

Reply

Received 20 January 2011; in final form 20 January 2011

We thank the journal for the opportunity to respond to the submission by Borak *et al.* (2011) regarding our recent papers on the Diesel Exhaust in Miners Study (DEMS). The primary concern of Borak *et al.* (2011) related to the precision of the historical exposure estimates at the study mines, which they claimed are unduly susceptible to several sources of random variability. We agree that imprecision exists in our estimates, but we disagree with the conclusions Borak *et al.* (2011) drew on this issue.

In epidemiological studies of chronic disease, some exposure imprecision typically exists. Efforts to minimize and characterize the imprecision are essential, and we have described these efforts in detail in our papers on the exposure assessment procedures (Coble *et al.*, 2010; Stewart *et al.*, 2010; Vermeulen *et al.*, 2010a,b). Borak *et al.* (2011) made the frequently seen error of assuming that any imprecision invalidates the epidemiological findings. Instead, the most useful scientific approach is to ask whether the exposure assessment method contributes to scientific understanding of the issue and whether, with the data available, a better approach could have been selected. In our study, we believe that the answer to the first question is clearly 'yes'. No epidemiological study to date has attempted to quantitatively characterize historical diesel exposures using such an extensive body of measurement data. Our answer to the second question is 'no'; we believe that given the available information, an optimal scientifically sound strategy was devised to reconstruct historical diesel exhaust (DE) exposures. The following are specific comments on a number of points raised by Borak *et al.* (2011) that we believe are erroneous and misleading.

Borak *et al.* (2011) identified several concerns about the individual respirable elemental carbon (REC) and carbon monoxide (CO) measurements. When considering imprecision, it is important to distinguish between error associated with any single measurement value and that associated with an aver-

age based on multiple measurements. The former is important for control and compliance evaluations, but the latter is more appropriate for epidemiological exposure assessment. The precision of a summary statistic increases most quickly with the addition of the first few measurements; as measurements continue to be added, the change in precision becomes increasingly less. Our estimates of average REC exposure level by mine and job were based on an average of 44, and a minimum of 5 (Coble *et al.*, 2010) personal measurements. These values served as 'anchor points' that were adjusted for changes over time in diesel equipment usage and ventilation rates based on a regression analysis of the historical CO measurements. The mine-specific regression models were based on hundreds of CO measurements per mine (range 248–2361) (Vermeulen *et al.*, 2010a). As such, except for a few estimates (<1% of the person-years), no aspect of the DEMS exposure assessment was based on a single, or even a few, measurements.

Borak *et al.* (2011) also raised concerns about the correlation between CO and REC. CO and REC emissions from individual diesel engines vary depending on engine speed and load. In our study, however, we were not investigating the emissions correlations for individual engines. Rather, we were investigating the correlations in the air concentrations of these contaminants measured in different areas of the facilities (Vermeulen *et al.*, 2010b), where multiple diesel engines were operating. Furthermore, the factor analyses indicated that CO and other gaseous components of DE and REC loaded on a presumed 'diesel exhaust' factor most strongly. This finding indicates that CO was a useful proxy for DE and, because it also was the component of DE most frequently measured historically in the study, it was the best DE surrogate for the historical DE exposure extrapolation.

Borak *et al.* (2011) indicated that we did not describe the CO measurements. It is unclear what more Borak *et al.* (2011) were expecting beyond what

was provided in Vermeulen *et al.* (2010a). In table 1, we followed the general practice of presenting, by mining facility, the number of measurements, the geometric means, and the geometric standard deviations of the CO measurements used in the modeling and in the evaluation of the models.

Borak *et al.* (2011) were incorrect in stating that we used fleet horsepower (HP) to predict CO and REC emissions. Rather, we used annual changes in the ratio of HP to ventilation rates to predict annual changes in DE, as indicated by CO. These ratios were then used to adjust reference REC means to derive historical REC estimates. The diesel equipment inventories showed fairly stable use of diesel equipment, with only incremental changes as new equipment was brought in or old equipment was retired. It was not essential, therefore, for inventories to be available for every year because the annual estimates could be reasonably interpolated from the existing inventories. In addition, the HP and ventilation summaries formed the basis for discussions with long-term workers at the study facilities and were adjusted if needed, further reducing the impact of any imprecision in the HP and ventilation estimates. Similar to single measurements, the imprecision in the HP and ventilation rate at a specific location on a given day will be high, but these estimates were based on expected averages. Thus, we are confident that we have provided reasonable estimates of the average relative changes in the ratio of HP and ventilation rate over the years of the study. In Vermeulen *et al.* (2010a), we provided estimates of this imprecision with our beta coefficients of the regression models. In spite of this imprecision, the associations between the main indicators of DE underground (HP and ventilation rates) and historical CO concentrations were strong and statistically significant.

Borak *et al.* (2011) were concerned that many of the confidence intervals for the REC measurements included 0. This is, of course, a reflection of the imprecision in the measurements. In this case, however, it would seem unlikely that the 'true' level of DE would be 0 in a confined environment where large diesel engines were running. As such, the most likely exposure levels would still be a measure of central tendency, such as the mean.

In conclusion, there is always uncertainty in estimates based on exposure reconstruction. We recognized this and provided considerable data to characterize the degree of uncertainty. We are confident that our approach is an improvement over previous efforts to quantify DE exposure in epidemiological studies.

This conclusion was supported by the evaluation of our exposure assessment methods. First, we compared our 1976 CO estimates to CO measurements taken in 1976 that were not used in the modeling of the historical trends (Vermeulen *et al.*, 2010a). The relative bias of the estimates was similar to that reported in other studies that have attempted to validate historical exposure estimates (Hornung *et al.*, 1994; Burstyn *et al.*, 2002; Stewart *et al.*, 2003; Astrakianakis *et al.*, 2006). Borak *et al.* (2011) suggested that if the confidence intervals were taken into account, these estimates would be different. Again, this is unlikely, as the variation in the 1976 measurements is independent from the estimated concentrations. As such, the estimate of the mean relative bias is unlikely to change. We also evaluated our methods by conducting three different sensitivity analyses to investigate the robustness of our estimation assumptions. We found high correlations between the three sets of alternative REC estimates and our primary REC estimates (Spearman rho = 0.87–0.99) (Stewart *et al.*, 2010).

The final concern of Borak *et al.* was that imprecision of estimates will lead to false-positive findings in our etiologic analyses. Borak *et al.* (2011), however, appear to have confused sampling variation with nondifferential misclassification of exposure. All epidemiological findings are subject to sampling error, which refers to the fact that we 'see' only one realization of data. A relative risk computed from data from a single study can, by chance alone, be greater or less than the true (unknown) relative risk. This idea is clearly demonstrated in the papers by Jurek *et al.* (2005) and Sorahan and Gilthorpe (1994), which Borak *et al.* (2011) mistakenly cited in support of their argument that our etiologic results could create false-positive findings. These papers, however, examined the effects of sampling variation, which is largely captured by confidence limits on the relative risk estimates, and which is not the same as nondifferential misclassification. Nondifferential misclassification of exposure is typically due to the observed exposures deviating from the true exposures independent of disease status (i.e. exposure misclassification is operating in the same way in the cases and the controls). In this case, the consequences of nondifferential misclassification can only be expressed in terms of its expected effect (i.e. the 'average' impact over repeated samplings of study data). In this, it almost always tends to result in false-negative, not false-positive, findings (Pearce *et al.*, 2007; Blair *et al.*, 2009). Borak *et al.* (2010) cited Dosemeci *et al.* (1990), who described how it is possible to increase an observed disease risk in

a lower exposure category by nondifferentially misclassifying subjects from a higher exposure category to the lower exposure category. An increase in a relative risk in the highest exposure category by nondifferential misclassification, however, cannot occur because there is no higher exposure category to serve as a source of misclassified individuals. In contrast, if no exposure category has an elevated relative risk (because there is no exposure–disease association), nondifferential misclassification has no effect, i.e. it cannot create a false-positive association because no category has an elevated disease risk. If, on the other hand, there is a true exposure–response gradient, nondifferential misclassification from a higher to a lower exposure category can increase the relative risk in the lower exposure category, but its effect would tend to distort a true monotonic gradient. This tends to diminish confidence that an association exists and increases the chance of a false-negative conclusion. Thus, the paper by Dosemeci *et al.* (1990) presents a scenario where nondifferential misclassification actually increases the likelihood of declaring a false-negative effect.

The study and historical monitoring data and descriptive information used for the development of historical quantitative exposure estimates of REC, along with the extensive description of the estimation process and the evaluation of the methods, are major strengths of the DEMS. We do not believe, nor do we claim, that our approach is without error. Rather, we believe that the procedure used was a sound assessment strategy to estimate historical exposure levels, as indicated by comparison to independent data, and is a significant improvement over earlier procedures and a contribution to the science of exposure assessment in general.

PATRICIA A. STEWART

*Division of Cancer Epidemiology and Genetics,
National Cancer Institute, Building EPS, Room
8006, 6120 Executive Blvd, Rockville, MD 20892,
USA*

JOSEPH B. COBLE

*Division of Cancer Epidemiology and Genetics,
National Cancer Institute, Building EPS, Room
8006, 6120 Executive Blvd, Rockville, MD 20892,
USA*

ROEL VERMEULEN

*Division of Cancer Epidemiology and Genetics,
National Cancer Institute, Building EPS, Room
8006, 6120 Executive Blvd, Rockville, MD 20892,*

*Utrecht University—IRAS, PO Box 80176,
Utrecht 3508 Td, the Netherlands, USA*

AARON BLAIR

*Division of Cancer Epidemiology and Genetics,
National Cancer Institute, Building EPS, Room
8006, 6120 Executive Blvd, Rockville, MD 20892,
USA*

JAY LUBIN

*Division of Cancer Epidemiology and Genetics,
National Cancer Institute, Building EPS, Room
8006, 6120 Executive Blvd, Rockville, MD 20892,
USA*

MICHAEL ATTFIELD

*Division of Respiratory Disease Studies, National
Institute of Occupational Safety and Health, 1095
Willowdale Road, Morgantown, WV 26505, USA*

DEBRA T. SILVERMAN

*Division of Cancer Epidemiology and Genetics,
National Cancer Institute, Building EPS, Room
8006, 6120 Executive Blvd, Rockville, MD 20892,
USA*

E-mail: silvermd@mail.nih.gov

REFERENCES

- Astrakianakis G, Seixas NS, Camp JE *et al.* (2006) Modeling, estimation and validation of cotton dust and endotoxin exposures in Chinese textile operations. *Ann Occup Hyg*; 50: 573–82.
- Blair A, Saracci R, Vineis P *et al.* (2009) Epidemiology, public health, and the rhetoric of false positives. *Environ Health Perspect*; 117: 1809–13.
- Borak J, Bunn WB, Chase GR *et al.* (2011) Comments on the diesel exhaust in miners study. *Ann Occup Hyg*; 55: 339–42.
- Burstyn I, Boffetta P, Burr GA *et al.* (2002) Validity of empirical models of exposure in asphalt paving. *Occup Environ Med*; 59: 620–4.
- 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: 759–73.
- Dosemeci M, Wacholder S, Lubin J. (1990) Does non-differential misclassification of exposure always bias a true effect toward the null value? *Am J Epidemiol*; 132: 746–8.
- Hornung RW, Greife AL, Stayner LT *et al.* (1994) Statistical model for prediction of retrospective exposure to ethylene oxide in an occupational mortality study. *Am J Ind Med*; 25: 825–36.
- Jurek AM, Greenland S, Maldonado G *et al.* (2005) Proper interpretation of non-differential misclassification effects: expectations vs observations. *Int J Epidemiol*; 34: 680–7.

- Pearce N, Checkoway H, Kriebel D. (2007) Bias in occupational epidemiology studies. *Occup Environ Med*; 64: 562–8.
- 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.
- Stewart P, Coble J, Vermeulen R *et al.* (2010) The Diesel Exhaust in Miners Study: I. Overview of the exposure assessment process. *Ann Occup Hyg*; 54: 728–46.
- Stewart PA, Lees PS, Correa A *et al.* (2003) Evaluation of three retrospective exposure assessment methods. *Ann Occup Hyg*; 47: 399–411.
- Vermeulen R, Coble JB, Lubin J *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.

doi:10.1093/annhyg/mer006