and work environment (such as use of hearing protection, tools and sources of noise). For the worker time-at-task estimate, workers were asked to complete an activity–time log immediately after the end of their work shift on which they noted the tasks conducted and the time spent at each task. If the difference in the estimated total time and actual time worked was >10 min, the technician assisted workers in recalling the tasks and revising task times. For the supervisor assessment, at least one supervisor of the workers monitored was asked to provide an estimate of the average time-at-task for a typical worker in the job group. The supervisors reviewed a list of tasks associated with the job group and estimated the amount of time that employees typically spent on each task during an average workday. The technician ensured that the total task time was equivalent to the length of a typical workday for that job group.

Calculation of full-shift and task-based full-shift exposures

The full-shift dosimeter time history records (start and end times) were synchronized with the actual work start and end times. Full-shift noise levels were calculated for each worker (dosimeter) based on the 1-s averaging of noise levels using the standard formula for averaging noise exposures,

$$L_{\text{eq-FS}}(\text{dB}) = q \times \log_{10}\frac{1}{N}\left(\sum_{i=1}^{N} 10^{L_{\text{pi}}/q}\right), \quad (1)$$

where, $L_{\text{eq-FS}}$ is the full-shift noise exposure, L_{pi} are the 1-s average noise levels downloaded from the dosimeter measured over $i = 1$ to N seconds over the full shift and q is the exchange rate (3 dB) divided by $\log_{10}$ of 2.

The STTLs recorded in the field sampling forms were used in conjunction with the information on time-at-task to calculate eight indices of the task-based full-shift noise exposure based on the combinations of the three different methods for assessing time-at-task and the three methods of assigning noise exposures to tasks (Table 1). For the two indices based on the direct measurement of task STTLs, the time-at-task was matched with STTL measured on the same person, day and the specific occurrence of a task. However, not all the occurrences of a task within a day were monitored, hence a hierarchical approach was used to fill in the next best exposure data starting with: mean task level from the person on the same day, mean task level from all workers performing the same task for the same job on the day, mean task level from all workers performing the same task for the same job on for all the days monitored and mean task level from all workers performing the same task for all the jobs and days monitored. For the remaining six indices based on a summary noise exposure level, the overall GM and AM of the tasks were assigned to the time-at-task. Since noise measurements are reported in the dB scale which incorporates a reference level and the log scale, task measurements were converted back to the log-normal pascal^2 scale, and the AM and GM of the tasks were calculated in pascal^2 and then these summary measures were converted back to the dB scale. It should be noted that the AM of the tasks calculated in pascal^2 and converted back to dB is equivalent to the task average calculated in dB using equation 1.

Table 1. Indices of task-based full-shift exposures

Task-based TWA indices	Time-at-task	Task noise exposure
DOD-measured	Direct observations diary	Measured task noise levels
DOD-AM	Direct observations diary	AM task noise levels
DOD-GM	Direct observations diary	GM task noise levels
WKD-measured	Worker diary	Measured task noise levels
WKD-AM	Worker diary	AM task noise levels
WKD-GM	Worker diary	GM task noise levels
SUP-AM	Supervisor summary	AM task noise levels
SUP-GM	Supervisor summary	GM task noise levels

The eight indices of task-based full-shift noise exposures were calculated in dB using the standard noise formula,

$$L_{\text{eq-TBFS}}(\text{dB}) = q \times \log_{10}\left(\frac{1}{T}\sum_{i=1}^{k} t_i \times 10^{L_{\text{pi}}/q}\right), \quad (2)$$

where, L_{pi} is the directly measured STTL, the overall GM or the AM of the task; t_i is the time-at-task from the direct observation of tasks time, the worker reporting of task time or the supervisor assigned task time for $i = 1$ to k tasks and T is the total time worked by the worker in the job on the day of monitoring or the duration of a typical workday for the job group. The use of directly measured STTL in equation 2 is preferable over the use of either the overall AM or GM because these summary measures may not be representative of individual workers task exposure on any given day when task exposures are highly variable. It is also noteworthy that the task-based full-shift noise exposure using task GM in equation 2 is mathematically not equivalent to the full-shift dosimetry measurement (calculated using equation 1 based on 1-s averaging of noise levels). However, the GM-based indices may have some value in epidemiologic investigations particularly if the association between exposure and dose or heath affects is a non-linear function of exposure (Seixas *et al.*, 1988; Smith, 1992). Thus, the choice of GM or AM in creating exposure metrics for use in epidemiology may in part depends on the mechanism relating exposure to outcome.

Job indices

Indices of certain job characteristics were created to investigate aspects of jobs that lead to good or poor agreement between the task-based methods and the full-shift noise dosimetry. An index of job complexity was calculated based on the number of tasks performed by workers on the day of sampling as reported in the direct observation of tasks. The index was categorized into approximately equal size tertiles of high (more than or equal to seven), medium (five to six) and low (less than or equal to four) tasks performed per day. Job mobility was rated as the percent of the workday (range: 0–100%) that a typical worker was judged by the industrial hygienist to be mobile [i.e. not working on tasks within a six foot (1.8 m) radius of a work position]. Jobs mobility was then categorized into approximately equal size tertiles of low ($\leq$30%) moderate ($>$30 to $\leq$40%) and high ($\geq$40%) mobility. Job exposure variability was calculated as the coefficient of variation (percent) of the full-shift noise dosimetry measurements and was categorized into tertiles of low ($<$2%), moderate (2–4%) and high ($>$4%) exposure variability.

Statistical analysis

Statistical analyses were performed using PC-SAS version 9.1 (SAS Institute, Cary, NC, USA). The distributions of all the full-shift TWA indices were evaluated graphically using probability plots and summary statistics were calculated. To assess the degree of agreement between the eight task-based indices of full-shift noise exposure and the full-shift noise dosimetry measurements, Bland–Altman plots of differences including absolute and percent difference (bias) and ratios of the measurements were calculated. Scatter plots and Bland–Altman plots of differences were prepared in SigmaPlot 9.01 (Systat Software Inc., San Jose, CA, USA). The CCC and its 95% confidence intervals were calculated in SAS according to the method of moments described by Lin (1989), using a macro made available by Crawford *et al.* (2007) (see download information in Appendix 1). The CCC estimates from the macro were cross-checked against values from a second SAS macro provided by Lin *et al.* (2002) (see download information in Appendix 1), which also calculated the precision and accuracy components of CCC. The total deviation index ($\text{TDI}_{0.95}$) which is a measure of the absolute value of the difference (in dBA) below which 95% of the differences lie and the coverage probability ($\text{CP}_{3\text{-dBA}}$) which describes the fraction (percent) of the measurement differences that lie within 3 dBA were also calculated in SAS as described by Lin *et al.* (2002) using their macro. To examine the impact of job characteristics on the degree of agreement between the task-based method and the full-shift measurements, the CCC, accuracy and precision coefficients were calculated stratified by the job characteristic indices of job mobility, complexity and variability as well as by facility.

RESULTS

Out of a total of 361 dosimeter measurements collected on 148 workers, 198 dosimeter measurements on 128 workers were matched to the eight task-based full-shift noise exposures and used to assess the agreement between the two measurement methods. One array of matched measurements was available for 90 workers, while repeated measurements were available for the remaining 38 workers. The distribution of the full-shift noise dosimetry data was approximately normal based on probability plots, but the task-based full-shift indices showed strong tendencies toward bimodal or multimodal distributions, consistent with findings from other studies (Smith *et al.*, 1991; Nicas and Spear, 1993a). The log-transformed indices of task-based full-shift exposures did not fit the normal distribution any better than the untransformed exposure indices. Descriptive summary statistics for the three facilities using all the dosimeter measurements collected are reported in Table 2. Full-shift mean noise exposures were 84.5, 87.8 and 95.8 dBA for the three facilities and were most variable for the heavy equipment repair facility and least variable for the aluminum cans and bottle cap facility. The equipment repair facility also had the highest median scores for all three indices of job characteristics (i.e. complexity, mobility and variability), whereas the aluminum cans and bottle cap facility had the lowest median scores.

Assessment of agreement

Table 3 reports the results of simple calculations of the means of the difference (bias) between the dosimetry-based and task-based indices including the percent difference, ratio of the two methods, standard deviation of the difference (precision) and a measure of accuracy calculated as: $\left[100-\sqrt{\text{bias}^2+\text{precision}^2}\right]$. This accuracy measure is generally used to compare two paired measurements (Hornung, 1991; Reeb-Whitaker *et al.*, 2004). The results reported in Table 3 do not show a consistent pattern in these measures for the various task-based indices. However, all four measures identified the DOD-AM (see Table 1 for abbreviation descriptions for the various task-based indices of full-shift exposures) index as having the best agreement with the dosimetry data and the SUP-AM index as having the poorest agreement. All task-based indices showed a high degree of accuracy (95.1–96.7%) and very little bias (range: 0.24 to −2.82 dBA).

The results of the Bland–Altman method are reported in Figs 1–3. The top rows (scatter plots) show the spread of the data around the unity line, which marks perfect agreement. The bottom rows

Table 2. Descriptive summary of the facilities

Plant descriptions	Number of departments	Number of workers	Number of jobs	Number of tasks	Median complexity score	Median mobility score	Median variability score	Dosimeter (dBA)	
								n	Mean (SD)
Polystyrene containers facility	5	69	13	53	5 (2–10)	40 (15–80)	2.5 (2.1–5.5)	162	87.8 (4.5)
Aluminum cans and bottle caps facility	3	49	13	44	4 (2–8)	25 (15–80)	0.8 (0.5–2.9)	116	95.8 (2.8)
Heavy equipment repair facility	5	30	12	35	6 (3–12)	40 (30–60)	4.3 (2.7–7.3)	83	84.5 (5.7)

(difference plots) show the mean difference as well as the distribution of the differences around the zero line (no difference/bias). Generally, a small negative bias was observed for most of the task-based indices, but there is a wide distribution of the individual differences within the task-based indices. The supervisor assessment-based indices (Fig. 3) clearly show the poorest agreement; however, the distinctions among the other tasks-based indices (Figs 1 and 2) are not very clear, making it difficult to rank or differentiate among several competing alternative measurement methods.

The estimates of CCC from the two macros yielded identical agreement coefficients, validating the use of these macros (Table 4). These results suggest that the best agreement was achieved by DOD-measured exposure index, followed by DOD-AM, WKD-measured and WKD-GM. These four indices had similar CCC values with overlapping 95% confidence intervals. The SUP-AM-based index

Table 3. Regression analysis, ratio and difference

TB indices[a]	Ratio (range)	Difference[b] (dBA) (%)[c]	Std of difference[d] (dBA)	Accuracy (%)[e]
DOD-measured	0.98 (0.86–1.15)	−1.73 (−1.97)	3.37	96.2
DOD-AM	1.00 (0.89–1.17)	0.24 (0.40)	3.31	96.7
DOD-GM	0.97 (0.83–1.09)	−2.82 (−3.16)	3.33	95.6
WKD-measured	0.99 (0.82–1.13)	−0.74 (−0.80)	3.74	96.2
WKD-AM	1.01 (0.90–1.14)	1.01 (1.30)	3.74	96.1
WKD-GM	0.98 (0.82–1.08)	−1.88 (−2.06)	3.38	96.1
SUP-AM	1.03 (0.90–1.20)	2.19 (2.72)	4.38	95.1
SUP-GM	0.99 (0.86–1.14)	−0.93 (−0.92)	3.82	96.1

[a] Task-based (TB) indices described in Table 1.

[b] $\text{Difference(Bias)} = \frac{1}{N}\sum \text{TWA}_{\text{task-based full-shift}} - \text{TWA}_{\text{dosimetry full-shift}}$

[c] $\%\,\text{Difference} = \left(\frac{1}{N}\sum \frac{\text{TWA}_{\text{task-based full-shift}} - \text{TWA}_{\text{dosimetry full-shift}}}{\text{TWA}_{\text{dosimetry full-shift}}}\right) \times 100$

[d] $\text{Std of difference}\big(\text{Precision}\big) = \text{Standard deviation of}\big(\text{TWA}_{\text{task-based full-shift}} - \text{TWA}_{\text{dosimetry full-shift}}\big)$

[e] $\%\,\text{Accuracy} = 100 - \left(\sqrt{(\text{Bias})^2 + (\text{Precision})^2}\right)$

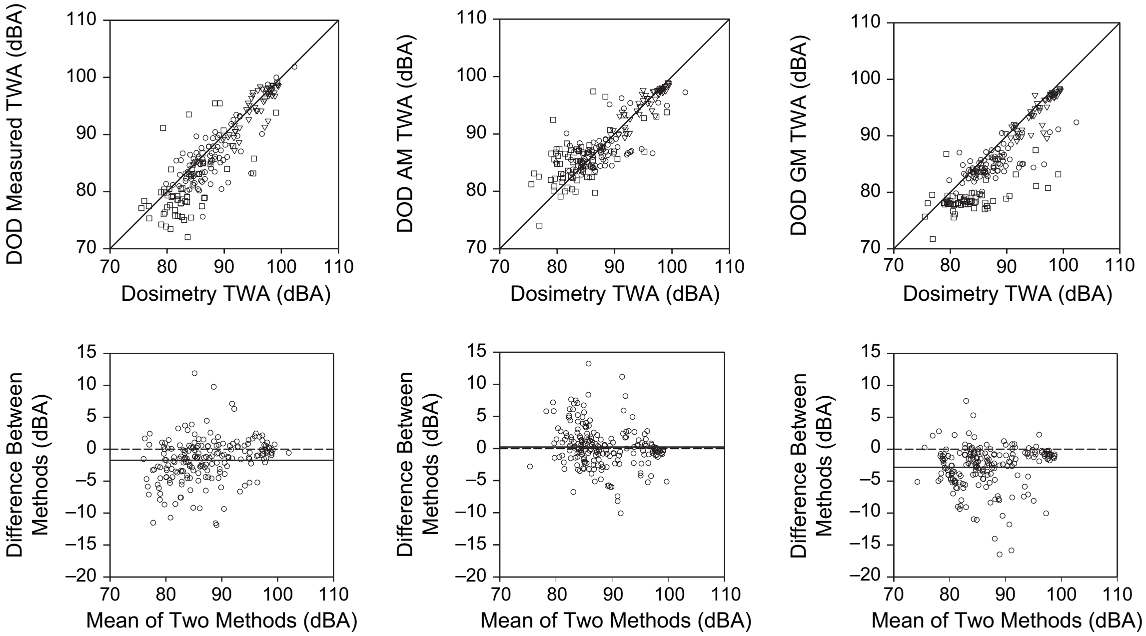

Fig. 1. Bland–Altman plots of agreement and differences for the direct observation-based full-shift exposure indices. See Table 1 for abbreviations. Open circles, polystyrene containers facility; open triangles, aluminum cans and bottle cap facility; open squares, heavy equipment repair facility; dashed line, no bias reference line; solid line, mean difference line.

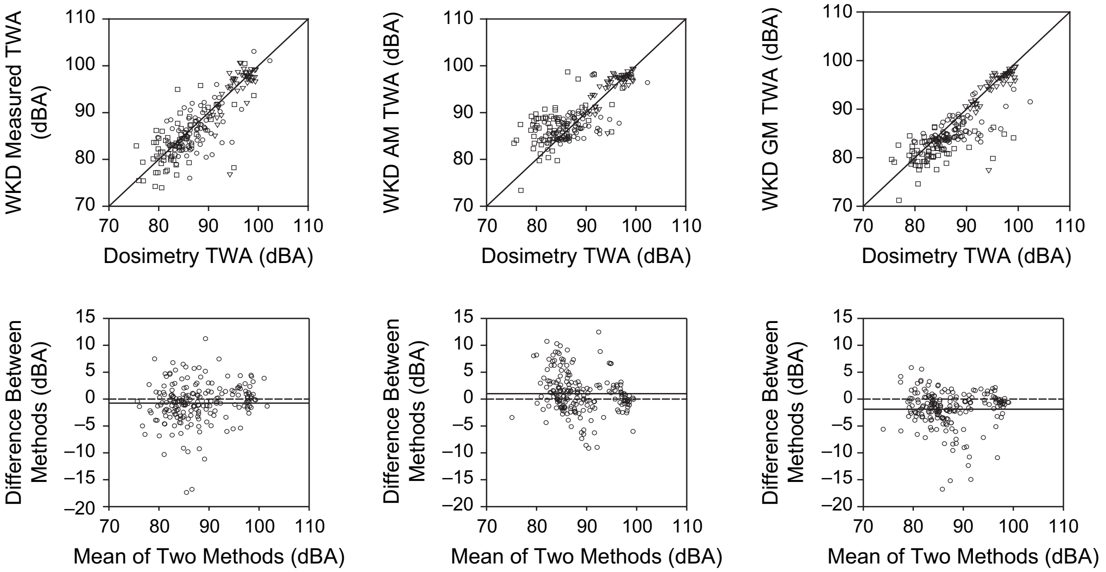

Fig. 2. Bland–Altman plots of agreement and differences for the worker diary-based full-shift exposure indices. See Table 1 for abbreviations. Open circles, polystyrene containers facility; open triangles, aluminum cans and bottle cap facility; open squares, heavy equipment repair facility; dashed line, no bias reference line; solid line, mean difference line.

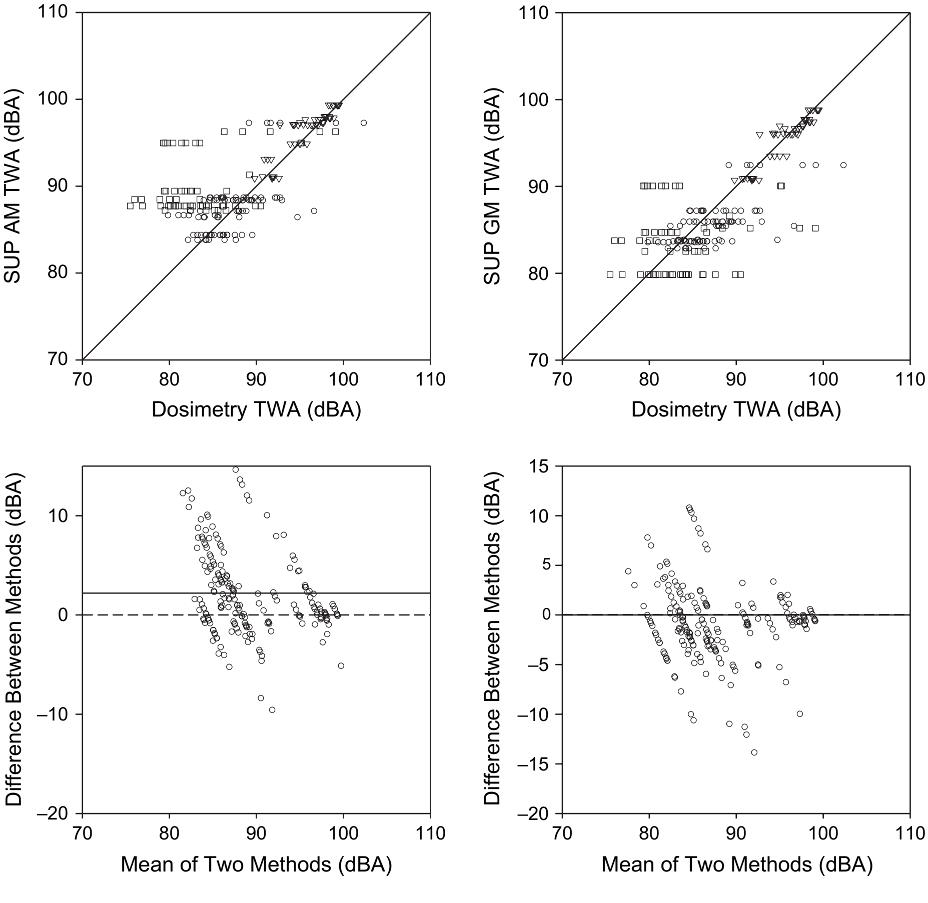

Fig. 3. Bland–Altman plots of agreement and differences for the supervisor assessment-based full-shift exposure indices. See Table 1 for abbreviations. Open circles, polystyrene containers facility; open triangles, aluminum cans and bottle cap facility; open squares, heavy equipment repair facility; dashed line, no bias reference line; solid line, mean difference line.

showed the poorest agreement. Generally, the direct observation-based indices performed the best, followed by the worker diary-based and then the supervisor assessment-based indices. Within direct observation- or worker diary-based methods, the indices based on directly measured task exposure performed better than the summary-based indices (GM or AM); the latter gave inconsistent agreement results with sometimes GM- or AM-based indices showing superior agreement.

Results of the coefficients of precision and accuracy, the two components of the CCC, are also reported in Table 4. The precision coefficients were similar to the CCC in pattern, while no particular pattern was observed with the accuracy coefficient. Generally, a high degree of accuracy was observed for all the task-based indices, suggesting very little average deviation from the concordance (agreement) line. Most of the source of disagreement arose from a lack of precision. The high values for the CCC, precision and accuracy coefficients for the direct observation- and worker diary-based indices are encouraging, but the TDI and CP measures suggest that caution is needed. The $TDI_{0.95}$ suggests that 95% of the differences were <6.5 dBA at best for the DOD-AM index and <9.6 dBA at worst for the SUP-AM index. For noise measurements, this represents a large degree of difference in exposure between the dosimeter-based and task-based indices. The $CP_{3\text{-dBA}}$ shows that a little >50% ($CP_{3\text{-dBA}}$ range: 45%–63%) of the differences were <3 dBA, a value representing the exchange rate between dBA and sound energy doubling.

Table 5 contains results of the CCC, precision and accuracy coefficients stratified by job complexity. An increasing trend in agreement was observed with decreasing job complexity categories. The highest degree of agreement by all three measures, CCC, precision and accuracy, were observed for low-complexity job category. A majority of the gain in CCC across the complexity classification arose from improvement in the precision measure, with the accuracy measure changing little from the values for high and medium categories. It is noteworthy that all the task-based indices including those based on supervisor assessment showed this high degree of agreement for the low-complexity category. Similarly, results for job mobility show that the highest degree of agreement was achieved for low job mobility category while the degree of agreement was similar for medium and high job mobility categories (data not shown). For low job mobility category, high CCC was observed for all the task-based exposure indices (CCC range: 0.82–0.91), including a high

Table 4. CCCs, precision, accuracy and coverage statistics

TB indices[a]	CCC (SE)	Precision coefficient	Accuracy coefficient	$TDI_{0.95}$	$CP_{3\text{-dB}}$
DOD-measured	0.84 (0.020)	0.88	0.96	7.43	0.56
DOD-AM	0.83 (0.021)	0.84	0.99	6.51	0.63
DOD-GM	0.78 (0.025)	0.86	0.91	8.55	0.48
WKD-measured	0.82 (0.023)	0.83	0.99	7.47	0.57
WKD-AM	0.77 (0.028)	0.79	0.97	7.60	0.56
WKD-GM	0.80 (0.024)	0.84	0.95	7.59	0.55
SUP-AM	0.62 (0.039)	0.70	0.90	9.61	0.45
SUP-GM	0.78 (0.028)	0.79	0.98	7.71	0.55

$n = 198$ for all rows. Precision = Pearson correlation coefficient.
[a]Task-based (TB) indices described in Table 1.

Table 5. CCCs for job complexity

TB indices[a]	High-complexity jobs			Medium-complexity jobs			Low-complexity jobs		
	CCC (SE)	Precision	Accuracy	CCC (SE)	Precision	Accuracy	CCC (SE)	Precision	Accuracy
DOD-measured	0.75 (0.058)	0.78	0.96	0.84 (0.025)	0.89	0.95	0.92 (0.028)	0.95	0.97
DOD-AM	0.69 (0.067)	0.71	0.97	0.84 (0.027)	0.85	0.99	0.95 (0.019)	0.95	1.00
DOD-GM	0.50 (0.077)	0.67	0.75	0.82 (0.028)	0.89	0.92	0.94 (0.018)	0.98	0.96
WKD-measured	0.62 (0.084)	0.62	1.00	0.86 (0.023)	0.88	0.98	0.92 (0.030)	0.93	0.98
WKD-AM	0.57 (0.087)	0.60	0.95	0.79 (0.033)	0.82	0.97	0.94 (0.025)	0.94	1.00
WKD-GM	0.50 (0.088)	0.59	0.85	0.85 (0.025)	0.88	0.96	0.96 (0.015)	0.98	0.98
SUP-AM	0.53 (0.084)	0.62	0.85	0.61 (0.052)	0.69	0.89	0.79 (0.071)	0.81	0.97
SUP-GM	0.57 (0.086)	0.61	0.93	0.80 (0.034)	0.80	0.99	0.93 (0.026)	0.94	1.00

$n = 56$ for high-complexity, $n = 114$ for medium-complexity and $n = 28$ for low-complexity jobs.
[a]Task-based (TB) indices described in Table 1.

degree of precision (precision range: 0.86–0.93) and accuracy coefficients (accuracy range: 0.95–1.0). Similar results were obtained for job variability which showed the highest degree of agreement for low-variability job category across all task-based indices (data not shown).

Finally, we examined the CCC stratified by facility (data not shown). Generally, the highest degree of agreement was observed for the aluminum cans and bottle cap facility which also had the least variable exposure (although the highest mean exposures) and the fewest tasks (Table 2). The poorest agreement was observed for the equipment repair facility which had the highest exposure variability and most number of tasks. However, no pattern was observed among the various task-based indices for any of the three facilities with different indices showing better agreement at the three facilities.

DISCUSSION

In recent years, there has been a renewed interest in the task-based exposure assessment method. The task-based method offers an important advantage over the traditional full-shift method by providing the opportunity to directly identify high-exposure tasks for targeted controls. For example, in a study of lead exposures among construction workers during bridge rehabilitation, Goldberg *et al.* (1997) reported significant differences in exposure levels associated with tasks, which were partially obscured when full-shift exposures were calculated making the selection of respiratory protection less clear. Both task exposure levels and time-at-task can impact full-shift exposure levels. For example, Andersson and Rosén (1995) showed that in carpentry operations, the task with the highest exposures had less impact on daily exposures than the task with lower exposures but conducted for a larger proportion of time. In studies of construction trades workers, noise exposure levels were found to differ significantly among tasks or tools used, but did not significantly differ among trades (Neitzel *et al.*, 1999; Seixas *et al.*, 2001). The task-based exposure assessment strategy focuses on the relative importance of task exposure levels and task time in prioritizing control interventions, thus providing distinct advantages over the full-shift method, particularly for jobs with high variation in task exposure levels within a day, highly variable proportion of time-at-task between days and highly variable time-at-task between workers (Seixas *et al.*, 2003).

Task-based exposure measurements can be used in conjunction with the time-at-task to obtain estimates of daily or long-term average exposures for use in epidemiologic investigations. However, there is little empirical evidence in the literature comparing the task-based to the full-shift monitoring methods in estimating the full-shift exposures or the average exposures of occupational groups. Nicas and Spear (1993a,b) presented a task-based model that suggests that task-based methods may be more efficient (in terms of sampling) and may yield more precise estimates of the mean occupational group exposures than estimates based on full-shift measurements, especially when task means and the time-at-task are highly variable. However, Smith *et al.* (1997) showed through simulations that the task-based method does not improve the precision of the mean of an occupational group when the variation of the time-at-task is appropriately accounted for. Benke *et al.* (2000) compared estimates of cumulative exposure based on a job exposure matrix (JEM) and a task exposure matrix (TEM) and found significant difference between the two methods. They suggest that exposure estimates based on TEM can reduce exposure misclassification in epidemiologic studies and have used the TEM-based estimates to examine associations with respiratory symptoms and lung-function changes in aluminum smelters and bauxite miners (Beach *et al.*, 2001; Fritschi *et al.*, 2003). Woskie *et al.* (2008) found that a task-based exposure estimation algorithm was a better exposure index than estimates based on exposure surrogates such as job title or event counts in an exposure–response model of cross-week changes in lung function among auto body shop workers. In evaluating the utility of the task-based method for use in epidemiologic research, a number of questions remain unanswered including (i) the degree of agreement between task-based and full-shift measures of exposure, (ii) whether task-based or full-shift exposure measurements can estimate the means of occupational groups more precisely and with larger contrast between groups and (iii) the performance of task-based and full-shift exposure estimates in epidemiologic exposure–response relationships. In this study, we addressed the issue of the degree of agreement between the task-based and full-shift estimates of daily exposure using noise exposure data collected from three industrial facilities and examined various statistical methods to assess the degree of agreement.

Choice of task-based exposure metric

To compare our results to the published literature, we estimated R^2 for the comparison of the full-shift dosimetry to the task-based estimates by squaring the precision coefficient in Table 4, which showed moderate to high degree of agreement (R^2: 0.49–0.77). Indices based on direct observations (R^2: 0.7–0.77) yielded better agreement than worker reports (R^2: 0.63–0.71) or supervisor assessments (R^2: 0.49–0.62), which are all higher than those reported for construction workers (R^2: 0.62) (Reeb-Whitaker *et al.*, 2004). Likewise, Seixas *et al.* (2003) found low to moderate overall agreement

(R^2: 0.10–0.55) between the task-based estimates and full-shift noise exposure in the construction industry depending on the specificity of task definition, with substantial improvement in the degree of agreement (R^2: 0.89–0.90) with increasing specificity of tasks which resulted in less variable estimates of task exposures. One major difference between the present study and the study reported by Seixas *et al.* (2003) is that the task-based noise levels in their study were derived directly from the dosimetry data used in the comparison, whereas in our study, the task-based estimates were independent of the dosimetry data.

Results of the overall CCC suggests that task-based indices based on either direct observation or worker reports agreed well with the full-shift dosimetry data, while indices based on supervisor assessment had lower degree of agreement. This suggests that the worker diary-based approach is as good as the more time-intensive direct observation method. Reeb-Whitaker *et al.* (2004) reported a substantial degree of agreement between worker-reported and researcher-observed tasks performed ($\kappa = 0.67$). Similarly, Neitzel *et al.* (1999) and Seixas *et al.* (2001) reported high degree of agreement between worker recording and researcher observations of tasks with kappa values of $\kappa = 0.87$ and 0.86, respectively. In addition, when the workers were asked to recall their time-at-task 6 months after they had initially recorded them, the study found a high degree of accuracy in the recall of percent time-at-task (range: 53–100%, median: 91%). The number of tasks performed by workers per day ranged from one to four tasks (mean = 2.5 tasks day^{-1}), out of a total of 22 tasks (Reeb-Whitaker *et al.*, 2004). Benke *et al.* (2000) found increasing differences between cumulative exposure based on TEM and JEM with increasing number of tasks per job. In our study, the workers performed a large number of tasks (range: 2–12, median: 5) which could affect their recall performance. However, the similarities in agreement measures for the direct observation- and worker diary-based indices suggest that worker estimates of time-at-task were similar to direct observations.

Generally, a high degree of accuracy was observed for all the task-based indices suggesting very little average deviation from the 45° concordance line and a small bias. Most of the source of disagreement arose from a lack of precision suggesting higher variability in the difference between the individual measurement types. Similar results were reported by Reeb-Whitaker *et al.* (2004) who observed average bias of 1.5 dBA, accuracy of 97% and R^2 (precision) of 0.62 between dosimetry-based and worker diary-based full-shift noise exposures. In our study, task-based indices based on the AM generally yielded larger coefficients of accuracy than indices based on GM, suggesting that the AM is a better representation of task exposure than the GM. However, the GM-based indices yielded larger coefficients of precision than the AM-based indices, which was also replicated with the larger R^2 for the GM-based indices than AM-based indices. Hence a consistent pattern in the CCC was not observed because of this switch in the coefficients of precision and accuracy for the AM- and GM-based indices. When the interest is in obtaining a group mean estimate, the choice of a task-based index with the highest coefficient of accuracy may be of interest (AM), whereas if the interest is in estimating individual workers' exposures, an index that maximizes the precision coefficient may be of interest (GM).

A number of factors may contribute to the error associated with the task-based indices including errors in the estimates of the STTLs, worker and supervisor assessment of tasks performed and the time-at-task and the variation in task exposure levels between and within workers when task summary exposures are used. Reducing these sources of error would likely impact the coefficient of precision more than the coefficient of accuracy and would further improve the coefficient of agreement (CCC). This is evident in the present study which suggests that when mean task levels are highly variable, the resulting agreement measures (coefficient of precision and CCC) are low because the summary measure may not be representative of individual worker's task level on any given day. Improved estimates of the STTLs may be obtained through multiple regression models of the determinants of task exposures as has been reported by Seixas *et al.* (2003). One future direction could be to evaluate whether improvements in task exposure estimates relative to the methods explored to date could be achieved by predicting task exposures using these models.

The classification of measurements by job characteristics yielded significant improvements in the degree of agreement for the low job mobility, low job complexity and low job variability categories, but poorer agreement for the high categories. The coefficients of accuracy did not vary much among the categories for any of the job characteristics, but the precision coefficients improved for the low categories and worsened for the high categories. Generally, the indices based on AM had the highest coefficients of accuracy suggesting that very little bias in the estimate of mean exposure by the job category based on the task-based indices. The high degree of precision in the low categories of job characteristics for all the task-based indices suggests that the STTLs for these categories were measured with less error than for the high or medium categories. The estimates of task exposures were obtained for a very short duration of the task (30 s to 2 min) and may not be representative of high-mobility, high-variability and high-complexity tasks. In instances where the entire duration of a task can not be monitored, an important objective may be to understand exposure variability within tasks, which in turn can guide the choice of task sampling duration.

Measures of agreement

Carrasco and Jover (2003) have shown, using the variance components method, that the ICC and the CCC are one and the same, and both clearly have desirable agreement characteristics (absolute agreement), especially for comparing multiple alternative methods. The CCC as defined by Lin (1989) as the degree of departure from the 45° concordance line is conceptually more appealing, intuitive and easier to understand. Furthermore, CCC as computed by the method of moments can be decomposed into its components of precision and accuracy, which is very useful in understanding the sources of disagreement. Some limitations of the CCC as computed by the method of moments include the inability to: consider more than two methods at a time, adjust for confounding by covariates and account for repeated measurements. However, these limitations are accounted for when the CCC is calculated via the method of variance components which allows the incorporation of covariates that may affect the degree of agreement, as well as the comparison of multiple methods at once and accounting for repeated measurements (Carrasco and Jover, 2003). Carrasco and Jover (2003) also propose alternative measures of precision (residual variance) and accuracy (estimate of between method variance) that can be obtained from the variance components approach to describe the sources of disagreement.

General considerations for a task-based strategy

Although there is increased interest in the task-based exposure assessment strategy in recent years, little guidance is available on when and how to devise a task-based exposure assessment strategy. While our study indicated that the best agreement was achieved for jobs with low variability, complexity and mobility, the task-based strategy will likely not be beneficial under these circumstances. Task-based strategy is ideal for jobs associated with variable task exposure levels and time-at-task, such as jobs that are more mobile, more complex and have more variable full-shift exposures. In mobile, highly complex jobs for example in the construction industry, dosimeter measurements suggest high variability in full-shift exposures for some construction trades (Neitzel *et al.*, 1999; Kerr *et al.*, 2002; Seixas *et al.*, 2005) suggesting that characterizing these jobs with accuracy and precision would require a large number of samples. Thus, a task-based approach, although containing error, may still be more practical and perhaps cost effective than a large-scale dosimeter sampling.

Perhaps, the most important initial step is the definition and identification of tasks. In our study, considerable effort was devoted toward identifying and defining tasks during the pilot study. In defining tasks, there is a need to strike a balance between the specificity of the task and the practicality of recognizing and assessing the task exposure levels and task time. Exposures within task are also likely to vary, and hence it is also important to identify and collect information of possible factors that may impact task exposure levels.

A number of factors are important in the task-based strategy that can influence the degree of agreement between task-based estimates and full-shift measurements, including the precision and accuracy of task exposures and time-at-task, the duration of task preformed compared to the duration of task sampling and whether to sample for constant or variable duration. Although the longer a task is monitored, the more accurate the estimate will be, the use of grab samples is a cost effective approach to implementing a task-based strategy. Hence, the key is to determine the degree to which exposure variability within a task necessitates longer task monitoring. Clearly, highly mobile or variable tasks will likely require longer task-monitoring duration than the less mobile or variable tasks. Better understanding of the exposure variability within task is required to devise a monitoring plan which focuses on sampling tasks with high exposures, high exposure variability and/or conducted for a significant proportion of the time. Another aspect of the task-based strategy is the assessment of time-at-task. A practical strategy is worker assessment of time-at-task since the observer assessment approach is time and resource intensive. Several studies have noted high degree of agreement in the recording of tasks and the time-at-task by workers compared to the observation by researchers for fairly few tasks per job (Seixas *et al.*, 2003; Reeb-Whitaker *et al.*, 2004). However, a high degree of task specificity may require direct observation while less-specific tasks may be recorded with adequate accuracy by workers. The degree of agreement is likely dependent upon the specificity and the number of tasks and may be diminished with larger number of tasks per job. In the present study, the worker report-based indices of full-shift exposures agreed as well as the direct observation-based indices with the dosimetry data, even though a large number of tasks were identified. However, the direct comparison on the direct observation and worker reports of tasks and time-at-task has not been conducted yet, which may further shed light on the sources of disagreement with full-shift dosimetry measurements. Overall, conducting a well-planned task-based exposure assessment strategy requires significant resources. Thus, there is a need to balance the potential benefits of this strategy with the amount of resources required.

CONCLUSION

Overall, our results suggest that the task-based exposure assessment strategy can be used to obtain estimates of full-shift noise exposures that are in good agreement with the full-shift dosimetry measurements. The worker report-based indices yielded similar agreement to the direct observation-based indices, which achieved the best degree of agreement. Hence reliable estimates can be obtained using the less resource-intensive method of worker diary and task means. The task-based indices showed a high degree of accuracy (small bias) with most of the disagreement arising from a lack of precision (variability of the differences). Methods which improve the estimates of task exposure levels will likely improve the precision as well as the overall agreement. While the task-based methods can be used to estimate full-shift exposures which compare well with the full-shift dosimetry data, it is not clear how well these estimates will reflect long-term average exposures or mean group exposures. Therefore, careful consideration is warranted before embarking on a task-based exposure assessment strategy. Our results also demonstrate the utility of a number of methods to assess agreement, which together provide a comprehensive evaluation of the degree of agreement, the sources of disagreement and the boundaries within which most differences are contained and facilitates informed decision making among alternative measurement methods.

FUNDING

National Occupational Research Agenda. The study was funded in part by a NIOSH contract to the University of Massachusetts Lowell (212-2005-M-14358); NIOSH NORA project.

Acknowledgements—The authors would like to thank and acknowledge Dr Robert Bertrand and Ninon Bertrand (Bertrand Johnson Acoustique Inc.) for their contribution to field sampling and the identification of study facilities. We also thank Dr Michael Attfield from the Division of Respiratory Disease Studies at National Institute for Occupational Safety and Health, Morgantown, and Chucri Kardous from the Division of Applied Research Technology at National Institute for Occupational Safety and Health, Cincinnati, for their critical review of this manuscript. *Disclaimers*—The authors declare that they have no competing interests, financial or otherwise. Mention of a specific product or company does not constitute endorsement by the Centers for Disease Control and Prevention. The findings and conclusions in this report are those of the authors and do not necessarily represent the views of the National Institute for Occupational Safety and Health.

APPENDIX 1: SUMMARY OF SAS MACROS USED TO CALCULATE AND CROSS-CHECK THE CCC.

1. The macro is described in Crawford *et al.* (2007) and calculates the CCC and its confidence interval based on Lin (1989). It was downloaded from http://www.statisticaldisplays.org, on February 2008.
2. The macro is described in Lin *et al.* (2002) and calculates the CCC, TDI and CP and their one-sided 95% confidence limit. It was downloaded from http://www.uic.edu/∼hedayat/, on October 2007.

REFERENCES

Andersson IM, Rosén G. (1995) Detailed work analysis of control of exposure to airborne contaminants in the workplace. Appl Occup Environ Hyg; 10: 537–44.

Barnhart HX, Haber M, Song J. (2002) Overall concordance correlation coefficient for evaluating agreement among multiple observers. Biometrics; 58: 1020–7.

Beach JR, de Klerk NH, Fritschi L *et al.* (2001) Respiratory symptoms and lung function in bauxite miners. Int Arch Occup Environ Health; 74: 489–94.

Bedard M, Martin NJ, Kruger P *et al.* (2000) Assessing reproducibility of data obtained with instruments based on continuous measurements. Exp Aging Res; 26: 353–65.

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.

Bland JM, Altman DG. (1986) Statistical methods for assessing agreement between two methods of clinical measurements. Lancet; 1: 307–10.

Bland JM, Altman DG. (2003) Applying the right statistics: analyses of measurement studies. Ultrasound Obstet Gynecol; 22: 85–93.

Brueck SE, Stancescu D, Waters M. (2006) Industrial hygiene report for noise exposure monitoring surveys conducted at three manufacturing plants in Quebec, Canada (June 2003–January 2004). Cincinnati, OH: Centers for Disease Control and Prevention, National Institute for Occupational Safety and Health, Division of Surveillance, Hazard Evaluations and Field Studies, Industry Wide Studies Branch.

Carrasco JL, Jover L. (2003) Estimating the generalized concordance correlation coefficient through variance components. Biometrics; 59: 849–58.

Crawford SB, Kosinski AS, Lin HM *et al.* (2007) Computer programs for the concordance correlation coefficient. Comput Methods Programs Biomed; 88: 62–74.

Fritschi L, Sim MR, Forbes A *et al.* (2003) Respiratory symptoms and lung-function changes with exposure to five substances in aluminium smelters. Int Arch Occup Environ Health; 76: 103–10.

Goldberg M, Levin SM, Doucette JT *et al.* (1997) A task-based approach to assessing lead exposure among iron workers engaged in bridge rehabilitation. Am J Ind Med; 31: 310–8.

Haber M, Barnhart HX. (2006) Coefficients of agreement for fixed observers. Stat Methods Med Res; 15: 255–71.

Hager LD. (1998) Sound exposure profiling: a noise monitoring alternative. Am Ind Hyg Assoc J; 59: 414–8.

Hansen DJ, Whitehead LW. (1988) The influence of task and location on solvent exposures in a printing plant. Am Ind Hyg Assoc J; 49: 259–65.

Hornung RW. (1991) Statistical evaluation of exposure assessment strategies. Appl Occup Env Hyg; 6: 516–20.

Humann MJ, Donham KJ, Jones ML *et al.* (2005) Occupational noise exposure assessment in intensive swine farrowing systems: dosimetry, octave band, and specific task analysis. J Agromedicine; 10: 23–37.

Kalil AJ, Woskie SR, Holcroft C *et al.* (2004) Time variant exposure analysis (TVEA): a measurement tool for characterizing particulate exposure determinants in construction. J Occup Environ Hyg; 1: 816–25.

Kerr MJ, Brosseau L, Johnson CS. (2002) Noise levels of selected construction tasks. AIHAJ; 63: 334–9.

Kromhout H, Fransman W, Vermeulen R *et al*. (2004) Variability of task-based dermal exposure measurements from a variety of workplaces. Ann Occup Hyg; 48: 187–96.

Lin L. (2000) Total deviation index for measuring individual agreement with application in laboratory performance and bioequivalence. Stat Med; 19: 255–70.

Lin L, Hedayat AS, Sinha B *et al*. (2002) Statistical methods in assessing agreement: models, issues, and tools. J Am Stat Assoc; 97: 257–70.

Lin LI. (1989) A concordance correlation coefficient to evaluate reproducibility. Biometrics; 45: 255–68.

Linnet K. (1998) Performance of Deming regression analysis in case of misspecified analytical error ratio in method comparison studies. Clin Chem; 44: 1024–31.

Ludbrook J. (2002) Statistical techniques for comparing measurers and methods of measurement: a critical review. Clin Exp Pharmacol Physiol; 29: 527–36.

Mcgraw KO, Wong SP. (1996) Forming inferences about some intraclass correlation coefficients. Psychol Methods; 1: 30–46.

Meijster T, Tielemans E, Schinkel J *et al*. (2008) Evaluation of peak exposures in the dutch flour processing industry: implications for intervention strategies. Ann Occup Hyg; 52: 587–96.

Methner MM, McKernan JL, Dennison JL. (2000) Task-based exposure assessment of hazards associated with new residential construction. Appl Occup Environ Hyg; 15: 811–9.

Middendorf PJ, Lehocky AH, Williams PL. (1999) Evaluation and field calibration of the Miniram PDM-3 aerosol monitor for measuring respirable and total coal dust. Am Ind Hyg Assoc J; 60: 502–11.

Neitzel R, Yost M. (2002) Task-based assessment of occupational vibration and noise exposures in forestry workers. AIHA J; 63: 617–27.

Neitzel R, Seixas NS, Camp J *et al*. (1999) An assessment of occupational noise exposures in four construction trades. Am Ind Hyg Assoc J; 60: 807–17.

Nicas M, Spear RC. (1993a) A task-based statistical model of a worker's exposure distribution: Part II—application to sampling strategy. Am Ind Hyg Assoc J; 54: 221–7.

Nicas M, Spear RC. (1993b) A task-based statistical model of a worker's exposure distribution: Part I—description of the model. Am Ind Hyg Assoc J; 54: 211–20.

Nickerson CAE. (1997) A note on the concordance correlation coefficient to evaluate reproducibility. Biometrics; 53: 1503–7.

Nieuwenhuijsen MJ, Gordon S, Harris JM *et al*. (1995) Variation in rat urinary aeroallergen levels explained by differences in site, task and exposure group. Ann Occup Hyg; 39: 819–25.

Preller L, Heederik D, Kromhout H *et al*. (1995) Determinants of dust and endotoxin exposure of pig farmers: development of a control strategy using empirical modelling. Ann Occup Hyg; 39: 545–57.

Pronk A, Yu F, Vlaanderen J *et al*. (2006) Dermal, inhalation, and internal exposure to 1,6-HDI and its oligomers in car body repair shop workers and industrial spray painters. Occup Environ Med; 63: 624–31.

Reeb-Whitaker CK, Seixas NS, Sheppard L *et al*. (2004) Accuracy of task recall for epidemiological exposure assessment to construction noise. Occup Environ Med; 61: 135–42.

Ross AS, Teschke K, Brauer M *et al*. (2004) Determinants of exposure to metalworking fluid aerosol in small machine shops. Ann Occup Hyg; 48: 383–91.

Seixas NS, Robins TG, Moulton LH. (1988) The use of geometric and arithmetic mean exposures in occupational epidemiology. Am J Ind Med; 14: 465–77.

Seixas NS, Ren K, Neitzel R *et al*. (2001) Noise exposure among construction electricians. AIHAJ; 62: 615–21.

Seixas NS, Sheppard L, Neitzel R. (2003) Comparison of task-based estimates with full-shift measurements of noise exposure. AIHAJ; 64: 823–9.

Seixas N, Neitzel R, Sheppard L *et al*. (2005) Alternative metrics for noise exposure among construction workers. Ann Occup Hyg; 49: 493–502.

Smith RW, Sahl JD, Kelsh MA *et al*. (1997) Task based exposure assessment: analytical strategies for summarizing data by occupational groups. Am Ind Hyg Assoc J; 58: 402–12.

Smith TJ. (1992) Occupational exposure and dose over time: limitations of cumulative exposure. Am J Ind Med; 21: 35–51.

Smith TJ, Hammond SK, Hallock M *et al*. (1991) Exposure assessment for epidemiology: characteristics of exposure. Appl Occup Environ Hyg; 6: 441–7.

Susi P, Schneider S. (1995) A task-based exposure assessment model. Appl Occup Environ Hyg; 10: 395–9.

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.

Verma DK, Kurtz LA, Sahai D *et al*. (2003) Current chemical exposures among Ontario construction workers. Appl Occup Environ Hyg; 8: 1031–47.

Verma DK, Cheng WK, Shaw DS *et al*. (2004) A simultaneous job- and task-based exposure evaluation of petroleum tanker drivers to benzene and total hydrocarbons. J Occup Environ Hyg; 1: 725–37.

Vipperman JS, Bauer ER, Babich DR. (2007) Survey of noise in coal preparation plants. J Acoust Soc Am; 121: 197–205.

Virji MA, Woskie SR, Pepper L. (2009) Task-based lead exposures and worksite characteristics of bridge surface preparation and painting contractors. J Occup Environ Hyg; 6: 99–112.

White SA, van den Broek N. (2004) Methods for assessing reliability and validity for a measurement tool; a case study and critique using the WHO haemoglobin colour scale. Stat Med; 23: 1603–19.

Woskie SR, Bello D, Gore RJ *et al*. (2008) Comparison of task-based exposure metrics for an epidemiologic study of isocyanate inhalation exposures among auto body shop workers. J Occup Environ Hyg; 5: 588–98.