



Published in final edited form as:

Occup Environ Med. 2012 October ; 69(10): 752–758. doi:10.1136/oemed-2011-100524.

Comparison of two expert-based assessments of diesel exhaust exposure in a case-control study: Programmable decision rules versus expert review of individual jobs

Anjoeka Pronk^{1,2}, Patricia A. Stewart^{1,3}, Joseph B. Coble^{1,4}, Hormuzd A. Katki⁵, David C. Wheeler^{1,6}, Joanne S. Colt¹, Dalsu Baris¹, Molly Schwenn⁷, Margaret R. Karagas⁸, Alison Johnson⁹, Richard Waddell⁸, Castine Verrill⁷, Sai Cherala¹⁰, Debra T. Silverman¹, and Melissa C. Friesen¹

¹Occupational and Environmental Epidemiology Branch, Division of Cancer Epidemiology and Genetics, National Cancer Institute, Bethesda, Maryland, USA ²TNO, Zeist, The Netherlands ³Now at: Stewart Exposure Assessments, LLC, Arlington, Virginia, USA ⁴Now at: 1412 Harmony Lane, Annapolis, MD USA ⁵Biostatistics Branch, Division of Cancer Epidemiology and Genetics, National Cancer Institute, National Institutes of Health, DHHS, Bethesda, Maryland, USA ⁶Now at: Department of Biostatistics, Virginia Commonwealth University, Richmond, VA, USA ⁷Maine Cancer Registry, Augusta, Maine, USA ⁸Dartmouth Medical School, Lebanon, New Hampshire, USA ⁹Vermont Cancer Registry, Burlington, Vermont, USA ¹⁰New Hampshire Department of Health and Human Services, Concord, New Hampshire, USA

Abstract

Objectives—Professional judgment is necessary to assess occupational exposure in population-based case-control studies; however, the assessments lack transparency and are time-consuming to perform. To improve transparency and efficiency, we systematically applied decision rules to the questionnaire responses to assess diesel exhaust exposure in the New England Bladder Cancer Study, a population-based case-control study.

Methods—2,631 participants reported 14,983 jobs; 2,749 jobs were administered questionnaires (‘modules’) with diesel-relevant questions. We applied decision rules to assign exposure metrics based solely on the occupational history responses (OH estimates) and based on the module responses (module estimates); we combined the separate OH and module estimates (OH/module estimates). Each job was also reviewed one at a time to assign exposure (one-by-one review estimates). We evaluated the agreement between the OH, OH/module, and one-by-one review estimates.

Results—The proportion of exposed jobs was 20–25% for all jobs, depending on approach, and 54–60% for jobs with diesel-relevant modules. The OH/module and one-by-one review had moderately high agreement for all jobs ($\kappa_w=0.68-0.81$) and for jobs with diesel-relevant modules

Corresponding Author: Melissa C Friesen, Occupational and Environmental Epidemiology Branch, Division of Cancer Epidemiology and Genetics, National Cancer Institute, 6120 Executive Blvd, Room 8106 MSC 7240, Bethesda MD 20892-7240. Tel: (301) 594-7485; Fax: (301) 402-1819; friesenmc@mail.nih.gov.

Competing interest: none declared

Licence statement: “I *Melissa Friesen* The Corresponding Author has the right on behalf of all Contributors to seek publication by the BMJ Group of all content within the submitted Contribution or as later submitted (which includes without limitation any diagrams, photographs, other illustrative material, video, film or any other material howsoever submitted by any of the Contributors at any time and related to this article) and to grant the warranties all as fully set out here:

(http://group.bmj.com/products/journals/instructions-for-authors/wholly_owned_licence.pdf)

($\kappa_w=0.62-0.78$) for the probability, intensity, and frequency metrics. For exposed subjects, the Spearman correlation statistic was 0.72 between the cumulative OH/module and one-by-one review estimates.

Conclusions—The agreement seen here may represent an upper level of agreement because the algorithm and one-by-one review estimates were not fully independent. This study shows that applying decision-based rules can reproduce a one-by-one review, increase transparency and efficiency, and provide a mechanism to replicate exposure decisions in other studies.

Introduction

Professional judgment is usually a necessary component of retrospective occupational exposure assessment in population-based case-control studies because the wide variety of jobs and employers involved makes collecting exposure information from the work sites impossible. Instead, exposure assessors rely on participant-reported information on job title, employer, and work tasks collected using an occupational history (OH) questionnaire. In some studies, occupation- and industry-specific questionnaires (modules) are incorporated to better capture within-job exposure variability among participants that cannot be determined from the job title alone [1].

Exposure assessors consider their knowledge of the process, job tasks, and exposure concentrations within those processes and tasks, and anchor the assessment to other familiar jobs or jobs with specific information in the literature [2]. To aid the decision process, exposure assessors have summarized the published measurement data for several agents, including diesel exhaust [3]. The application of professional judgment to the study participants' work histories to obtain an exposure estimate in population-based studies, however, has not been described and thus is criticized for occurring in a 'black box' [4–6]. The lack of explicit rules makes it difficult to determine the consistency of expert-based assessments over time or jobs. Furthermore, without explicit rules, it is difficult to assess if similar exposure decisions would be made by other exposure assessors or for other study populations. The lack of explicit rules can also be inefficient, because it provides no mechanism to assess that same agent in another study.

The exposure assessor's 'black box' does not need to be opaque [4]. The occupational questions asked in population-based studies are generally collected in a structured format [1], to which decision rules may be developed and applied. The idea of using a transparent decision process is not new and has been applied in industry-based studies [7–10]. For example, Cherrie et al. [7–8] described and validated a structured, deterministic method to assess past concentrations based on the work tasks and environment. This method required information on the intrinsic emission of the pollutant, the handling or processing method of the source, and the efficiency of local ventilation. Unfortunately, this information is not easily derived from the types of questions that can be asked of study participants in population-based studies. Decision rules based on questionnaire responses were also the basis of a semi-quantitative exposure assessment approach in a cohort of asphalt pavers [10] and subsequent nested case-control study [9]. More recently, Fritschi et al. [11] extended this approach to population-based studies by developing a web-based application to automate part of the expert-based assessment by assigning exposure decisions based on response patterns in the questionnaires.

In this paper, we describe an approach to apply programmable decision rules to assess occupational exposure to diesel exhaust in a population-based case-control study of bladder cancer. Decision rules for three metrics – probability, intensity, and frequency of occupational exposure – and the confidence in each metric's rating were developed and programmed based on the participants' responses to OH questions and, for the subset of jobs

with modules that included diesel-exhaust related questions, responses to modules. Our primary objective was to evaluate the extent that the estimates from the programmable decision rules (hereafter, algorithm estimates) differed from estimates obtained from an expert review of each job. Overall, our aim was to develop and evaluate an approach that might improve the efficiency, consistency, and transparency of the exposure assessment process in population-based case-control studies and provide a mechanism to replicate the decision rules for other studies.

Methods

Study population and data collection

The study population comprised 1213 patients newly diagnosed with carcinoma of the urinary bladder from 2001 through 2004 and 1418 population controls from Maine, New Hampshire, and Vermont. This population has been previously described [12]; we describe only the aspects related to collecting occupational information here. Trained interviewers administered a structured interview to all participants. As part of the interview, all participants completed an occupational history section where they reported all jobs held for six months or more from the age of 16 years. The OH comprised open-ended questions that included job title, name and location of employer, type of service or product provided, year started and stopped, work frequency, principal work tasks and duties, the tools and equipment used, and the chemicals and materials handled. For each job, two additional questions were asked: 'While at this job, did you ever work near diesel engines or other types of engines' and 'While at this job, did you ever smell diesel exhaust or other types of engine exhaust'. The OH responses triggered additional modules for 50 occupations and 17 industries of *a priori* interest for a variety of exposures, including diesel exhaust, if the relevant job was held for a minimum of two years. The questionnaires can be obtained from the corresponding author.

Diesel exhaust exposure assessment

Exposure metrics—The same exposure metrics and definitions were used, regardless of exposure assessment approach. Probability of exposure was assessed based on the estimated proportion of workers exposed to diesel exhaust for the identified exposure scenario (i.e., combinations of reported task, job or industry, and decade) and on secular changes in the prevalence of diesel engine use. Four probability categories were defined: 0 for 0–4%; 1 for 5–49%; 2 for 50–79%; and 3 for 80% of workers exposed.

Intensity of exposure was assessed on a continuous scale as the average level of respirable elemental carbon (REC, in $\mu\text{g}/\text{m}^3$) in the workers' breathing zone while performing tasks with likely diesel exhaust exposure. The assignments were based on reported personal REC measurements abstracted from an extensive literature review that were most closely related to the identified exposure scenario [3]. When no published measurements were available, an estimated REC level was extrapolated from other diesel exhaust components or was based on similarity to jobs or other exposure scenarios for which exposure measurements were available. The intensity levels were assumed constant over time because insufficient data were available to address secular trends in concentrations.

Frequency of exposure was assessed on a continuous scale as the average number of hours per week exposed to diesel exhaust. When available, frequency was based on the participants' self-reported frequency in hours per week for a likely exposed task or job. Otherwise, the job was assigned to one of four categories (0, 1–19%, 20–49%, 50% of work time) and then the midpoint of the relevant category was multiplied by the hours per week worked.

The ordinal confidence ratings (1–3) for each of the probability metrics were assigned based on whether the specific diesel sources were directly mentioned in the response (e.g., rating = 3), inferred from the questionnaire responses (e.g., rating = 2), or mentioned in the literature as being relevant to diesel exposure (e.g., rating = 1). For the confidence rating for intensity, higher confidence ratings were assigned when exposure measurements for that scenario were reported in the literature. For the confidence rating for frequency, higher confidence ratings were assigned when the frequency was based on the participants' self-reported frequency of diesel exhaust-related questions.

Exposure scenarios—A team of exposure assessors (AP, JBC, PAS) used the diesel literature review, the questions asked in the modules, and their professional judgment to identify scenarios associated with diesel exhaust exposure [3]. Exposure scenarios included jobs with traffic exposure (e.g., parking attendants, drivers) and jobs and locations where the subject worked with, or near, diesel-fueled equipment and vehicles (e.g., areas of shipyards, construction sites, farms). For each exposure scenario, a consensus rating was reached for each of the above-mentioned exposure metrics. See Appendix A for these scenarios and their assigned exposure metrics.

Extracting information from questionnaires—To relate the different patterns of questionnaire responses to the aforementioned exposure scenarios, we first undertook two data extraction tasks to code the questionnaire responses into a usable format. First, we coded the free-text OH responses to the questions on job title, industry, tasks and duties, tools and equipment, and chemicals and material into standardized variables. This was done by searching the free-text OH responses using string-based filters for common terms (e.g., truck, vehicle, generator), and their typical misspellings, that described diesel exhaust exposure scenarios and then coding these terms into standardized categorical variables. The free-text responses were manually scanned to identify potential search terms. Three groups of variables were extracted (see supplementary materials, Appendix A for definitions of the groups):

- *Diesel-powered equipment* (e.g., vehicles, generators) was extracted primarily from the job duties and equipment questions (Cat. A).
- *Traffic-exposed jobs* (e.g., parking attendant, driver) were extracted primarily from the job title and job duties questions (Cat. B).
- *Diesel exhaust-exposed job locations* (e.g., a construction site or a warehouse) were extracted primarily from the industry question and categorized into locations with *likely* (Cat. C) and *plausible* (Cat. D) bystander exposure to diesel exhaust.

Second, we manually reviewed the 6,196 questions asked in the modules and identified 553 questions in 29 of the 67 modules that directly or indirectly provided information on tasks, equipment, and locations that resulted in potential exposure to diesel exhaust (diesel-relevant modules). The remaining modules and questions captured occupational information related to other exposure agents, such as solvents. Examples of questions providing information on exposure scenarios include 'how much time did you spend doing city driving', and 'how often did you use a tool or equipment that was powered by diesel fuel'. See supplementary materials for more details (Appendix A).

Application of exposure scenarios to study subjects—One exposure assessor (AP) related each permutation of patterns in the extracted questionnaire responses to an exposure scenario and the scenario's assigned exposure metrics. The exposure metrics were assigned using a series of *if... then...* statements, where the 'if' portion of the statement provided the conditions reported in the questionnaire that related to an exposure scenario and the

'then' portion provided the team-based exposure decision. Two types of algorithm estimates were assigned. The first set of estimates was assigned to all jobs based solely on the exposure scenario derived from the OH responses in combination with the supplemental diesel-related OH questions (OH estimates). The second set of estimates was assigned only for those jobs with a diesel-relevant module and was based solely on the module responses (module estimates). After the decision rules were programmed, a second exposure assessor (PAS) reviewed a small subset of job records to determine whether important diesel exhaust-exposed jobs or tasks were missed during the development of the decision rules. The rules were modified where necessary.

For jobs with a diesel-relevant module, we then derived an OH/module estimate that considered both the OH and module estimates. We assigned the module estimate for the six metrics as the final estimates if a job's probability rating and the probability confidence rating were equal to or higher than the OH estimate probability and confidence ratings. Otherwise, an exposure assessor (AP) reviewed the entire participant's responses to assign the OH/module estimate. We gave greater weight to the module estimate because the modules contained more specific information on the subjects' work tasks.

One-by-one review estimate—Approximately one year after the team-based exposure decisions were made, an industrial hygienist (PAS) reviewed each participant's OH and module responses one at a time to assign exposure estimates for the same exposure metrics. During this review the industrial hygienist did not have access to the final algorithm-based estimates at the job-level, but did have access to the consensus ratings at the exposure scenario-level.

Statistical analyses

Statistical analyses were conducted using Stata S.E. 11.2 (StataCorp LP, College Station, Texas, 2009). We calculated descriptive statistics for each exposure metric for each of the three sets of estimates: the OH, the OH/module, and one-by-one review estimates. The intensity estimates were categorized as follows: 0 for no exposure; 1 for >0 – <5 ; 2 for 5 – <20 ; and 3 for $20 \mu\text{g m}^{-3}$ REC. The frequency estimates were categorized as follows: 0 for none; 1 for >0 – <8 ; 2 for 8 – <20 ; and 3 for 20 hours per week.

Two overall comparisons were made: 1) the OH estimates were compared with the OH/module estimates for jobs with a diesel-relevant module; and 2) the OH/module estimates were compared with the one-by-one review estimates for all jobs and for jobs with and without a diesel-relevant module. The categorical metrics were compared using the proportion of agreement, kappa (κ), and quadratically-weighted kappa (κ_w). The continuous metrics were compared using the Spearman correlation statistic (ρ). Patterns in the differences between exposure assessment approaches were identified by examining cross-tabulations of the categorical metrics and examining the direction of disagreement, the off-diagonals. The proportion of agreement by rating category was calculated for each metric (number of participants assigned category i in each exposure assessment approach divided by the number of participants assigned category i in either approach).

For each exposure assessment approach, we calculated a cumulative exposure estimate for each study participant by multiplying each job's REC intensity estimate by the frequency estimate and the number of years in that job; we then summed the products across all jobs held by the subject. Cumulative exposure was calculated twice for each exposure assessment approach: the first calculation treated any job with a probability rating greater than 0 as exposed; the second calculation changed the intensity and frequency estimate to 0 for jobs with a probability rating of 1, thus treating only jobs with a probability rating of 2 or 3 as

exposed. We then compared the cumulative exposure metrics from the three assessment approaches using the Spearman correlation statistic.

Results

Overview of algorithm assignment process

For 82% (n=12,358) of the 14,983 jobs, only the OH responses (including the two supplementary diesel questions) were used to assign algorithm estimates. Module responses were available for 64% (n=9,515) of the jobs; however, only 17% (n=2,603) of all jobs were associated with diesel-related modules. The jobs with these modules were assigned both an OH and an OH/module estimate. Only 41 jobs had insufficient information to assign any algorithm estimate. Of the 2,603 jobs with both an OH and module assignment, 82% were assigned the module estimate as the final OH/module estimate, because the module assignments' probability and confidence ratings were equal or higher than the OH algorithm rating. In the review of the remaining jobs, 17% were assigned the module estimate; 37% were assigned the OH estimate; and 46% were assigned revised estimates that considered both the OH and module responses together.

Descriptive Statistics

The overall proportion of jobs identified as exposed (probability rating > 0) to diesel exhaust ranged from 19.7 to 24.7%, depending on the approach (Table 1). The highest proportion of exposure was identified using the one-by-one review and the lowest prevalence was identified using the OH algorithm. When we restricted the exposure definition to jobs assigned a probability rating of 2 or 3, the three methods identified a similar overall proportion of exposure (16.4–17.7%). The proportion of exposed jobs increased to 54.7–60.3% in the subset with responses to diesel-related modules, with the highest proportion observed with the one-by-one review. The pattern changed in this subset when we restricted the comparison to those jobs assigned a probability rating of 2 or 3, with the OH/module estimates identifying the highest proportion of exposed jobs. In the subset of jobs without a diesel-relevant module, the proportion of exposed jobs ranged from 11.8–17.3% for any exposure and 9.2–9.8% for probability rating 2 or 3, depending on the assessment approach.

The means and standard deviations of each metric for jobs rated exposed were generally similar across the approaches, but there were some differences (Table 1). The mean intensity and frequency estimates were lower in the OH estimates than in the OH/module estimates. The means for all metrics were higher for jobs with diesel-relevant modules than for jobs without diesel-relevant modules, with the exception of the intensity confidence rating.

OH vs. OH/module estimates

In the subset of jobs with diesel-relevant modules, the agreement between the OH and OH/module estimates was moderately high (proportion of agreement: 71–84%; κ : 0.58–0.68; κ_w : 0.50–0.74) for all metrics except for the confidence metrics (κ_w : 0.35–0.56) (Table 2). The Spearman correlation between the OH and OH/module estimates evaluated on a continuous scale was 0.73 for the frequency metric and 0.53 for the intensity metric.

OH/module vs. one-by-one review estimates

Overall, the OH/module and one-by-one review estimates had a moderately high agreement for the categorical probability, frequency, and intensity metrics (proportion of agreement: 84–87%; κ : 0.56–0.66; κ_w : 0.68–0.81) (Table 3) and for the continuous intensity (Spearman $\rho=0.72$) and frequency ($\rho=0.70$) metrics. For the confidence metrics, the agreement was poor to moderate for all jobs ($\kappa_w = 0.39–0.44$) and poor for jobs with diesel-relevant modules (κ_w : 0.15–0.32). Overall, the agreement was similar for the jobs with diesel-

relevant modules ($\kappa_w = 0.62\text{--}0.78$; $\rho=0.61\text{--}0.66$) and the jobs without diesel-relevant modules ($\kappa_w = 0.60\text{--}0.73$; $\rho=0.64\text{--}0.72$).

The one-by-one review resulted in more ratings of 3 than the OH/module approach for probability (13.3 vs. 12.5%) and intensity (2.4 vs. 1.3%), but not for frequency (6.5% vs. 7.5%) (Table 4). The largest difference in the assigned scores between approaches was that the one-by-one review rated the probability rating for 1,100 jobs ($5.0+1.2+1.2=7.4\%$) as exposed that were rated unexposed by the OH/module approach. Most of these differences in the probability rating were rated 1 by the one-by-one review. The one-by-one review rated 455 jobs (3.0%) as unexposed that were exposed by the OH/module approach.

From the cross-tabulations in Table 4 we calculated the proportion of agreement by rating category. The unexposed category consistently had the highest agreement between approaches for all three metrics (87–88%). Among the exposed categories, the agreement was better in the highest exposure category than in the middle exposure categories for probability (14%, 19%, and 69%, for ratings 1, 2, and 3, respectively) and for frequency (31%, 20%, and 42%, for ratings 1, 2, and 3, respectively). For intensity, the agreement decreased across ratings, with agreements of 51%, 50%, and 35% for ratings 1, 2 and 3, respectively.

Comparisons of cumulative exposure

For all subjects, the Spearman correlation was 0.92 between the cumulative exposure estimates from the OH and OH/module approaches and 0.87 between the cumulative exposure estimates from the OH/module and one-by-one review approaches. For the exposed subjects, the correlation decreased to 0.66 for the OH vs. OH/module estimates and to 0.72 for the OH/module vs. one-by-one review estimates. Restricting the definition of exposed to those jobs with a probability rating of 2 or 3 resulted in little change in the relationship between the cumulative exposure measures (all subjects: OH vs. OH/module: $\rho=0.92$; OH/module vs. one-by-one review: $\rho=0.83$; exposed subjects: OH vs. OH/module: $\rho=0.65$; OH/module vs. one-by-one review: $\rho=0.72$).

Discussion

In this paper, we described an decision rule-based approach to assign diesel exhaust exposure in a population-based case-control study to increase the transparency and reproducibility of the decision rules. The moderately high agreement between the OH/module algorithm approach and a one-by-one review suggests that using programmable decision rules may provide comparable exposure estimates to a one-by-one expert review for diesel exhaust exposure.

OH vs. OH/module estimates

The moderate agreement between the assessments based solely on the OH responses (including the two additional diesel-related questions) and the assessments based on both the OH and module responses for jobs with diesel-relevant modules ($\kappa_w = 0.5\text{--}0.7$) indicates that, as suspected, the module information contributed information that influenced the exposure decision beyond what was available from the occupational history information. The jobs with diesel-relevant modules had a much higher prevalence of diesel exhaust exposure (55–60%) than jobs without diesel-relevant modules (12–17%), indicating that the diesel-relevant questions in the modules were probably targeting the most relevant and prevalent jobs.

OH/module vs. one-by-one review estimates

Probability, intensity, and frequency—The algorithm approach identified fewer jobs exposed to diesel exhaust (20%) than the one-by-one review (25%). Most of this difference was due to jobs rated as having a low probability of exposure in the one-by-one review compared with unexposed by the OH/module approach. Thus, we suspect that OH/module approach likely captured only the most common exposure scenarios reported in the questionnaires and used a strict definition of exposure; in contrast, the one-by-one review likely captured less frequently reported exposure scenarios and likely used a more inclusive definition of exposure. The OH/module approach's stricter definition of exposure is appropriate in a case-control study, because previous studies have shown that exposure definitions that define exposure status more strictly minimize attenuation in exposure-response associations when the exposure prevalence is low [13–14], albeit with potential loss of statistical power. For example, an exposure-response association between trichloroethylene exposure and non-Hodgkin lymphoma was only observed with a more strict exposure definition (subjects with medium or high probability of exposure) [15]. Capturing several levels of probability, however, allows us to evaluate the effects of varying our exposure definition in evaluations of the disease risks.

Overall, there was generally moderate to high agreement between the OH/module and one-by-one review estimates (overall and for jobs with or without diesel-relevant modules: $\kappa_w = 0.6$ – 0.8), indicating that the decision rule-based approach was able to capture most of the relevant exposure information from the occupational histories and modules, especially for the higher probability ratings. This agreement was generally similar or higher than the agreement between any two raters in previous studies (median $\kappa = 0.6$) [5 16–19], and much better than the poor agreement previously found between job exposure matrices and expert-based approaches ($\kappa = 0.4$ – 0.6) [20–27]. Our comparisons by exposure category showed very high agreement in the unexposed category for probability, intensity and frequency (87–88%) and variable agreement across the exposed categories (14–69%). This pattern has also been observed in previous comparisons between exposure assessment approaches [17 19–20 28–30].

There are several explanations for the potential differences between the decision rule-based and one-by-one review approaches. The differences likely represent the inability of the algorithm approach to identify less frequently-reported exposure scenarios, which was suggested by the greater prevalence of jobs with a probability rating of 1 in the one-by-one review. The differences may represent the ability of the one-by-one review to identify subtle contextual differences in a participant's exposure scenario that could be distinguished only when the entire response pattern was viewed. The differences may represent new information about diesel exhaust exposure that was identified after the decision rules were developed but considered in the one-by-one review. On the other hand, the differences may reflect inconsistencies in applying the decision rules in the one-by-one review. We plan to review these differences carefully in the future to add to and refine the algorithm-based decision rules where necessary.

Confidence metrics—We found poor agreement between approaches for the confidence metrics. The lack of agreement may reflect differences in how the confidence definitions were applied, despite using a common definition; or differences due to information found after the decision rules were developed. The confidence metric is potentially valuable in sensitivity analyses of exposure-response associations [31–32], although the probability metric has also been used for this purpose [15]. Our finding that different exposure assessment approaches results in such different confidence metrics suggests that we may need to consider other approaches to more consistently capture uncertainty in exposure

decisions. One such approach is to ask exposure assessors to assign the likelihood that the estimate falls within each exposure category, rather than provide a single score, thus providing an exposure distribution. This approach has been useful in retrospective exposure assessment efforts in industry-based studies [33] and in evaluating current exposure hazards in the workplace [34], but has not yet been used in case-control studies.

Limitations

This study has several limitations. First, we had no gold standard against which to evaluate the exposure assessment approaches so we cannot make any conclusions about any of the approaches' accuracy. We can only identify where the approaches were similar or different. Previous studies have shown that the accuracy of exposure assessors, compared to exposure measurements, can vary widely, from very good to very poor depending on the agent [5 19 35–40]. We plan to evaluate the robustness of these different expert-based approaches in future analyses on exposure-disease associations. Second, the exposure intensity levels and the category cut points should be interpreted cautiously and are meant to indicate relative, rather than absolute, differences in exposure between one job and another job. Third, the exposure assessor that conducted the one-by-one review had also participated in the decision rule-based assessment. To create some distance, the one-by-one review was conducted blind to the final exposure estimates, occurred over a year later, and was undertaken by an exposure assessor that had not extracted the free-text OH responses into standardized variables and had not developed the program that assigned the questionnaire response patterns to the exposure scenario decisions. Because the algorithm and one-by-one approach was not independent, the agreement levels reported here should be interpreted cautiously and may be an upper limit of agreement. On the other hand, the agreement between the two approaches may have been higher had new information been obtained between the development of the decision rules and the one-by-one review. The next step will be to compare the natural variability in expert ratings for diesel exhaust and compare multiple raters' ratings to the algorithm estimates to more fully understand the performance of the decision rule-based approach.

In addition, diesel exhaust may not be representative of other exposures and the differences between approaches may be greater for other agents. The exposure scenarios where exposure to diesel exhaust occurs are generally straightforward to identify and include exposure to traffic; driving, working with, or repairing heavy vehicles and equipment; and working in locations where heavy equipment are likely to be used. Diesel exhaust also has an offensive odor. Diesel fuel is also a commercial product likely known by many laymen and it cannot be interchanged with other types of fuel. Thus, the recognition, remembrance, and accuracy of reporting diesel fueled equipment may be greater than that for other hazards. In addition, this study included two supplemental exhaust-related questions in the OH that were asked of all participants for all jobs, which aided us in discriminating between unexposed and possibly exposed jobs (see supplementary Figure A.1). These screening questions may not be available in all studies or for all agents.

Conclusions

We successfully applied algorithms to assign transparent and reproducible decision rules to assess occupational exposure to diesel exhaust in a case-control study of bladder cancer. Our programmable, decision rule-based approach also provides a mechanism to apply the decision rules for diesel exhaust to other studies, increasing efficiency. Despite the good agreement seen here, we observed differences between approaches that warrant additional examination, which may further refine the decision rules and improve our ability to reproduce a one-by-one expert review. The impact of these differences on exposure-response associations will be examined in future analyses. Diesel exhaust may represent a

best case example, however, so the use of this approach needs to be examined and evaluated for a wide range of exposure agents.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

The research was funded by the Intramural Research Program of the National Institutes of Health, National Cancer Institute, Division of Cancer Epidemiology and Genetics.

References

1. Stewart PA, Stewart WF, Siemiatycki J, et al. Questionnaires for collecting detailed occupational information for community-based case control studies. *Am Ind Hyg Assoc J*. 1998; 59:39–44. [PubMed: 9438334]
2. Logan P, Ramachandran G, Mulhausen J, et al. Occupational exposure decisions: can limited data interpretation training help improve accuracy? *Ann Occup Hyg*. 2009; 53:311–24. [PubMed: 19332428]
3. Pronk A, Coble J, Stewart PA. Occupational exposure to diesel engine exhaust: a literature review. *J Expo Sci Environ Epidemiol*. 2009; 19:443–57. [PubMed: 19277070]
4. Kromhout H. Commentary. *Occupational and Environmental Medicine*. 2002; 59:594–594.
5. Teschke K, Olshan AF, Daniels JL, et al. Occupational exposure assessment in case-control studies: opportunities for improvement. *Occup Environ Med*. 2002; 59:575–93. discussion 594. [PubMed: 12205230]
6. Burstyn I. The ghost of methods past: exposure assessment versus job-exposure matrix studies. *Occup Environ Med*. 2011; 68:2–3. [PubMed: 21075766]
7. Cherrie JW, Schneider T. Validation of a new method for structured subjective assessment of past concentrations. *Annals of Occupational Hygiene*. 1999; 43:235–245.
8. Cherrie JW, Schneider T, Spankie S, et al. A new method for structured, subjective assessments of past concentrations. *Occupational Hygiene*. 1996; 3:75–83.
9. Agostini M, Ferro G, Olsson A, et al. Exposure assessment for a nested case-control study of lung cancer among European asphalt workers. *Ann Occup Hyg*. 2010; 54:813–23. [PubMed: 20861450]
10. Burstyn I, Boffetta P, Kauppinen T, et al. Estimating exposures in the asphalt industry for an international epidemiological cohort study of cancer risk. *Am J Ind Med*. 2003; 43:3–17. [PubMed: 12494417]
11. Fritschi L, Friesen MC, Glass D, et al. OccIDEAS: Retrospective Occupational Exposure Assessment in Community-Based Studies Made Easier. *J Environ Public Health*. 2009; 2009:957023. [PubMed: 20041014]
12. Colt JS, Karagas MR, Schwenn M, et al. Occupation and bladder cancer in a population-based case-control study in Northern New England. *Occup Environ Med*. 2011; 68:239–49. [PubMed: 20864470]
13. Dosemeci M, Stewart PA. Recommendations for reducing the effects of exposure misclassification on relative risk estimates. *Occup Hyg*. 1996; 3:169–176.
14. Kromhout H, Vermeulen R. Application of job-exposure matrices in studies of the general population: some clues to their performance. *European Respiratory Review*. 2001; 11:80.
15. Purdue MP, Bakke B, Stewart P, et al. A case-controlled study of occupational exposure to trichloroethylene and non-Hodgkin lymphoma. *Environ Health Perspect*. 2010
16. Correa A, Min YI, Stewart PA, et al. Inter-rater agreement of assessed prenatal maternal occupational exposures to lead. *Birth Defects Res A Clin Mol Teratol*. 2006; 76:811–24. [PubMed: 17044050]
17. Mannetje A, Fevotte J, Fletcher T, et al. Assessing exposure misclassification by expert assessment in multicenter occupational studies. *Epidemiology*. 2003; 14:585–92. [PubMed: 14501274]

18. Steinsvag K, Bratveit M, Moen BE, et al. Inter-rater agreement in the assessment of exposure to carcinogens in the offshore petroleum industry. *Occup Environ Med.* 2007; 64:582–8. [PubMed: 17227837]
19. Friesen MC, Coble JB, Katki HA, et al. Validity and Reliability of Exposure Assessors' Ratings of Exposure Intensity by Type of Occupational Questionnaire and Type of Rater. *Ann Occup Hyg.* 2011
20. Bhatti P, Stewart PA, Linet MS, et al. Comparison of occupational exposure assessment methods in a case-control study of lead, genetic susceptibility and risk of adult brain tumours. *Occup Environ Med.* 2011; 68:4–9. [PubMed: 20798009]
21. Benke G, Sim M, Fritschi L, et al. Comparison of occupational exposure using three different methods: hygiene panel, job exposure matrix (JEM), and self reports. *Appl Occup Environ Hyg.* 2001; 16:84–91. [PubMed: 11202032]
22. Luce D, Gerin M, Berrino F, et al. Sources of discrepancies between a job exposure matrix and a case by case expert assessment for occupational exposure to formaldehyde and wood-dust. *Int J Epidemiol.* 1993; 22 (Suppl 2):S113–20. [PubMed: 8132384]
23. Nam JM, Rice C, Gail MH. Comparison of asbestos exposure assessments by next-of-kin respondents, by an occupational hygienist, and by a job-exposure matrix from the National Occupational Hazard Survey. *Am J Ind Med.* 2005; 47:443–50. [PubMed: 15828074]
24. Rybicki BA, Johnson CC, Peterson EL, et al. Comparability of different methods of retrospective exposure assessment of metals in manufacturing industries. *Am J Ind Med.* 1997; 31:36–43. [PubMed: 8986252]
25. Stengel B, Pisani P, Limasset JC, et al. Retrospective evaluation of occupational exposure to organic solvents: questionnaire and job exposure matrix. *Int J Epidemiol.* 1993; 22 (Suppl 2):S72–82. [PubMed: 8132397]
26. Stucker I, Bouyer J, Mandereau L, et al. Retrospective evaluation of the exposure to polycyclic aromatic hydrocarbons: comparative assessments with a job exposure matrix and by experts in industrial hygiene. *Int J Epidemiol.* 1993; 22 (Suppl 2):S106–12. [PubMed: 8132383]
27. Peters S, Vermeulen R, Cassidy A, et al. Comparison of exposure assessment methods for occupational carcinogens in a multi-centre lung cancer case-control study. *Occup Environ Med.* 2011; 68:148–53. [PubMed: 20871102]
28. Fritschi L, Nadon L, Benke G, et al. Validation of expert assessment of occupational exposures. *American Journal of Industrial Medicine.* 2003; 43:519–522. [PubMed: 12704624]
29. Benke G, Sim M, Forbes A, et al. Retrospective assessment of occupational exposure to chemicals in community-based studies: validity and repeatability of industrial hygiene panel ratings. *Int J Epidemiol.* 1997; 26:635–42. [PubMed: 9222790]
30. Tielemans E, Heederik D, Burdorf A, et al. Assessment of occupational exposures in a general population: comparison of different methods. *Occup Environ Med.* 1999; 56:145–51. [PubMed: 10448321]
31. Gold LS, Stewart PA, Milliken K, et al. The relationship between multiple myeloma and occupational exposure to six chlorinated solvents. *Occup Environ Med.* 2010
32. Rajaraman P, Stewart PA, Samet JM, et al. Lead, genetic susceptibility, and risk of adult brain tumors. *Cancer Epidemiol Biomarkers Prev.* 2006; 15:2514–20. [PubMed: 17164378]
33. Ramachandran G, Banerjee S, Vincent JH. Expert judgment and occupational hygiene: application to aerosol speciation in the nickel primary production industry. *Ann Occup Hyg.* 2003; 47:461–75. [PubMed: 12890655]
34. Hewett P, Logan P, Mulhausen J, et al. Rating exposure control using Bayesian decision analysis. *J Occup Environ Hyg.* 2006; 3:568–81. [PubMed: 16998991]
35. de Cock J, Kromhout H, Heederik D, et al. Experts' subjective assessment of pesticide exposure in fruit growing. *Scand J Work Environ Health.* 1996; 22:425–32. [PubMed: 9000309]
36. Friesen MC, Demers PA, Spinelli JJ, et al. Validation of a semi-quantitative job exposure matrix at a Soderberg aluminum smelter. *Ann Occup Hyg.* 2003; 47:477–84. [PubMed: 12890656]
37. Hawkins NC, Evans JS. Subjective estimation of toluene exposures: A calibration study of industrial hygienists. *Applied Industrial Hygiene.* 1989; 4:61–68.

38. Hertzman C, Teschke K, Dimich-Ward H, et al. Validity and reliability of a method for retrospective evaluation of chlorophenolate exposure in the lumber industry. *Am J Ind Med.* 1988; 14:703–13. [PubMed: 3232688]
39. Kromhout H, Oostendorp Y, Heederik D, et al. Agreement between qualitative exposure estimates and quantitative exposure measurements. *Am J Ind Med.* 1987; 12:551–62. [PubMed: 3687951]
40. Post W, Kromhout H, Heederik D, et al. Semi-quantitative estimates of exposure to methylene chloride and styrene: The influence of quantitative exposure data. *Applied Occupational and Environmental Hygiene.* 1991; 6:197–204.

What this paper adds

- Expert-based exposure assessment in case-control studies is criticized for its lack of transparency, but the structured format of occupational history and job- and industry-specific modules makes it possible to develop and program transparent exposure decision rules.
- The application of programmable decision rules was greatly aided by extracting information from the study participants' free-text responses to open-ended questions into standardized variables, which likely improved the capture of within-job differences in work tasks.
- The programmable decision rules resulted in exposure estimates that had a moderately high agreement with exposure estimates obtained from a one-by-one expert review of job records that used similar decision rules. This agreement may be an upper level of agreement because the two approaches were not completely independent.
- The good agreement between the OH/module and one-by-one review approaches provides support for using transparent, programmable decision rules for diesel exhaust, and perhaps other exposures and other studies.

Table 1
Descriptive statistics of the exposure metrics by exposure assessment approach.

Assessment Approach	Jobs exposed to diesel exhaust			Exposure metrics for jobs with probability > 0												
	Jobs with any prob. of exposure ^a	%	N	Jobs with medium or high prob. of exposure ^b	%	N	Level (0-3)	Mean (SD)	Conf. (1-3)	Level ^c	Mean (SD)	Conf. (1-3)	Level ^d	Mean (SD)	Conf. (1-3)	Frequency
<i>All jobs (n=14,961)</i>																
OH algorithm ^e	2,944	19.7	2,461	2,461	16.5	2,461	2.4 (0.8)	2.1 (0.8)	2.1 (0.9)	5.6 (6.0)	2.1 (0.9)	2.1 (0.9)	16 (16)	1.3 (0.5)	1.3 (0.5)	1.3 (0.5)
OH/module algorithm	3,039	20.3	2,640	2,640	17.7	2,640	2.5 (0.7)	2.4 (0.8)	2.2 (0.9)	6.4 (11.8)	2.2 (0.9)	2.2 (0.9)	17 (16)	1.5 (0.7)	1.5 (0.7)	1.5 (0.7)
One-by-one review	3,684	24.7	2,452	2,452	16.4	2,452	2.2 (0.9)	2.0 (0.8)	2.5 (0.7)	7.4 (13.5)	2.5 (0.7)	2.5 (0.7)	14 (15)	1.6 (0.6)	1.6 (0.6)	1.6 (0.6)
<i>Jobs with diesel-relevant module (n=2,603)</i>																
OH algorithm ^e	1,489	57.2	1,247	1,247	47.9	1,247	2.4 (0.8)	2.3 (0.8)	2.2 (0.8)	6.1 (6.5)	2.2 (0.8)	2.2 (0.8)	19 (17)	1.4 (0.5)	1.4 (0.5)	1.4 (0.5)
OH/module algorithm	1,424	54.7	1,435	1,435	55.1	1,435	2.8 (0.6)	2.8 (0.5)	2.3 (0.8)	7.9 (16.1)	2.3 (0.8)	2.3 (0.8)	17 (18)	1.8 (0.8)	1.8 (0.8)	1.8 (0.8)
One-by-one review	1,571	60.3	1,313	1,313	50.4	1,313	2.7 (0.6)	2.7 (0.6)	2.2 (0.8)	7.7 (15.4)	2.2 (0.8)	2.2 (0.8)	21 (14)	1.8 (0.8)	1.8 (0.8)	1.8 (0.8)
<i>Jobs without diesel-relevant module (n=12,358)</i>																
OH algorithm ^e	1,455	11.8	1,192	1,192	9.7	1,192	2.3 (0.7)	2.0 (0.9)	2.1 (0.9)	5.1 (5.4)	2.1 (0.9)	2.1 (0.9)	13 (14)	2.1 (0.9)	2.1 (0.9)	2.1 (0.9)
OH/module algorithm	1,468	11.9	1,205	1,205	9.8	1,205	2.3 (0.7)	2.0 (0.9)	2.1 (0.9)	5.1 (5.4)	2.1 (0.9)	2.1 (0.9)	13 (14)	2.1 (0.9)	2.1 (0.9)	2.1 (0.9)
One-by-one review	2,136	17.3	1,139	1,139	9.2	1,139	1.9 (0.9)	1.8 (0.7)	1.7 (0.7)	6.8 (12)	1.7 (0.7)	1.7 (0.7)	12 (14)	1.5 (0.6)	1.5 (0.6)	1.5 (0.6)

% proportion of jobs; N, number of jobs; prob., probability.

^aJobs assigned a probability of exposure rating of 1, 2 or 3.

^bJobs assigned a probability of exposure rating of 2 or 3.

^cin $\mu\text{g m}^{-3}$ REC

^din hours per week

^eThe OH algorithm approach was not able to assign exposure estimates for 22 jobs.

Table 2

Agreement between the algorithm estimates (categorical metrics) based solely on the occupational history information (OH estimates) and those based on both the occupational history and module information (OH/module estimates) for jobs with a diesel-relevant module.

Jobs with diesel-relevant module (n=2,603)			
	% agree.	κ	κ_w
Probability	73	0.59	0.69
Frequency	71	0.58	0.74
Intensity	84	0.68	0.50
Probability Conf.	63	0.31	0.35
Intensity Conf.	67	0.44	0.56
Frequency Conf.	61	0.43	0.37

agree., agreement; conf., confidence rating; κ , kappa; κ_w , weighted kappa using quadratic weights.

Agreement between the algorithm estimates (categorical metrics) based on both the occupational history and the estimates from a one-by-one expert review.

Table 3

	All jobs (n=14,961)		Jobs with diesel-relevant module (n=2,603)		Jobs without diesel-relevant module (n=12,358)	
	% agree.	κ	% agree.	κ_w	% agree.	κ_w
Probability	86	0.62	0.81	0.78	87	0.52
Frequency	84	0.56	0.73	0.62	88	0.55
Intensity	87	0.66	0.68	0.66	89	0.58
Probability Conf.	74	0.39	0.44	0.15	78	0.45
Intensity Conf.	74	0.42	0.54	0.32	78	0.46
Frequency Conf.	72	0.44	0.61	0.28	78	0.51

agree., agreement; conf., confidence; κ , kappa; κ_w , weighted kappa.

Table 4

Comparison of the OH/module estimates to the OH estimates and the one-by-one review estimates for the categorical exposure metrics.

Metric	OH/Module Estimate	Proportion of Jobs (%) (n=14961)				Total
		0	1	2	3	
Probability	0	72.3	5.0	1.2	1.2	79.7
	1	0.7	1.3	0.3	0.4	2.7
	2	1.2	1.5	1.3	1.2	5.2
	3	1.1	0.5	0.4	10.5	12.5
	Total	75.4	8.2	3.1	13.3	100.0
Intensity	0	72.5	5.9	0.3	0.9	79.7
	1	2.5	11.2	1.2	0.4	15.3
	2	0.3	0.7	2.6	0.1	3.7
	3	0.3	0.0	0.1	0.9	1.3
	Total	75.6	17.9	4.2	2.4	100.0
Frequency	0	73.8	4.5	0.6	0.8	79.7
	1	2.8	4.9	0.8	0.9	9.4
	2	0.7	0.7	1.2	0.7	3.4
	3	1.0	1.0	1.3	4.2	7.5
	Total	78.4	11.1	4.0	6.5	100.0