Ann. Occup. Hyg., Vol. 55, No. 3, pp. 339–342, 2011

Letter to the Editor

Comments on the Diesel Exhaust in Miners Study

Received 15 September 2010; in final form 16 November 2010

We are writing in response to four recent articles describing the historical reconstruction of diesel exhaust exposures in underground mines (Diesel Exhaust in Miners Study or DEMS) for use in epidemiological analyses of exposure–response relationships between mining-associated diesel exposure and health effects, e.g. lung cancer (Coble *et al.*, 2010; Stewart *et al.*, 2010; Vermeulen *et al.*, 2010a,b). We have particular concerns about the estimation of historical respirable elemental carbon (REC) levels for underground miners. Those estimates were back extrapolated using historical carbon monoxide (CO) area measurements and estimated levels of historical 'adjusted horsepower (HP)' of diesel fleets to estimate levels of CO as surrogates of diesel exhaust (DE). As discussed below, those measures are at best imprecise and may not be valid measures of DE. Accordingly, we have concerns about the validity of the articles' conclusions, which we believe are insufficiently justified.

The following communication briefly describes those concerns.

1. Measurements of CO by colorimetric tubes are imprecise and may be unreliable. The historical reconstruction relied on CO data obtained in the DEMS survey, a 1994 Feasibility study, and a compilation of MSHA data from 1976 to 2001 (MIDAS). In the DEMS and Feasibility studies, CO measurements were obtained using long-term (i.e. 8 h) colorimetric tubes. The MIDAS surveys used short-term colorimetric tubes (i.e. 5–15 min) or gas chromatography, but the numbers determined by each of those methods were not described.

The precision of CO colorimetric tubes is limited, especially at low exposure levels. Prior to the NIOSH certification program, colorimetric and length-of-stain detector tubes were considered 'inaccurate' (Perkins, 1997). Early studies reported that over the range from 25 to 100 p.p.m., all commercially available CO detector tubes were worse than ±25% of the true value, while only some brands yielded results within ±50% of the true value (Morgenstern *et al.*, 1970). CO tubes are currently required to yield results ±35% the true value at 12.5 p.p.m., but precision and accuracy decline as concentrations approach the limit of their recommend use ($\approx$5 p.p.m.).

Because of imprecision, inaccuracy, and observer variability, historical and current authorities agree that CO colorimetric tubes should only be used to detect the presence of CO and for range finding purposes, not for quantitative measurements (WHO, 1976; Stern and Mansdorf, 1999; Todd, 2003) and that after CO detection 'a more accurate . . . method' should be used (Lodge, 1988). In 1976, the year MIDAS surveys began, WHO recommended that detector tubes only be used 'for estimating the concentration of CO at concentrations above 5 mg/m^3, (i.e. 4.35 p.p.m.) (WHO, 1976). Accordingly, it is notable that the great majority of CO measurements in the DEMS reports were below 4.35 p.p.m..

We are also concerned by a table footnote indicating that some CO values were 'corrected for measurement technique (detector tube versus bistable)', but the reports described neither the method for and effects of such 'correction', nor the number of samples so 'corrected'. Thus, the CO measurements are subject to still further uncertainty.

2. The majority of CO measurements were below the recommended range for use of colorimetric tubes. The reports do not describe the actual CO measurements across the seven mines, but summary statistics were provided for samples obtained at the underground production face. Across the seven mines in the DEMS survey, the geometric means of CO samples ranged from 0.8 to 4.5 p.p.m.. Among historical production face CO measurements, the geometric means across all mines grouped by decade 'typically' ranged from $\leq$1 to 3 p.p.m..

1219-AB86-COMM-20-2 Borak

The median analytical limit of detection (LOD) for CO was reported to be 0.3 p.p.m. for the DEMS survey; 15% of the DEMS survey production face samples were <LOD. The analytical LOD for historical CO samples was reported to be 1.0 p.p.m.. Among all of the CO production face samples used in facility-specific models (i.e. DEMS survey plus historical surveys), 20–60% of samples were <LOD.

To gain perspective on the actual distribution of CO samples, one of us (T.A.H.) independently obtained the 1998 CO results from Mine A in the DEMS survey. Of 26 measurements, only 1 was in the 25–100 p.p.m. range, for which detector tubes are required to be within ±25%, while 9 of 26 (35%) were <2 p.p.m.. Thus, nearly all of the measurements were in the range for which the detector tube precision is expected to be worse than ±35%. More notably, Mine A had the highest levels of CO of the seven mines. It seems reasonable to assume that each of the other mines had even greater proportions of CO samples in the very low concentration range for which detector tubes are least precise and not recommended.

3. Fleet HP is a poor predictor of CO and REC emissions. The exposure reconstruction relied upon estimates of the 'adjusted HP' of each mine's historical diesel fleet to estimate CO emissions, which were then extrapolated to estimate REC levels. However, fleet inventories were available for only 'a few years'; for the missing years, the inventories themselves were estimated. HP ratings were known for 80% of the equipment; for others, HP was also estimated. Total annual HP was summed for each mine and then 'adjusted' by further estimating the percentage of work shifts that the equipment was used.

Our particular concern is that there is 'no universal relation between CO and particulate matter (PM)' across an engine fleet; to the contrary, evidence suggests that the CO/PM relationship is 'unique for each engine type and perhaps for each engine' (Clark *et al.*, 1999b). Studies of diesel equipment in underground mines revealed no consistent relationship between engine power and either CO or EC (Davies, 2000, 2002). Individual engine types had a wide range of EC and CO emissions, and a given engine's emissions varied widely based on operating conditions. For example, the same engine emitted nearly four times as much EC and nearly three times as much CO before maintenance as compared to after maintenance (Davies, 2000).

Wide variability of emissions has been shown among in-use heavy duty vehicles built under identical regulatory standards, including some with identical engines (Yanowitz *et al.*, 2000). For example, EPA Certification Data for large heavy-duty off-road diesel engines tested in 2003 indicate that the ratio of CO/PM emission rates (reported as g/bHP-hr or g/kW-hr) varied over a range more than 100-fold, from <0.1 to >19 (US EPA, 2010). Engine emission rates are also driver- and route-related. Tests of a diesel bus over five dynamometer cycles found the CO/PM ratio varied from 12.88 to 38.39 (Clark *et al.*, 1999b). In other tests, aggressive 'pedal behavior' increased CO emissions nearly 3.5-fold and PM emissions 2.4-fold compared to non-aggressive behavior (Clark *et al.*, 1999a).

Thus, even if the numerous required estimations proved factually correct, use of fleet HP to estimate DE emissions is probably not justifiable because HP is not a good predictor of individual engine emissions.

4. CO and REC levels are not strongly correlated. Despite the inherent imprecision and inaccuracy of CO measurements, historical reconstruction of DE exposures relied on the CO–REC relationship. Unfortunately, that relationship was not strong. To the contrary, linear regression, performed over 168 pairs of log-transformed underground production face samples, indicated a mean Pearson correlation coefficient for REC on CO of only 0.41, the 'weakest' correlation with any of the gaseous emissions in the study (Vermeulen *et al.*, 2010b, p. 769). Moreover, the correlation was highly variable; correlation coefficients ranged from 0.05 to 0.77 across the seven mines. The DEMS authors described that correlation as 'only moderate'. Likewise, in a review of dynamometer testing of heavy-duty diesel vehicles, Yanowitz et al. (2000) described the relationship between PM and CO ($r^2 = 0.45$) as only 'somewhat correlated'.

The DEMS authors specifically noted 'considerable heterogeneity' in the association between CO and REC, but it was essentially ignored: 'The observed heterogeneity . . . was likely due to facility- and measurement location-specific circumstances . . . . However, this level of detail was not available on an individual measurement level and we therefore were not able to explore this issue further.' Thus, the historical reconstruction was based on a relationship that was 'only moderate' and further burdened by uncharacterized heterogeneity.

5. The historical exposure reconstruction has substantial uncertainty. As noted above, the key elements of the historical exposure reconstruction are subject to significant imprecision and uncertainty. Therefore, the results of the reconstruction (estimated historical REC exposure levels) derived by multiplicative interactions of those measurements greatly compound the uncertainty. Unfortunately, the four reports do not describe the magnitude of that uncertainty, and the reported data are insufficient

to independently determine the total effect of the compounded uncertainties.

Some insight is gained by examining one element of the complex reconstruction—comparing model-predicted levels of CO for 1976–1977 versus average levels observed at the production face in an independent (MESA) survey at six of the mines (table 3 of Vermeulen *et al.*, 2010a). The relative differences of the estimated arithmetic means compared to the arithmetic means of the measured levels (i.e. [(measured − estimated)\measured] × 100%) ranged from −25 to +49%. If the authors had taken into account the confidence intervals (CIs) around the 'measured' means and/or the estimated arithmetic means, a substantially larger range of the already large relative differences would have been observed.

Thus, rather than supporting the exposure reconstruction, these relative differences cast serious doubts on the validity and applicability of that process. Moreover, comparison of the mine-specific arithmetic mean CO levels from the MESA survey (table 3 of Vermeulen *et al.*, 2010a) versus the corresponding means computed from the geometric means and standard deviations (table 1 of Vermeulen *et al.*, 2010a) raises doubts about the assumption of log normality used in the reconstruction.

Additional insight can be gained by considering the likely range of CIs around the model-estimated values of historical REC. Those estimates were derived by the following formula (see Vermeulen *et al.*, 2010a, p. 779):

$$\mathrm{REC}_{ik} = \mathrm{REC}_{kR} \times \mathrm{RELtrend}_i$$

where REC_{ik} is the REC exposure estimate for year i and job k; REC_{kR} is the reference REC exposure for job k; and $\mathrm{RELtrend}_i$ is the CO exposure estimate for year i relative to the reference concentration. To determine the range of CIs, it is necessary to determine the variance of the REC_{ik} estimates, a process that requires values for at least the following terms: $\mathrm{Var}(\mathrm{REC}_{kR})$; $\mathrm{Var}(\mathrm{RELtrend}_i)$; Mean of REC_{kR}; Mean of $\mathrm{RELtrend}_i$; and $\mathrm{Covar}(\mathrm{REC}_{kR}, \mathrm{RELtrend}_i)$. Unfortunately, only two of those values, $\mathrm{Var}(\mathrm{REC}_{kR})$ and Mean of REC_{kR}, are found in the reports.

In order to calculate the likely CI, we assigned a value to $\mathrm{RELtrend}_i$ of 3.0 (i.e. 300%), equal to the 'median of the maximum relative increases among the operations' Vermeulen *et al.,* 2010a, p. 781; the range of reported maximum relative increases was 100–685%. And, for simplicity, we assumed that $\mathrm{Var}(\mathrm{RELtrend}_i)$ and $\mathrm{Covar}(\mathrm{REC}_{kR}, \mathrm{RELtrend}_i)$ were both zero. We then estimated the 95% CI for each of the 49 mine-specific job titles described in table 4 of Vermeulen *et al.* (2010b). For 30 of 49 jobs (61%), the 95% CI (i.e. the reported arithmetic mean for the job ±2 SD, calculated as the square root of the variance) included zero. In other words, 61% of the calculated values of REC_{ik} were not statistically different from zero. [We note that the values for $\mathrm{Var}(\mathrm{RELtrend}_i)$ and $\mathrm{Covar}(\mathrm{REC}_{kR}, \mathrm{RELtrend}_i)$ are very likely non-zero, and consequently the estimated proportion of calculated values not statistically different from zero would be even larger.]

Our analysis of these two elements of a more complex exposure reconstruction indicates substantial uncertainty about the estimated historical REC exposure values. We suspect that the estimated historical REC values have such broad CIs that point estimates for most job- and year-specific exposure levels cannot be statistically distinguished. Unfortunately, the DEMS reports do not provide the data needed to fully evaluate that possibility.

If we are correct, then the estimated REC values would also not provide adequate quantitative information for calculating dose–response relationships. However, because response data (e.g. lung cancer rates) have not yet been reported, it is not yet possible to determine the impact of exposure assessment uncertainties on estimated dose–response relationships.

6. In summary, the methods used for historical exposure reconstruction are subject to substantial, but uncharacterized uncertainty and the data reported in these studies are insufficient to evaluate that uncertainty. Our initial efforts to understand the reported data suggest that the historical reconstruction yields exposure estimates that lack precision, are likely to be associated with significant misclassification, and may be little better than qualitative estimates. It is likely that the imprecision of CO colorimetric tubes and problems in using HP to estimate CO (and DE) emissions explain why CO has not been previously used as a DE surrogate in epidemiological studies (e.g. MSHA, 2001; US EPA, 2002; Hesterberg *et al.*, 2006; Gamble, 2010).

We are concerned that future reliance on such data to estimate dose–response relationships will lead to significant misclassification and invalid conclusions. We are also concerned that such misclassification will be excused as 'non-differential' and therefore likely to underestimate risk; that view is not justified (e.g. Dosemici *et al.*, 1990; Sorahan and Gilthorpe, 1994; Jurek *et al.*, 2005). In contexts such as these, the effects of misclassification cannot be predicted and should not be ignored.

JONATHAN BORAK
Departments of Epidemiology and Medicine,
Yale University,
234 Church Street, New Haven, CT 06510, USA

WILLIAM B. BUNN
Department of Public Health, Northwestern
School of Medicine, and Navistar, Inc., 303 E.
Wacker Dr., Chicago, IL 60601, USA

GERALD R. CHASE
Consultant in Statistics & Epidemiology, 6058
South Pike Dr., Larkspur, CO, USA

THOMAS A. HALL
Consultant in Industrial Hygiene, 2400 Big Cedar
Ave, Edmond, OK, USA

H. John HEAD
Behre Dolbear, 999 Eighteenth St., Denver, CO
Denver, CO, USA

THOMAS W. HESTERBERG
Medical, Health Management and Safety, Navistar,
Inc., 303 E. Wacker Dr. Chicago, IL, USA

GREG SIRIANNI
Department of Fire Science, University of New
Haven, 300 Orange Ave., West Haven, CT, USA

THOMAS J. SLAVIN
Medical, Health Management and Safety, Navistar,
Inc., 303 E. Wacker Dr. Chicago, IL, USA

REFERENCES

Clark NN, Gautam M, Rapp BL *et al.* (1999a) Diesel and CNG transit bus emissions characterization by two chassis dynamometer laboratories: results and issues (report #1999-01-1469). Dearborn, MI: The Engineering Society for Advancing Mobility Land Sea Air and Space (SAE).

Clark NN, Jarrett RP, Atkinson CM. (1999b) Field measurements of particulate matter emissions, carbon monoxide, and exhaust opacity from heavy-duty diesel vehicles. J Air Waste Manage Assoc; 49: 76–84.

Coble JB, Stewart PA, Vermeulen R *et al.* (2010) The diesel exhaust in miners study: II. Exposure monitoring surveys and development of exposure groups. Ann Occup Hyg; 54: 747–61.

Davies B. (2000) Application of an elemental carbon analyser to the measurement of particulate levels from diesel vehicles (project number: C7014). Australia: Australian Coal Association Research Program.

Davies B. (2002) Diesel particulate control strategies at some Australian underground coal mines. Am Ind Hyg Assoc J; 63: 554–8.

Dosemici M, Wacholder S, Lubin J. (1990) Does nondifferential misclassification of exposure always bias a true effect toward the null value? Am J Epidemiol; 132: 746–8.

Gamble J. (2010) Lung cancer and diesel exhaust: a critical review of the occupational epidemiology literature. Crit Rev Toxicol; 40: 189–244.

Hesterberg TW, Bunn WB III, Chase GR *et al.* (2006) A critical assessment of studies on the carcinogenic potential of diesel exhaust. Crit Rev Toxicol; 36: 727–6.

Jurek AM, Greenalnd S, Maldonado G *et al.* (2005) Proper interpretation of non-differential misclassification effects: expectations vs observations. Int J Epidemiol; 34: 680–7.

Lodge JP. (1988) Methods of air sampling and analysis. Boca Raton, FL: CRC Press.

Morgenstern AS, Ash RM, Lynch JR. (1970) The evaluation of gas detector tube systems. I. Carbon monoxide. Am Ind Hyg Assoc J; 31: 630–2.

MSHA(2001) Diesel particulate matter exposure of underground metal and nonmetal miners: final rule. Fed Reg; 66: 5706–910.

Perkins JL. (1997) Modern industrial hygiene. New York, NY: Van Nostrand Reinhold.

Sorahan T, Gilthorpe MS. (1994) Non-differential misclassification of exposure always leads to an underestimate of risk: an incorrect conclusion. Occup Environ Med; 51: 839–40.

Stern MB, Mansdorf SZ. (1999) Applications and computational elements of industrial hygiene. Boca Raton, FL: CRC Press.

Stewart PA, Coble JB, Vermeulen R *et al.* (2010) The diesel exhaust in miners study: I. Overview of the exposure assessment process. Ann Occup Hyg; 54: 728–46.

Todd LA. (2003) Direct-reading instruments for determining concentrations of gases, vapors, and aerosols. In: Dinardi SR, ed. The occupational environment—its evaluation, control and management. Fairfax, VA: American Industrial Hygiene Association.

US EPA. (2002) Health assessment document for diesel engine exhaust (EPA/600/8-90/057F). Washington, DC: U.S. Environmental Protection Agency.

US EPA. (2010) Heavy trucks, buses, and engines: engine certification data. Available at http://www.epa.gov/otaq/certdata.htm#largeeng.

Vermeulen R, Coble JB, Lubin JH *et al.* (2010a) The diesel exhaust in miners study: IV. Estimating historical exposures to diesel exhaust in underground non-metal mining facilities. Ann Occup Hyg; 54: 774–88.

Vermeulen R, Coble JB, Yereb D *et al.* (2010b) The diesel exhaust in miners study: III. Interrelations between respirable elemental carbon and gaseous and particulate components of diesel exhaust derived from area sampling in underground non-metal mining facilities. Ann Occup Hyg; 54: 762–73.

WHO. (1976) Selected methods of measuring air pollutants (publication No. 24). Geneva: World Health Organization.

Yanowitz J, McCormick RL, Graboski MS. (2000) In-use emissions from heavy-duty vehicle emissions. Environ Sci Technol; 34: 729–40.

doi:10.1093/annhyg/mer005