Validation of a Semi-quantitative Job Exposure Matrix at a Söderberg Aluminum Smelter

M. C. FRIESEN1,2*, P. A. DEMERS2, J. J. SPINELLI1 and N. D. LE1

1Cancer Control Research, British Columbia Cancer Agency, 201–601 West Broadway, Vancouver, BC V5Z 4C2, Canada; 2School of Occupational and Environmental Hygiene, University of British Columbia, Room 372, 2206 East Mall, Vancouver, BC V6T 1Z3, Canada

Received 16 October 2002; in final form 17 March 2003

Objectives: We tested the validity of a job exposure matrix (JEM) for coal tar pitch volatiles (CTPV) at a Söderberg aluminum smelter. The JEM had been developed by a committee of company hygienists and union representatives for an earlier study of cancer incidence and mortality. Our aim was to test the validity and reliability of the expert-based assignments.

Methods: Personal CTPV exposure measurements \( (n = 1879) \) overlapped 11 yr of the JEM. The arithmetic mean was calculated for 35 job/time period combinations (35% of the exposed work history), categorized using the original exposure intervals, and compared with the expert-based assignments.

Results: The expert-based and the measurement-based exposure assignments were only moderately correlated (Spearman’s \( \rho = 0.42 \); weighted \( \kappa = 0.39, CI 0.10–0.69 \)). Only 40% of the expert-based medium category assignments were correctly assigned, with better agreement in the low (84%) and high (100%) categories. Pot operation jobs exhibited better agreement (\( \rho = 0.60 \)) than the maintenance and pot shell repair jobs (\( \rho = 0.25 \)). The mid-point value of the medium category was overestimated by 0.3 mg/m\(^3\).

Conclusions: The expert-based exposure assignments may be improved by better characterizing the transitions between exposure categories, by accounting for exposure differences between pot lines and by re-examining the category mid-point values used in calculating the cumulative exposure. Lack of historical exposure measurements often requires reliance on expert knowledge to assess exposure levels. Validating the experts’ estimates against available exposure measurements may help to identify weaknesses in the exposure assessment where improvements may be possible, as was shown here.

Keywords: coal tar pitch volatiles; cohort studies; occupational exposure; polycyclic aromatic hydrocarbons; reproducibility of results; retrospective studies

INTRODUCTION

In the primary aluminum smelting industry, exposure to coal tar pitch volatiles (CTPV) has been associated with elevated risks of bladder cancer (Theriault et al., 1981, 1984; Armstrong et al., 1986; Spinelli et al., 1991; Ronneberg and Andersen, 1995; Ronneberg et al., 1999; Romundstad et al., 2000a,b) and lung cancer (Andersen et al., 1982; Gibbs, 1985; Armstrong BG et al., 1994; Ronneberg and Andersen, 1995). CTPV is a complex mixture, including known carcinogens such as polycyclic aromatic hydrocarbons (PAHs), that are generated during aluminum production (International Agency for Research on Cancer, 1984).

We are conducting a 15 yr update of a study of mortality and cancer incidence at a vertical stud Söderberg aluminum smelter, requiring the exposure assessment to be updated (Spinelli et al., 1991). The original exposure assessment was conducted by a committee of industrial hygienists and union health and safety representatives. However, several authors have pointed out significant limitations in using assignments made by expert raters, even when they are very familiar with plant operations and have access to sampling data (Kromhout et al., 1987; Hertzman et al., 1988; Hawkins and Evans, 1989;
METHODS AND PROCEDURES

Original exposure assessment and work history

A committee of two company industrial hygienists and two union health and safety representatives assigned CTPV exposures by consensus (Spinelli et al., 1991). They based their estimates on their knowledge of historical operational and technological changes and available CTPV measurements [measured as benzene soluble materials (BSM)], although it is not known to what extent the measurements were used. The period of plant operation, from start of production in 1954 to 1988, was divided into 13 intervals corresponding to the union contract periods. The committee assigned categorical exposure estimates for each distinct combination of job title and time interval to create a JEM. Four categories of exposure to CTPV were used: no exposure, low exposure (<0.2 mg/m³ BSM), medium exposure (0.2–1.0 mg/m³ BSM) and high exposure (>1.0 mg/m³ BSM). The mid-points of each exposure category (0, 0.1, 0.6 and 1.5 mg/m³ BSM) were used to calculate cumulative exposure. Jobs with the highest CTPV exposure primarily occurred before 1970 and included cell operators (pot men) and other jobs located within the pot lines. Jobs with no exposure to CTPV were administrators and workers in casting and power generation.

The proportion of the study work history assigned to each exposure category varied with the time period (Fig. 1). Two transition points, 1977 and 1983, mark large changes in the relative proportion of the exposure categories. These transitions coincide with substantial technological changes that were implemented in the pot rooms in the mid-1970s–1980s to reduce exposure levels (Table 1). By 1983 almost all of the exposed individuals (96%) were assigned to the low exposure category.

The original JEM contained just over 9400 unique job/department combinations, of which 3800 represented job/department combinations assigned to the low, medium or high exposure categories. Most of these combinations were not unique. After we eliminated variations in spellings and abbreviations that existed in the original JEM, there were 930 unique exposed job/department combinations. Of these, 100 job/department combinations accounted for 81% of the work history contributed by exposed individuals. Each unique job was assigned to an organizational work group representing their main activities: reduction operations (production of molten aluminum), pot lining and pot shell repair, maintenance, casting (molten aluminum to ingots), wharf/transportation and administration.

Prior to expanding the exposure assessment, we assessed the validity of the original exposure assessment using exposure measurements as the gold standard. We also compared the level of agreement between the two methods, which assumes neither a gold standard. Previous studies validating expert-based exposure estimates have used air or biological monitoring collected specifically for the validation study (Hertzman et al., 1988; Post et al., 1991; de Cock et al., 1996). Designing a specific sampling campaign ensures that exposure measurements are collected for all exposed jobs, and, where relevant, may permit comparisons of exposure ratings on a greater level of detail, such as on a task level. However, these studies are limited to comparisons for the current exposure conditions and may not capture how well the expert raters assessed exposure changes due to technological change. Designing a specific sampling campaign was not appropriate for this validation study because it was undertaken 15 yr after the job exposure matrix (JEM) was created.

To test the validity of the original exposure assessment, we used an 11 yr period of personal CTPV exposure measurements that overlapped the expert-based exposure assignments. These exposure measurements were available to the committee, but we do not know to what extent these measurements were used in their exposure assignments. During this period the company was implementing a major modernization initiative to reduce CTPV exposures. Thus, we have an opportunity to examine the ability of the committee to assess exposure reductions in a period of technological change. The aim of this study was to examine how the committee’s exposure assignments (expert-based) compare with exposure estimates based solely on exposure measurements (measurement-based). The results of this validation study will be used to determine what future work will be needed to improve or re-calibrate the original exposure assessment for CTPV prior to expanding it forward.

Teschke et al., 1989; Post et al., 1991; de Cock et al., 1996). In particular, they suggest that while committees of expert raters appear to be able to achieve reasonable agreement among themselves, their classifications may actually be poorly correlated with directly measured exposure levels, depending on the quantity of exposure data available at the time of classification and the characteristics of the contaminant. With 25 yr of personal CTPV exposure measurements collected at the study smelter since 1975, we have an opportunity to improve upon the original exposure assessment to better assess the exposure–response relationships between CTPV exposures at this smelter and cancer incidence and mortality.

To test the validity of the original exposure assessment, we used an 11 yr period of personal CTPV exposure measurements that overlapped the expert-based exposure assignments. These exposure measurements were available to the committee, but we do not know to what extent these measurements were used in their exposure assignments. During this period the company was implementing a major modernization initiative to reduce CTPV exposures. Thus, we have an opportunity to examine the ability of the committee to assess exposure reductions in a period of technological change. The aim of this study was to examine how the committee’s exposure assignments (expert-based) compare with exposure estimates based solely on exposure measurements (measurement-based). The results of this validation study will be used to determine what future work will be needed to improve or re-calibrate the original exposure assessment for CTPV prior to expanding it forward.
Exposure measurements

Between 1975 and 2001, 2624 personal BSM exposure measurements were collected by the company [n = 2295 (87%), hereafter referred to as company measurements] and by a regulatory agency [n = 329 (13%), hereafter referred to as compliance measurements]. All samples were collected using 37 mm sampling cassettes with fibreglass filters, which were desorbed with benzene. Company BSM exposure measurements collected pre-1982 were analyzed using a moving wire detector (Pye LCM-2) to carry the benzene extract to a flame ionization detector (Alcan Arvida Research Laboratory Method, undated). For company measurements after 1982 and all compliance measurements, the benzene was evaporated and the residue was weighed (Alcan Kitimat Laboratories Standard Method 2020, 1983; Workers’ Compensation Board of British Columbia Method 3350). No information was available regarding the comparability of these analytical methods. We tested the assumption of method equivalency by comparing exposure levels based on the source of the measurements during the analyses.

We excluded personal exposure measurements with a sampling duration of less than 6 h (n = 10, all compliance samples) so that measurements were representative of full-shift exposures. Samples less than the detection limit (method detection limit 0.1 mg/sample) were assigned an exposure 0.707 times...
the detection limit \( n = 51 \) compliance measurements) (Hornung and Reed, 1990). The CTPV exposure measurements encompassed 36 different jobs. Exposure measurements were coded to their respective time period in the study, using 1 July as the cut-off between time periods.

**Scope of validation study**

The validation included four time periods from 1977 to 1988, which excludes 4.5% of the data which was collected prior to 1 July 1977 and 23% of the data which was collected after 1 July 1988. Limited data was available for the time period 1975–1977, however, the majority of the measurements had been collected by the compliance agency to test its analytical method. Significant limitations had been found with their analytical method, so they were excluded from consideration in the validation study.

We used a strict approach for calculating arithmetic means for job/time period cells in the JEM because these arithmetic means were to be used as the ‘gold standard’ to examine the accuracy of expert-based exposure assessment. We set a minimum requirement of 10 measurements for each calculated arithmetic mean based on Mulhausen and Damiano (1998), who found that a plateau is reached in estimating the mean and standard deviation somewhere between six to ten measurements and that a ‘reasonable approximation of an exposure distribution often is possible with about 10 measurements’ for similarly exposed groups. We grouped adjacent time periods to meet the 10 measurement minimum if the time periods had been assigned the same expert-based exposure category. Job/grouped time period combinations that remained with less than 10 measurements were excluded from analysis in this validation study.

Arithmetic means were calculated for 35 job/grouped time period combinations (Table 2). The arithmetic mean exposures were categorized using the same measurement-based category assignments as the standard. The level of agreement between the two methods, which does not designate one method as the standard, was calculated using weighted \( \kappa \) (Armstrong BK et al., 1994). Analyses were also stratified by job type (pot operations jobs versus maintenance/pot shell repair jobs). The bias in using the category mid-points for calculating cumulative exposure was examined by calculating the mean of the differences between the mid-point and the arithmetic mean for each job/time period. The standard deviation of the differences was calculated as a measure of the precision.

The effect of source of exposure measurement (company versus compliance) and pot line was significant in linear regression while controlling for time period and job. As such, we examined whether the magnitude of these differences would result in different measurement-based category assignments by comparing arithmetic means calculated using both sources of data versus using only company data. There was insufficient data to use compliance measurements alone.

We tested the effect of exposure differences in different pot lines using one-way analysis of variance on natural log-transformed measurements. The pot lines were grouped as follows: lines 1 and 2, lines 3–5 and lines 7 and 8. Each pot line is a group of pots connected in series on the same electrical circuit. These line groupings were chosen to reflect similarities in technology and building ventilation properties. The smelter has used these same groupings administratively as working units and the work history differentiates between pot line groups for pot room operations jobs for most time periods.

All analyses were conducted using SPSS version 10.1 (SPSS Inc., Chicago, IL), except \( \kappa_W \), which was calculated using SAS version 8.0 (SAS Institute Inc., Cary, NC).

**RESULTS**

**Validity and reliability of expert-based assessment**

There were numerous discrepancies between the expert-based and measurement-based assignments (Table 2). The expert-based exposure category assignments were only moderately correlated with the measurement-based assignments (Spearman’s \( \rho = 0.42, P = 0.012 \)). There was moderate agreement between the two methods \( (\kappa_W = 0.39, 95\% \ CI 0.10–0.69) \). Sixty-six percent of the expert-based exposure assignments were correctly assigned based on the measurement-based assignments. The proportion varied by category, with good performance in the low category (84%, 16/19), but much poorer performance
in the medium category (40%, 6/15). The expert-based exposure assessment overestimated eight job/time period combinations and underestimated four. The committee was able to estimate exposures for pot operation jobs better than for maintenance/pot shell repair jobs (proportion correctly assigned, 71 versus 57%, respectively) (Table 3). The mid-points of the expert-based exposure categories were overestimated by 0.3 for the medium category (Table 4). Interpretation of the high exposure category is limited as it was based on only one job/time period.

### Influence of the exposure measurements source

Compliance measurements were available for 15 of the job/time period combinations. While measurement source was significant in linear regression, only one of the 15 job/time period combinations would result in a change in the measurement-based exposure.
category if only company measurements were used. For cell operator 1980–1983 the company-only arithmetic mean of 0.19 mg/m³ fell just below the cut-off point of 0.20 mg/m³ between the low and medium categories. Using only company data would result in a discrepancy in the two methods that reduces the level of agreement (Spearman’s ρ = 0.37, P = 0.03 and κW = 0.33, CI 0.02–0.65, proportion correctly assigned 63%). The maximum difference between using combined data versus company only data was 0.06 mg/m³.

Comparison of pot line exposures

Pot line exposure differences had not been considered as an initial grouping/splitting feature in the validation study because the expert-based assignments did not vary with pot line location. We tested this assumption for job/time period combinations with sufficient exposure measurements to differentiate pot line exposures using a minimum of five measurements. The 10 measurement criteria was relaxed for this sub-analysis, as the groups were expected to be more homogeneous and we would not be able to test this assumption otherwise. Exposure differences between pot lines were significant (P < 0.05) for eight of the 10 job/time period combinations assessed (Table 5), although only three would result in a difference in the measurement-based exposure category between pot line groups. Exposure differences between pot lines would be expected due to different implementation dates of technological

Table 3. Correlation, reliability and proportion correctly assigned comparing the expert-based assessment with the measurement-based assessment by type of job

<table>
<thead>
<tr>
<th>Job type</th>
<th>n</th>
<th>Correlationa</th>
<th>κW (CI)b</th>
<th>Proportion (%) correctly assigned in original exposure assessment</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Low</td>
</tr>
<tr>
<td>Pot line operations jobs</td>
<td>21</td>
<td>0.60P</td>
<td>0.52 (0.11–0.84)</td>
<td>80% (8/10)</td>
</tr>
<tr>
<td>Maintenance/pot shell repair</td>
<td>14</td>
<td>0.25</td>
<td>0.09 (–0.29–0.46)</td>
<td>89% (8/9)</td>
</tr>
</tbody>
</table>

nP < 0.05.

Table 4. Bias and precision in category mid-point values used in expert-based cumulative exposure assessment based on calculated job/time period arithmetic means

<table>
<thead>
<tr>
<th>Expert-based exposure category</th>
<th>Category mid-pointa</th>
<th>n</th>
<th>Biasb</th>
<th>Precisionc</th>
<th>Mediand</th>
</tr>
</thead>
<tbody>
<tr>
<td>Low</td>
<td>0.1</td>
<td>19</td>
<td>0.0</td>
<td>0.07</td>
<td>0.1</td>
</tr>
<tr>
<td>Medium</td>
<td>0.6</td>
<td>15</td>
<td>0.3</td>
<td>0.5</td>
<td>0.2</td>
</tr>
<tr>
<td>High</td>
<td>1.5</td>
<td>1</td>
<td>0.5</td>
<td></td>
<td>1.1</td>
</tr>
</tbody>
</table>

aMid-point of category used for cumulative exposure calculations.
bΣ(category mid-point – job/time period arithmetic mean)/number of job/time period combinations.
cStandard deviation of bias.
dMedian of the calculated job/time period arithmetic means for the specified expert-based exposure category.

Table 5. Differences in arithmetic mean (AM), geometric mean (GM) and number of measurements (n) for jobs by pot line group and time period using one-way analysis of variance (ANOVA) on natural log-transformed measurements

<table>
<thead>
<tr>
<th>Job</th>
<th>Time period</th>
<th>Lines 1 and 2</th>
<th>Lines 3–5</th>
<th>Lines 7 and 8</th>
<th>ANOVA P value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Exhaust maintenance</td>
<td>1980–1983a</td>
<td>0.26</td>
<td>0.23</td>
<td>11</td>
<td>0.10</td>
</tr>
<tr>
<td></td>
<td>1983–1985</td>
<td>0.17</td>
<td>0.14</td>
<td>12</td>
<td>0.05</td>
</tr>
<tr>
<td></td>
<td>1985–1988</td>
<td>0.11</td>
<td>0.10</td>
<td>5</td>
<td>0.00</td>
</tr>
<tr>
<td>Pot room repairman</td>
<td>1983–1988a</td>
<td>0.20</td>
<td>0.18</td>
<td>13</td>
<td>0.21</td>
</tr>
<tr>
<td>Cell operator</td>
<td>1977–1980</td>
<td>0.59</td>
<td>0.41</td>
<td>25</td>
<td>0.32</td>
</tr>
<tr>
<td></td>
<td>1980–1983</td>
<td>0.22</td>
<td>0.18</td>
<td>119</td>
<td>0.22</td>
</tr>
<tr>
<td></td>
<td>1983–1985</td>
<td>0.18</td>
<td>0.17</td>
<td>58</td>
<td>0.14</td>
</tr>
<tr>
<td></td>
<td>1985–1988</td>
<td>0.16</td>
<td>0.12</td>
<td>140</td>
<td>0.10</td>
</tr>
<tr>
<td>Anode operator</td>
<td>1980–1983a</td>
<td>0.17</td>
<td>0.16</td>
<td>9</td>
<td>0.55</td>
</tr>
<tr>
<td></td>
<td>1985–1988</td>
<td>0.07</td>
<td>0.07</td>
<td>7</td>
<td>0.08</td>
</tr>
</tbody>
</table>

aResults in change in measurement-derived categories by pot line group.
DISCUSSION

The findings of this study support the rigorous use of exposure measurements over expert-based exposure assessment whenever possible. Even though the committee of company and union representatives had access to exposure measurements, we found only moderate correlation between the expert-based and measurement-based exposure assessments.

The time period assessed was a key period of rapidly changing exposure levels in which CTPV exposures decreased by approximately 75%. As such, it is remarkable that we found that the level of agreement between methods was similar to that of Post et al. (1991), where the raters achieved 69 and 77% agreement, dependent on the chemical, when assessing current conditions with limited exposure measurements. Unlike Post and colleagues, however, we were limited to a comparison for only the highly exposed jobs as we were reliant on measurements already collected by the company and regulatory agency. The jobs not assessed were mainly in the low exposure category and were typically jobs that spent minimal time in the pot lines. It is unlikely they would be a significant source of additional discrepancy.

We used a strict approach in the calculation of the measurement-based assignments by setting a minimum number of samples and using only full-shift personal measurements in order to achieve the most accurate standard possible. Using a more lenient approach, i.e. using a minimum of three measurements and not grouping time periods, resulted in a lower correlation between methods but a similar proportion correctly assigned (Spearman’s $\rho = 0.37$, proportion correctly assigned 67%, data not shown). The decrease in correlation with a more lenient approach suggests that additional error may be introduced by using less stable mean estimates.

Even full-shift measurements may result in a less than perfect ‘gold standard’. The use of compliance agency data, non-random sampling conditions and other factors may influence the variance components and the mean exposure (Stewart et al., 1996). For instance, autocorrelation is possible, in particular in campaigns for regulatory purposes, in which sample collection is typically limited to within the same week and work shift and may underestimate the true variance if the variability in process (e.g. number of operating pots, composition of pitch mixture), weather conditions and other work conditions were not captured (Mulhausen and Damiano, 1998). In this study, the inclusion of compliance measurements only affected one job/time period categorical exposure assignment. In addition, some of the discrepancies between methods may be accounted for by exposure distributions that overlap exposure categories, such as from large within-job variability (Kromhout et al., 1987).

As expected, there were differences in exposure between pot line groups due to different implementation dates of technology and building improvements. However, these exposure differences may not be important for the mortality study because the exposure levels over historical periods were much higher than those observed here. In addition, prior to 1975 and after 1988 pot line conditions were much more stable.

Since the validation study was limited to the later years of the study, extrapolation of results to earlier time periods may be limited because the measures of agreement may be time dependent. In particular, the high exposure category results should be considered cautiously as only one job/time period was assigned to that category for this time period. Intuitively, one expects that the time periods included in the validation study would have better agreement with exposure measurements compared with earlier time periods, as was shown by Teschke et al. (1996). The committee members had personal familiarity with the plant for these time periods and exposure measurements were available. Any misclassification seen here has the potential to be more severe in earlier time periods where no measurements are available and the committee had to rely on written reports of technological changes and plant conditions and interviews with senior workers.

Choosing which value to use for each exposure category for cumulative exposure calculation in the absence of historical measurements, in particular, is difficult to address. The analyses presented here suggest that there is potential for overestimation of the cumulative exposures by using the category midpoints. However, the bias in the mean category value is time dependent and unfortunately cannot be calculated for earlier time periods. The influence of the category mean values will be assessed through a sensitivity analysis with the original cohort in future work.

The results of this validation study show that subjectively estimating exposures in a period of significant technological change is difficult. The discrepancies in exposure category assignments seen in these later years are likely to be unimportant for the original cancer study as the lag periods used in analyses would eliminate most of these years. However, exposures in this time period will be influential in calculating cumulative exposure in the follow-up study, especially for workers who began working in the 1970s and who have since accumulated over 25 yr of exposure. The magnitude of the impact of these discrepancies on the exposure–response relationship in the original and the updated studies will be explored in future work. The expan-
pression of the job exposure matrix for the study update provides an opportunity for these discrepancies to be addressed. Future steps to improve the original exposure assessment will focus on characterizing more accurate dates for transitions between exposure categories, accounting for exposure differences between pot lines, re-evaluating the choice in mean values for the exposure categories and dividing the low exposure category into more categories for future time periods, as 96% of the exposed individuals now fall into this category.

CONCLUSIONS

Although this validation study only covered a limited time period, it was a key period when exposure levels were rapidly changing. Even when exposure measurements were available, the expert-based exposure assessment method resulted in significant differences in exposure assignments compared with a measurement-based method. Lack of historical exposure measurements often requires relying on expert knowledge to assess exposure levels. As seen in this study, validating the experts’ estimates against available exposure measurements may help to identify weaknesses in the exposure assessment and thus recalibration of the assessment may be possible to improve its accuracy.

Acknowledgements—The authors thank the joint Alcan/Canadian Autoworkers Union (representing the aluminum smelter workers) advisory committee of Lyle D’Hondt, Dr Vern Davis, Josee Goupil, Marc-Andre Lavoie, Gord Lechner and Brent Morrison for their assistance and for their comments on an earlier version of this paper that led to improvements in the presentation of the material. We also thank Barry Boudreau and Jim Thorne at Alcan and Bob Stewart and the staff of the Research Secretariat at the Workers’ Compensation Board of British Columbia (WCB) for their help in gathering exposure measurements. We also thank Donna Kan, Barbara Jamieson, Janet Pitts, and Raymond Fang at the British Columbia Cancer Agency for data entry and programming support. This research was partially supported by grants from Alcan and the WCB.

REFERENCES
