Ann. Occup. Hyg., Vol. 56, No. 1, pp. 80–91, 2012
Published by Oxford University Press
on behalf of the British Occupational Hygiene Society
doi:10.1093/annhyg/mer080

Combining a Job-Exposure Matrix with Exposure Measurements to Assess Occupational Exposure to Benzene in a Population Cohort in Shanghai, China

MELISSA C. FRIESEN[1]*, JOSEPH B. COBLE[1,2], WEI LU[3], XIAO-OU SHU[4], BU-TIAN JI[1], SHOUZHENG XUE[1], LUTZEN PORTENGEN[5], WONG-HO CHOW[1], YU-TANG GAO[6], GONG YANG[4], NATHANIEL ROTHMAN[1†] and ROEL VERMEULEN[1,5†]

[1]Occupational and Environmental Epidemiology Branch, Division of Cancer Epidemiology and Genetics, National Cancer Institute, Bethesda, MD 20892, USA; [3]Shanghai Municipal Center for Disease Control, 1380 Zhongshan Road, Shanghai, People's Republic of China; [4]Division of Epidemiology, Department of Medicine, Vanderbilt Epidemiology Center, Vanderbilt University School of Medicine, Nashville, TN 37203-1738, USA; [5]Environmental and Occupational Health Division, Institute for Risk Assessment Sciences, University of Utrecht, Utrecht, The Netherlands; [6]Department of Epidemiology, Shanghai Cancer Institute, Shanghai, People's Republic of China

Received 31 March 2011; in final form 18 July 2011; published online 5 October 2011

Background: Generic job-exposure matrices (JEMs) are often used in population-based epidemiologic studies to assess occupational risk factors when only the job and industry information of each subject is available. JEM ratings are often based on professional judgment, are usually ordinal or semi-quantitative, and often do not account for changes in exposure over time. We present an empirical Bayesian framework that combines ordinal subjective JEM ratings with benzene measurements. Our aim was to better discriminate between job, industry, and time differences in exposure levels compared to using a JEM alone.

Methods: We combined 63 221 short-term area air measurements of benzene exposure (1954–2000) collected during routine health and safety inspections in Shanghai, China, with independently developed JEM intensity ratings for each job and industry using a mixed-effects model. The fixed-effects terms included the JEM intensity ratings for job and industry (both ordinal, 0–3) and a time trend that we incorporated as a b-spline. The random-effects terms included job (n = 33) and industry nested within job (n = 399). We predicted the benzene concentration in two ways: (i) a calibrated JEM estimate was calculated using the fixed-effects model parameters for calendar year and JEM intensity ratings; (ii) a job-/industry-specific estimate was calculated using the fixed-effects model parameters and the best linear unbiased predictors from the random effects for job and industry using an empirical Bayes estimation procedure. Finally, we applied the predicted benzene exposures to a prospective population-based cohort of women in Shanghai, China (n = 74 942).

Results: Exposure levels were 13 times higher in 1965 than in 2000 and declined at a rate that varied from 4 to 15% per year from 1965 to 1985, followed by a small peak in the mid-1990s. The job-/industry-specific estimates had greater differences between exposure levels than the calibrated JEM estimates (97.5th percentile/2.5th percentile exposure level, ${}_{BG}R_{95B}$: 20.4 versus 3.0, respectively). The calibrated JEM and job-/industry-specific estimates were moderately correlated in any given year (Pearson correlation, r_p = 0.58). We classified only those jobs and

*Author to whom correspondence should be addressed. Tel: +1-301-594-7485; fax: +1-301-402-1819;
e-mail: friesenmc@mail.nih.gov
[2]Present address: 1412 Harmony Lane, Annapolis, MD 21409, USA.
†Joint First Authors.

industries with a job or industry JEM exposure probability rating of 3 (>50% of workers exposed) as exposed. As a result, 14.8% of the subjects and 8.7% of the employed person-years in the study population were classified as benzene exposed. The cumulative exposure metrics based on the calibrated JEM and job-/industry-specific estimates were highly correlated (r_p = 0.88).

Conclusions: We provide a useful framework for combining quantitative exposure data with expert-based exposure ratings in population-based studies that maximized the information from both sources. Our framework calibrated the ratings to a concentration scale between ratings and across time and provided a mechanism to estimate exposure when a job/industry group reported by a subject was not represented in the exposure database. It also allowed the job/industry groups' exposure levels to deviate from the pooled average for their respective JEM intensity ratings.

Keywords: benzene; job-exposure matrix; mixed-effects models; retrospective exposure assessment

INTRODUCTION

Generic job-exposure matrices (JEMs) are often used in population-based epidemiologic studies of occupational risk factors. Generic JEMs link a range of occupations and industries to exposure metrics for exposure agents of interest (Hoar *et al.*, 1980; Kauppinen *et al.*, 1997, 2009; Kennedy *et al.*, 2000; Kromhout and Vermeulen, 2001; Teschke *et al.*, 2002; Pukkala *et al.*, 2005). Validation and reliability studies have shown that generic JEMs usually have poor sensitivity and fail to account for heterogeneity in exposure levels within jobs and across time (Kromhout and Vermeulen, 2001; Teschke *et al.*, 2002). Nevertheless, JEMs are often the only possible exposure assessment approach in population-based studies that have limited occupational information. Thus, efforts to develop new approaches to improve the performance of JEMs remain important.

The values of a JEM's exposure metrics are generally based on professional judgment, although exposure databases have supplemented the subjective ratings in some JEMs (Kauppinen *et al.*, 1997, 2009; Pukkala *et al.*, 2005). When measurement data are available, exposure assessors can override the subjective rating and directly use the measurement data or they can use the measurements as part of the evidence in making their exposure decisions. The approach used may depend on the number of measurements available and the purpose for which the measurements were collected (e.g. representative samples, inspection samples). However, the two sources of information—professional judgment and exposure measurements—have not been combined in a transparent systematic framework thus far in the development of generic JEMs.

In this paper, we describe a systematic approach to combine both subjective ratings of exposure from a benzene-specific JEM and a large database of inspection measurements to assess historical benzene exposure for a prospective population-based cohort of women in Shanghai, China (Zheng *et al.*, 2005). We combined the two sources of information using a mixed-effects model framework, which has been previously used to combine group- with subgroup- or individual-level information in other exposure contexts (Friesen *et al.*, 2006b; McCracken *et al.*, 2009). In this framework, the group variables (i.e. JEM rating) are incorporated as fixed effects and the subgroup variables (i.e. job–industry combination) are incorporated as random effects. By treating subgroup variables as random effects, each subgroup contributes to the estimate of the group mean to which it belongs while also providing a mechanism to obtain an estimate of the subgroup's deviation in the exposure level compared to the group's estimate. The magnitude of the deviation is obtained from the model's estimates of the best linear unbiased predictors (BLUPs) from each random effect. The BLUPs work as a 'shrinkage estimator' that 'shrinks' the estimate toward the group estimate when the measurements are sparse and/or highly variable and 'pulls' the estimate toward the individual or subgroup estimate when more measurements are available and/or the exposure variability is low. The resulting model can be used to estimate both a group and a subgroup exposure estimate for the study subjects: the group estimate uses only the model's fixed-effects parameters, whereas the subgroup estimate uses both the fixed-effects parameters and the BLUPs. The BLUPs are assumed normally distributed, with a between-subgroup variance. Thus, for most subgroups, the subgroup estimate will be nearly identical to the group estimate. However, the subgroup estimate will vary from the group estimate if its exposure estimate differs from the other subgroups within the same group.

This approach provides two critical improvements to the exposure assessment compared to using either data source alone. First, our model calibrated the ratings across categories and across time to a measurement scale while accounting for data clusters. This calibrated the JEM and provided a way to estimate exposure for jobs and industries that were not represented in the exposure database. Second, our model allowed the job-/industry-exposure levels to deviate from their respective JEM intensity ratings by using the BLUP estimates from the job and industry random-effects model terms (obtained using an empirical Bayes procedure) to calculate a job-/industry-specific benzene exposure estimate. By combining the data sources, our aim was to better discriminate between time, job, and industry differences in benzene exposure in this study population compared to using only a JEM to estimate exposure.

METHODS

Study population

The Shanghai Women's Health Study is a population-based prospective cohort study of 74 942 women, aged 40–70 years, enrolled from seven communities in urban Shanghai between 1996 and 2000. This cohort has been described previously by Zheng *et al.* (2005). At enrollment, each woman completed a self-administered occupational history questionnaire, which included questions on the factory name, job title, description of work tasks, and the employment dates for all jobs held for at least 1 year since the age of 16. Each free-text job and employer reported in the occupational histories was standardized to codes in the Standard Chinese Classification of Industries and Occupations for the third national census (1982) by investigators from the Shanghai Cancer Institute. Other components of the enrollment questionnaire were administered by an interviewer and included demographic and lifestyle factors, medical history, residential history, and other non-occupational risk factors. Over 99% of the women were employed outside the home, with an average employment duration of 28 years at study enrollment.

Exposure sources

Job-exposure matrices. Investigators at the National Cancer Institute and Prevention (Shanghai CDC) used professional judgment to develop a job-specific and an industry-specific exposure matrix for benzene exposure in Shanghai, China. The jobs and industries were classified using the hierarchical coding scheme from the Standard Chinese Classification of Industries and Occupations for the third national census (1982), which included 306 job classes and 248 industry classes at the three-digit level. The job- and industry-exposure matrices were developed independently from each other, based on an approach described by Dosemeci *et al.* (1989). Thus, each work history record in the study was assigned both a job JEM rating and an industry JEM rating for each exposure metric.

For both the job- and industry-specific matrices, experts rated the probability (P_{job}, P_{ind}) and intensity (I_{job}, I_{ind}) of benzene exposure using an ordinal scale (0–3) for a single time point. The probability ratings were defined based on the estimated proportion of workers potentially exposed to benzene: 0 for no exposure, 1 for <5% exposed, 2 for 5–50% exposed, and 3 for >50% exposed. The intensity ratings were defined based on the 1980 maximum allowable concentration (MAC) for occupational benzene exposure in China (40 mg m^{-3}; 1 ppm = 3.25 mg m^{-3} at 20°C): 0 for very low or negligible exposure, 1 for above background but <10% of the MAC, 2 for 10–100% of the MAC, and 3 for above the MAC.

Exposure database. An exposure database was compiled from benzene short-term stationary (area) air measurements (n = 70 937) after removal of the duplicate, environmental, and non-routine measurements. The measurements were taken between 1953 and 2000 during inspections conducted by the Shanghai Health and Anti-epidemic Station, later renamed the Shanghai CDC, and abstracted from paper and electronic records maintained at municipal and district stations. The database included the sampling date, factory name, type of industry and job (coded using the same classification system used in the study population), sampling location, and air concentration. Monitoring records for 1990 and 1991 were missing. The factory names changed often, thus it was not possible to standardize the free-text factory name field. We were also unable to standardize the sample location field.

Data treatment

At the three-digit classification level, the inspection data included 108 of 171 exposed job classes (probability rating >0) and 129 of 167 exposed industry classes. Overall, there were 1028 job–industry combinations at the three-digit level in the inspection data. We combined similar jobs and industries to reduce the number of parameters that needed to be estimated while blind to the JEM ratings and measurements. This resulted in 33 job groups and 399 unique job/industry groups.

The estimated proportion of samples collected by powdered glass plate sampler versus syringes and analyzed by spectrophotometry versus gas chromatography in Shanghai, China, are reported in Fig. A1 in Supplementary Appendix 1 (available at *Annals of Occupational Hygiene* online). Generally, before 1980, historical benzene measurements in China were collected mainly by powdered glass plate sampler and analyzed by spectrophotometry. After 1980, the measurements were collected mainly by syringes and have been analyzed by gas chromatography. As this information was not recorded in the database, we used data from a separate benzene study under way at the National Cancer Institute to estimate each method's probability of use over time and adjusted the measurements accordingly (described in Supplementary Appendix 1, available at *Annals of Occupational Hygiene* online). This adjustment lowered pre-1980 exposure levels by an average of 15% (range 7–18%) and post-1980 exposure levels by an average of 0.6% (range 0–5%).

We treated any measurement with a value <3.25 mg m^{-3} as a sample below the limit of detection (LOD; 50% of measurements). We imputed values for these measurements using a previously described imputation procedure (Lubin *et al.*, 2004). The imputation procedure assumed that the data were log-normally distributed and used the measurements above the LOD to determine the parameters of that distribution. Values for each sample below the LOD were then randomly drawn from the exposure distribution. We repeated the imputation procedure five times, to obtain five data sets.

Statistical model

The model was developed using the 'proc mixed' restricted maximum likelihood method in SAS version 9.1 (SAS Institute Inc., Cary, NC, USA) using the structure shown in equation (1). The SAS code for the procedures below is included in Supplementary Appendix 2 (available at *Annals of Occupational Hygiene* online). We excluded 184 (0.3%) measurements where the JEM intensity rating for either job group or industry class was 0.

$$\mathrm{Ln}\left(Y_{\mathrm{adj}}\right)_{jifd} = \beta_0 + \sum \beta T + \sum \beta I_{\mathrm{job}} + \sum \beta\left(I_{\mathrm{job}} \times I_{\mathrm{ind}}\right) + bJ_j + bInd(J)_{ji} + bO_{jif} + \varepsilon_{jifd}. \quad (1)$$

The model terms were as follows:

- $\mathrm{Ln}(Y_{\mathrm{adj}})_{jifd}$ was the natural log-transformed benzene concentration (adjusted for analytical method) on *d*th day, in *f*th factory, in *i*th industry class, and *j*th job group.
- β_0 was the model intercept.
- *T* was the b-spline time trend terms obtained using 'proc transreg' to transform calendar year into a series of linear terms. We increased the numbers of knots from 1 upwards in increments of 1 until we observed no additional improvement in model fit based on the Akaike Information Criteria. We saw no improvement in model fit from increasing from 4 to 5 knots, thus we report only the model with a b-spline based on 4 knots and 3 degrees of freedom, which can be described by seven linear terms, β_1 to β_7 (number of knots + the number of degrees of freedom).
- I_{job} was a categorical variable for the JEM intensity ratings for job (β_1 to β_3).
- I_{ind} was a categorical variable for the JEM intensity ratings for industry class that we nested within the job rating I_{job} (β_1 to β_3). Preliminary analyses revealed that I_{job} explained more variability in the data than I_{ind} and that the joint effect of I_{job} and I_{ind} (nine unique combinations) could be described by including I_{job} as a main effect and parameterizing the effect of industry as a nested effect within jobs rated 2 or 3, but not 1. We use this simpler parameterization in the final model.
- *J* was a random-effect term for job class (b_1 to b_{33}).
- *Ind*(*J*) was a random-effect term for industry class nested within job class (b_1 to b_{399}).
- *O* was a random-effect term for the measurement occurrence number (b_1 to $b_{13\,282}$), which we defined as a unique sampling date–unstandardized factory name combination, and was used to account for correlation between repeated measurements collected within the same factory on the same sampling date (average number of measurements collected per occurrence: 4.1, maximum: 79).
- ε was the residual error (within occurrence).

J, Ind(*J*), *O*, and ε were assumed statistically independent and normally distributed with mean 0 and variances σ_{BJ}^2, $\sigma_{BI(J)}^2$, σ_{BO}^2, and σ_{WO}^2, representing the between-job, between-industry/-job, between-occurrence, and within-occurrence variance components, respectively. All variance components were assumed equal across job groups and industry classes using a uniform variance structure.

We obtained BLUP estimates of the coefficients for the job and job/industry random effects using an empirical Bayes procedure, where their unknown parameters were replaced by their restricted maximum likelihood estimates in the Bayesian calculations (Verbeke and Molenberghs, 2000). The above model was derived separately for each of the five imputed data sets and the results were then combined using 'proc mianalyze' to derive the overall parameter estimates, standard errors, and confidence limits for each model parameter, including the BLUP estimates for job and job/industry group.

Application of model to study population

The model was used to calculate two estimates of benzene exposure for the three-digit job/industry classifications reported in the study population. The first was a calibrated JEM estimate that was calculated from the fixed-effects parameters for year and JEM intensity ratings, which calibrated the JEM ratings while accounting for the clustering within the random effects. The second was a job-/industry-specific estimate that was calculated from the fixed effects and included the BLUP estimates for job group and industry class nested within job group, which allowed the exposure level to deviate from the pooled estimate for the JEM ratings.

We assigned the predicted benzene estimates from the models to a work history record from the Shanghai Women's Health Study only when one of the two conditions was met: (i) job probability rating = 3 or (ii) industry probability rating = 3 and job probability rating $\geq$ 1 because specificity is more important than sensitivity when the exposure prevalence is low (Flegal *et al.*, 1986; Dosemeci and Stewart, 1996; Kromhout and Vermeulen, 2001). All other work history records were assigned zero exposure. For the years before 1965, we assigned the 1965 exposure levels because of sparse data. For work history records with exposed job/industry groups that were not represented in the exposure database, we assigned a value of 0 to the job/industry BLUP estimate. We calculated the cumulative exposure of each study subject using both the calibrated JEM estimates and the job-/industry-specific estimates.

Comparison of model estimates

Comparing the calibrated JEM estimates to the job-/industry-specific estimates provides an estimate of the gain in information from using the BLUP estimates. For the subset of job/industry groups assigned as exposed based on the probability rating, we used the Pearson correlation statistic to examine the relationship between the calibrated JEM estimates and the job-/industry-specific estimates for the year 1980 (median year of employment in the study population). A high correlation would indicate minimal gain from the use of the BLUP estimates to calculate job-/industry-specific estimates. We also divided the 97.5th percentile of the predicted geometric means (GMs) by the 2.5th percentile as a measure of the ability of the model to discriminate between-job/-industry differences in exposure, which is approximately analogous to the ${}_{BG}R_{95}$ ratio described by Kromhout and Heederik (1995) to represent the exposure variability within a group. We also compared the effect of using the job-/industry-specific estimates versus the calibrated JEM estimates in the calculation of cumulative exposure metrics for the exposed study subjects in the Shanghai Women's Health Study using the Pearson correlation statistic.

RESULTS

Description of the data

The average method-adjusted benzene concentration in the data set was 43.3 mg m^{-3} [GM = 3.5 mg m^{-3}, geometric standard deviation (GSD) = 13.9, maximum = 1500 mg m$^{-3]}$). The measurements were distributed unevenly among the job and industry intensity ratings (Table 1). The majority of the measurements (77%) had an intensity rating of 3 for either job or industry; only 15% of the measurements had an intensity rating of 1 for either job or industry. The exposure measurements spanned nearly the entire study period in the Shanghai Women's Health Study (Fig. 1), although the measurements were sparse until the mid-1970s. The median year for the exposure measurements was 1988 and the median year for the employed person-years was 1980.

Benzene model

We report the model parameters for the fixed-effects model terms and the variance components in Table 2. Just under half of the database's large GSD was within-measurement occasion variance. When compared to the variance components from a model with only random effects (not shown), the time trend and JEM intensity ratings explained 43% of the between-job variability, 13% of the between-industry (nested within job) variability, and 14% of the between-occasion variability. Exposure levels increased with both increasing job and industry intensity rating. The highest job intensity rating was three times higher [3$\sim$1/exp($-$1.146)] than the lowest job intensity rating and the highest industry intensity rating was two times higher [2$\sim$1/exp($-$0.676)] than the lowest industry intensity rating.

Exposure levels in 1965 were 13 times higher than in 2000 (Fig. 2). The exposure levels declined at a rate of 4–15% per year from the mid-1960s to the mid-1980s, followed by a small peak in the mid-1990s. We observed a slight increase in exposures from 1954 (when the data began) to 1965, but the time trend in that period was relatively uncertain because only 5% of the measurements were collected in this

Table 1. Number of measurements used in model development by JEM intensity rating for job (I_{job}) and industry (I_{ind}).

	Number of measurements (%)			
	I_{ind} = 1	I_{ind} = 2	I_{ind} = 3	I_{ind} = all
I_{job} = 1	546 (1)	747 (1)	49 (0)	1342 (2)
I_{job} = 2	2138 (3)	11 223 (18)	9355 (15)	22 716 (36)
I_{job} = 3	5242 (8)	27 498 (44)	6239 (10)	38 979 (62)
I_{job} = all	7926 (13)	39 468 (63)	15 643 (25)	63 037 (100)

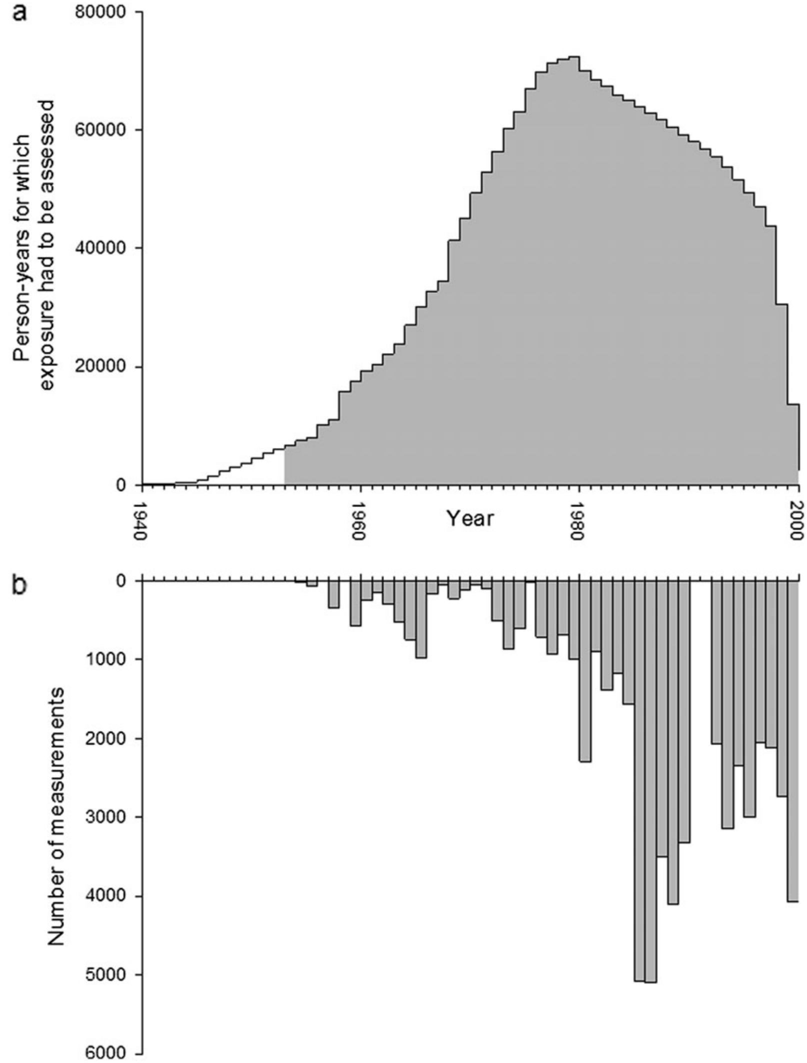

Fig. 1. Distribution of (a) employed person-years and (b) the number of available benzene measurements in the exposure database by calendar year. Graphical representation is based on Vlaanderen *et al.* (2010).

period. Because of this uncertainty, we assigned 1965 exposure levels for jobs held before 1965.

We assigned non-zero benzene exposure levels when the probability rating for job was 3 or when the probability rating for industry was 3 and the job probability rating was ≥1. Only 370 (36%) of the original 1028 three-digit job/industry groups in the measurement database met this criterion, but this subset accounted for 81.5% of the exposure measurements. We list the BLUP estimates and the 1980 exposure levels based on both the calibrated JEM and the job-/industry-specific estimates for this subset in Supplementary Table S1 (available at *Annals of Occupational Hygiene* online). The most frequently measured job group was painter (46 982 measurements, Standardized Occupation Code (SOC) 901), with an average predicted GM in 1980 of 4.7 mg m^{-3} based on the job-/industry-specific estimates and 3.8 mg m^{-3} based on the calibrated JEM estimates. The painters' exposure levels varied by industry. For example, the predicted job-/industry-specific exposure estimates in 1980 were 10.9 mg m^{-3} for painters in wooden ware manufacturing (n = 1405, Standardized Industry Code (SIC) 263), 7.7 mg m^{-3} in motor vehicle manufacturing (n = 1593, SIC 472), and 3.8 mg m^{-3} in the surface treatment of metals industry (n = 4850, SOC 472). These differences in exposure between painters in the same industry were captured with the industry random-effect term (the industry BLUPs were 0.44, 0.57,

Table 2. Fixed-effects model parameters and variance components.

Fixed-effects terms/variance components	Null model, parameter estimates[a]		Full model, parameter estimates[a]	
	Estimate	SE	Estimate	SE
Fixed-effects terms				
Intercept	1.16	0.11	0.682	0.234
Time trend terms[b]				
Spline term 1			2.510	0.301
Spline term 2			2.648	0.282
Spline term 3			2.904	0.218
Spline term 4			1.193	0.165
Spline term 5			2.551	0.184
Spline term 6			0.253	0.191
Spline term 7			Ref.	
Expert intensity rating (I)				
$I_{job} = 1$			−1.146	0.263
$I_{job} = 2$			−0.425	0.194
$I_{job} = 3$			Ref.	
$I_{job} = 2/3 \times I_{ind} = 1$			−0.676	0.151
$I_{job} = 2/3 \times I_{ind} = 2$			−0.483	0.138
$I_{job} = 2/3 \times I_{ind} = 3$			Ref.	
Variance components				
Between job	0.23	0.10	0.13	0.07
Between industry(job)	0.55	0.10	0.48	0.10
Between occurrence	3.97	0.07	3.40	0.06
Within occurrence	3.05	0.02	3.05	0.02
Total variance	7.80		7.06	

Ref., reference group; SE, standard error.
[a]The random-effects parameter estimates are presented in a separate table.
[b]See Supplementary Table S1 (available at *Annals of Occupational Hygiene* online) for values for spline terms 1–7 for each year.

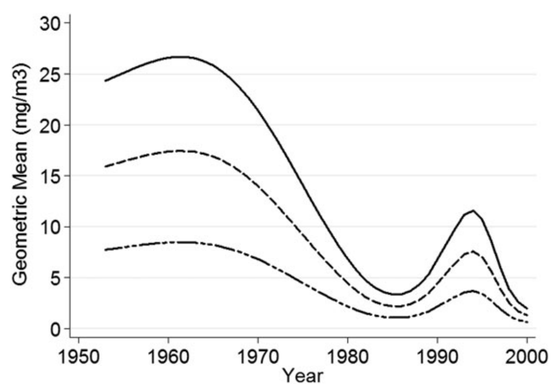

Fig. 2. Time trend in the predicted GM of benzene exposure, incorporating a linear extrapolation between 1986 and 2000 exposure levels, by job intensity rating (I_{job}). $I_{job} = 1$ (long dash line/short dash line); $I_{job} = 2$ (dashed line); $I_{job} = 3$ (solid line). Industry intensity rating was held constant at 3.

and −0.12 for these three industries, respectively), but the industry differences were not captured when the calibrated JEM estimates were used (4.2 mg m^{-3} for painters in each of these three industries). We illustrate the calculation of the job-/industry-specific exposure estimate using the example of painters in wooden ware manufacturing: the job-/industry-specific estimate is the product of the time trend multiplier (year 1980 = 6.8, Supplementary Table S1) multiplied by the antilogs of the model parameters for the job intensity rating [for job intensity = 3, exp(0) = 1], the industry probability rating [for industry intensity = 3, exp(0) = 1], the job BLUP [exp(0.04) = 1.04], and the industry BLUP [exp(0.44) = 1.55], resulting in a job-/industry-specific estimate of 10.9 mg m^{-3} (6.8 × 1 × 1 × 1.04 × 1.55 = 10.9). Machine operators in rubber production (n = 1084) and leather goods production workers (e.g. shoemakers using adhesives that include benzene, n = 5896) were also frequently monitored, with predicted job-/industry-specific GMs of 10.7 and 4.8 mg m^{-3} in 1980, respectively.

Application to study population

We show the distribution of employed person-years from hire year to study enumeration by the exposure probability ratings in Table 3; 14.8% of the subjects and 8.7% of the employed person-years were classified as benzene exposed based on our exposure criteria. For the exposed subjects, the mean duration of exposure was 17 years (maximum = 50 years; interquartile range, 9–24 years). For a small proportion of the exposed person-years, the exposure estimates were based on only the job group BLUP estimates (6.5%), rather than both the job and job/industry BLUP estimates.

Comparison of predicted benzene concentrations

We show the distribution of the predicted job-/industry-specific estimates for the year 1980 by job and industry intensity ratings in Fig. 3. The combination where both job and industry intensity ratings equaled 1 was not observed in the subset. We observed a large variability in assigned GMs within the JEM intensity ratings, with some degree of overlap between categories.

The job-/industry-specific estimates resulted in greater differences in exposure between job/industry groups ($_{BG}R_{95}$ = 20.4) than the calibrated JEM estimates ($_{BG}R_{95}$ = 3.0). The greater discrimination was not surprising because the fixed effects predicted only seven distinct exposure levels per year. For our reference year of 1980, the calibrated JEM estimates and the job-/industry-specific estimates were only moderately correlated (Pearson correlation, r_p = 0.58). For the exposed subjects, the correlation between the cumulative benzene metric based on the calibrated JEM estimates and the cumulative benzene metric based on the job-/industry-specific estimates was 0.88.

DISCUSSION

Table 3. Employed person-years in the Shanghai Women's Health Study by job (P_{job}) and industry (P_{ind}) probability rating.

	Employed person-years, in thousands (%)				
	P_{ind} = 0	P_{ind} = 1	P_{ind} = 2	P_{ind} = 3	P_{ind} = all
P_{job} = 0	273 (12.7)	467 (21.8)	181 (8.4)	27 (1.2)	823 (38.4)
P_{job} = 1	37 (1.7)	548 (25.6)	184 (8.6)	**29 (1.3)**	819 (38.2)
P_{job} = 2	5 (0.2)	58 (2.7)	178 (8.3)	**39 (1.8)**	384 (17.9)
P_{job} = 3	**2 (0.1)**	**31 (1.4)**	**69 (3.2)**	**15 (0.7)**	118 (5.5)
P_{job} = all	317 (14.8)	1104 (51.5)	613 (28.6)	110 (5.1)	2144 (100.0)

Bolded cells represent job/industry groups that were assigned exposure estimates.

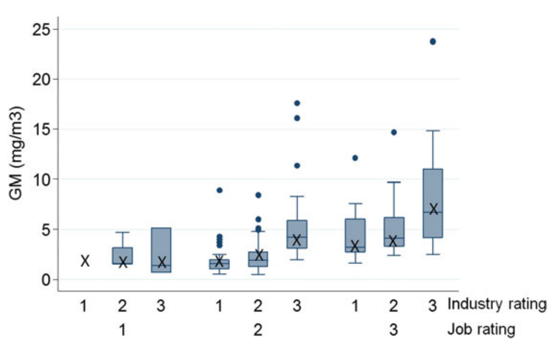

Fig. 3. Distribution of the predicted benzene exposure for job/industry groups with the same intensity ratings for the year 1980. The box shows the interquartile range (IQR) when the empirical Bayes terms were used to predict benzene concentration. The line within the box shows the median estimate, the whiskers show 1.5 × IQR, and the dots show the outliers. X shows the predicted concentration when the empirical Bayes terms were not used. The figure includes only the subset of job/industry groups that met the criteria for assigning a non-zero benzene exposure based on having a high exposure probability rating (>50% of workers exposed; n = 370).

In this paper, we presented a framework to maximize the information from two sources of exposure data, a benzene-specific JEM and exposure measurements, to derive benzene estimates for subjects in a population-based study. This approach allowed us to better discriminate between exposure levels by accounting for a 13-fold difference in exposure concentrations over four decades and a 20-fold difference in exposure concentrations between job/industry groups.

The framework used here is similar to the approaches used to combine individual-/subgroup-level and group-level exposure information using shrinkage estimators to maximize accuracy and precision (Seixas and Sheppard, 1996; Vermeulen *et al.*, 2003; Friesen *et al.*, 2006; McCracken *et al.*, 2009; Wild *et al.* 2002). In these examples and in this paper, the deviation of the individual (i.e. job/industry) mean from the group (i.e. JEM intensity rating) mean depended upon the number of measurements and their variability. The pooling of measurements at a group-level had the added benefit of allowing the assignment of a group-based exposure estimate to study subjects when no exposure information is available at the individual level. In this study, we used a tiered approach, starting from more detail to less detail, to assign benzene estimates to the study subjects. At the most detailed level, we assigned a job-/industry-specific exposure level when that job/industry group was represented in the exposure database. In its absence, we assigned a job-specific exposure level, which accounted for 6.5% of the exposed person-years. We could have also assigned a JEM intensity rating-specific estimate when the job was not represented in the exposure database; however, this scenario did not occur in this study given the extended exposure database that was created. The moderate correlation between the calibrated JEM estimates and job-/industry-specific estimates indicates that we were able to extract additional information from the exposure data set when we included the BLUP estimates beyond simply calibrating the JEM intensity ratings across time.

An alternate approach for combining expert ratings and measurements is to use a full Bayesian framework that uses the subjective ratings as exposure priors, which are then updated using measurement data to form a posterior distribution of the parameters of interest using Markov Chain Monte Carlo simulations (Ramachandran and Vincent, 1999; Ramachandran, 2001; Ramachandran *et al.*, 2003; Hewett *et al.*, 2006; Sottas *et al.*, 2009). Although it would have been possible to implement a Bayesian framework in this study, we did not do so for two reasons. First, studies that incorporated a Bayesian approach had study-specific information on the variability in the subjective ratings in the form of probability distributions of the expert inputs. In this study, we had no measure of the uncertainty or variability for the subjective ratings to provide informative priors and there was limited information in the literature to provide reasonable estimates for this context. Second, the full Bayesian approach requires specialized software and knowledge. Instead, we chose a data-driven approach that could be implemented in any software capable of mixed-effects (hierarchical) models and is a model structure that is within the tool kit of many research-based industrial hygienists.

Accounting for the changes in benzene exposure across time was essential to reduce exposure misclassification. Our time trend analyses provide a comprehensive evaluation of exposure time trends for benzene in China for a 4.5-decade time period. Previous trend analyses have been limited to summarizing the reported median benzene exposure concentrations by year from 1979 to 2001 based on published data, rather than routine inspection measurements (Liang *et al.*, 2005). The 13-fold exposure difference between 1965 and 2000 seen here parallels the exposure time trends reported in North American and European countries for other exposures, which generally observed a strong decline in exposures in the 1970s and early 1980s (Symanski *et al.*, 1998a,b; Teschke *et al.*, 1999; Kromhout and Vermeulen, 2000; Vermeulen *et al.*, 2000; Friesen *et al.*, 2006a; Lavoue *et al.*, 2006, 2008; Creely *et al.*, 2007; de Vocht *et al.*, 2008; Galea *et al.*, 2009; van Tongeren *et al.*, 2009). The decline in exposure levels was not linear on the natural or log-transformed scale. We captured this non-linearity using a b-spline time trend, but the data were too sparse to capture potential differences in time trends between jobs or industries. The period between 1965 and 1985 represented the period with the largest changes in exposure, with exposure levels declining between 4 and 15% per year. The small peak in exposure levels in the mid-1990s was somewhat unexpected. While this peak likely represents a period of higher exposure levels coinciding with Chinese economic reforms starting in the late 1980s through the beginning of the 1990s, it may also be (partially) a data artifact from changes in sampling method, analytical method or measurement strategy (as described in Supplementary Appendix 1, available at *Annals of Occupational Hygiene* online). Support for a data artifact can be found in the fact that the peak in exposure is driven by a decrease in the number of measurements $<$LOD while detectable levels continued to decline during this time period. Despite the uncertainty in exposure levels for this latter period, the cumulative exposures for the subjects were robust, with a correlation of 0.997 between the cumulative metrics with and without excluding the mid-1990s peak. Overall, this strong time trend indicates that a JEM based on a single time point would miss crucial exposure differences between subjects if the study years spanned this period of major change, as it did in our application to the Shanghai Women's Health Study.

The predicted exposure levels monotonically increased with increasing intensity rating for both job and industry. To assess the experts' intensity estimates in the JEM, we qualitatively compared the predicted GMs for 1980 (median study year) for the job intensity ratings to the cut points used by the experts in the JEM development. The 1980 predicted GM for job intensity ratings of 1, 2, and 3 were 2.5, 4.0, and 7.5 mg m^{-3}, respectively. Because the experts aimed to assign JEM intensity ratings based on the arithmetic mean (AM) of full-shift samples, we did not use the large data-driven GSD from our short-term samples to calculate an estimate AM. Instead, we calculated estimates of the corresponding AMs based on a GSD of 2.5, which was chosen based on the previously reported variance components for chemical gases and vapors (Kromhout *et al.*, 1993). Thus, these GMs approximately coincide with AMs of 4, 6, and 11 mg m^{-3} $\{AM = GM \times \exp[0.5 \times \ln(GSD)^2]\}$ (Aitchison and Brown, 1963). The predicted AMs were consistent with the experts' cut points for intensity ratings of 1 (AM < 4 mg m^{-3}) and 2 (AM $=$ 4–40 mg m^{-3}). However, the predicted AM was much lower than the cut point for intensity rating 3 (AM > 40 mg m^{-3}). The relation between the predicted AM and the JEM intensity categories will change across time, but this qualitative comparison suggests the experts were reasonably able to differentiate exposure levels at the study midpoint.

Overall, our predicted exposure levels were lower than exposure levels previously reported in a multi-industry multi-center benzene-exposed cohort (Dosemeci *et al.*, 1994). High levels of benzene exposure have also been reported in many workplaces (Rothman *et al.*, 1996; Vermeulen *et al.*, 2004; Liang *et al.*, 2005; Wang *et al.*, 2006; Liu *et al.*, 2009). Many of these reports relate to measurements collected during surveys investigating benzene-poisoning events or in workplaces targeted for molecular epidemiology studies because of known high exposure levels and thus these reports do not necessarily represent the average exposure scenario. The Shanghai CDC database used here includes many measurements with very high exposure concentrations that were similar to those reported in the above studies; however, on average, the exposure levels were lower than these previous reports. The lower levels seen here may be because Shanghai may have been quicker to adopt technological changes that lowered exposure levels and may have had more strict enforcement. Our lower exposure levels may have also resulted from limitations in the exposure data, which we describe below. Unfortunately, we could not validate the exposure models because there was no other comprehensive external data set available. This is a common limitation of most models developed for retrospective exposure assessment and previous validation efforts have been limited (Hornung *et al.*, 1994; Burstyn *et al.*, 2002; Friesen *et al.*, 2005; Vermeulen *et al.*, 2010). However, the models' benzene estimates were robust to assumptions and data treatment approaches.

The database of benzene measurements was a critical component of our framework, but it provided only an 'alloyed' gold standard of exposure (Wacholder *et al.*, 1993; Stewart *et al.*, 1996). The biggest limitation of this data set was that it included only short-term area samples, whereas personal full-shift measurements provide a better measure of the average exposure experienced by the workers in each job and industry. Area samples can both over- or underestimate the average personal exposure depending on the placement of the sampling device in relation to the source and the activity pattern of the worker. Routine sampling in China occurred most often in the far field but still close to the actual work process. This probably would have led to an underestimation of personal benzene exposure for jobs where the exposure occurred mostly in the near field and where workers were performing the exposed tasks for most of the day. In situations where this did not apply, the area samples might have overestimated the exposure. Another limitation is that this data set only represents the exposure levels of the factories that fell under the Shanghai CDC inspection program and might not have covered all factories in the region. For instance, small companies might not have been visited regularly and inspection visits of these smaller factories may have been limited to post-incident surveys. The large number of samples below the LOD provided some support that the measurements represented a wide range of exposure scenarios, rather than only worst-case scenarios.

There were also limitations to the data itself. First, it was uncertain whether exposure levels were higher or lower pre-1965 because of sparse data in this time period. As a result, we assumed a constant exposure level for jobs held pre-1965 (9% of exposed person-years). Second, the exposure database did not identify the measurement or analytical method used. As a result, we used Shanghai-specific data from another study to calibrate the measurements to the syringe measurement method and gas chromatography analysis method. Although these assumptions may change the magnitude of the exposure estimates, the rank order of the exposure estimates was nearly perfectly correlated when we varied our assumptions by changing how we treated the pre-1965 levels, by treating the mid-1990s exposure peak as real or as a data artifact, or by adjusting or not adjusting for the analytical method.

The ~15% exposure prevalence for benzene is not surprising for a cohort of Asian women as Asian women had higher employment rates in this time period, especially in manufacturing, than women in most other countries (ILO, 2010). While the true prevalence of benzene exposure among the women in this cohort is unknown, the prevalence of exposure that we estimated for this female study population is higher than the 8% prevalence observed for the controls in a separate Shanghai case–control study (Wong *et al.*, 2010) and is higher or consistent with the 4–18% prevalence of exposure in controls in previous European and US case–control studies (Seidler *et al.*, 2007; Wang *et al.*, 2009; Cocco *et al.*, 2010; Peplonska *et al.*, 2010). We could vary the proportion of exposure by changing the criteria for assigning exposure. Our *a priori* criteria were chosen to maximize specificity in the exposure assignment, which is recommended when the exposure prevalence is moderately low (Flegal *et al.*, 1986; Dosemeci and Stewart, 1996; Kromhout and Vermeulen, 2001). If increased sensitivity was desired, we could relax the criteria by including subjects whose job- and industry-exposure probability rating were both 2 (5–50% of workers exposed) as exposed. This would increase the exposed person-years in the study from 8.5 to 16.8% and the number of ever-exposed subjects from 14.8 to 27.8%. We will examine the robustness of exposure–disease associations to varying criteria for applying benzene estimates in future epidemiologic analyses in the Shanghai Women's Health Study.

In summary, we present a novel framework to combine a JEM with exposure measurements to maximize the exposure information. Our approach could be extended to other exposures, JEMs, and exposure databases. We were able to calibrate the ratings to a concentration scale between ratings and across time and provide a mechanism to estimate exposure when a job/industry group reported by a subject was not represented in the exposure database. We also allowed the job/industry groups' exposure levels to deviate from the pooled average for the JEM ratings, which would correct for some exposure misclassification in the JEM estimates if we assume the data are reasonably representative of exposure conditions. Although this approach cannot account for the inherent variability within jobs and industries, we expect that the greater discrimination in exposure levels from combining the JEM and exposure measurements should improve the exposure assessment in this study compared to using the JEM alone.

FUNDING

US National Institutes of Health (R37 CA70867); Intramural Research Program of the National Institutes of Health (contract N02 CP1101066).

Acknowledgements—The authors thank the research staff of the Shanghai Women's Health Study and the Shanghai CDC for their support and contributions to the study.

REFERENCES

Aitchison J, Brown JAC. (1963) The lognormal distribution. Cambridge, UK: Cambridge University Press.

Burstyn I, Boffetta P, Burr GA *et al*. (2002) Validity of empirical models of exposure in asphalt paving. Occup Environ Med; 59: 620–4.

Cocco P, t'Mannetje A, Fadda D *et al*. (2010) Occupational exposure to solvents and risk of lymphoma subtypes: results from the Epilymph case-control study. Occup Environ Med; 67: 341–7.

Creely KS, Cowie H, Van Tongeren M *et al*. (2007) Trends in inhalation exposure—a review of the data in the published scientific literature. Ann Occup Hyg; 51: 665–78.

de Vocht F, Vermeulen R, Burstyn I *et al*. (2008) Exposure to inhalable dust and its cyclohexane soluble fraction since the 1970s in the rubber manufacturing industry in the European Union. Occup Environ Med; 65: 384–91.

Dosemeci M, Li GL, Hayes RB *et al*. (1994) Cohort study among workers exposed to benzene in China: II. Exposure assessment. Am J Ind Med; 26: 401–11.

Dosemeci M, Stewart PA. (1996) Recommendations for reducing the effects of exposure misclassification on relative risk estimates. Occup Hyg; 3: 169–76.

Dosemeci M, Stewart PA, Blair A. (1989) Evaluating occupation and industry separately to assess exposures in case-control studies. Appl Ind Hyg; 4: 256–59.

Flegal KM, Brownie C, Haas JD. (1986) The effects of exposure misclassification on estimates of relative risk. Am J Epidemiol; 123: 736–51.

Friesen MC, Davies HW, Teschke K *et al*. (2005) Predicting historical dust and wood dust exposure in sawmills: model development and validation. J Occup Environ Hyg; 2: 650–8.

Friesen MC, Demers PA, Spinelli JJ *et al*. (2006a) From expert-based to quantitative retrospective exposure assessment at a Soderberg aluminum smelter. Ann Occup Hyg; 50: 359–70.

Friesen MC, Macnab YC, Marion SA *et al*. (2006b) Mixed models and empirical bayes estimation for retrospective exposure assessment of dust exposures in Canadian sawmills. Ann Occup Hyg; 50: 281–8.

Galea KS, Van Tongeren M, Sleeuwenhoek AJ *et al*. (2009) Trends in wood dust inhalation exposure in the UK, 1985–2005. Ann Occup Hyg; 53: 657–67.

Hewett P, Logan P, Mulhausen J *et al*. (2006) Rating exposure control using Bayesian decision analysis. J Occup Environ Hyg; 3: 568–81.

Hoar SK, Morrison AS, Cole P *et al*. (1980) An occupation and exposure linkage system for the study of occupational carcinogenesis. J Occup Med; 22: 722–6.

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.

ILO. (2010) Women in labour markets: measuring progress and identifying challenges. Geneva, Switzerland: Publisher is the International Labor Organization.

Kauppinen T, Heikkila P, Plato N *et al*. (2009) Construction of job-exposure matrices for the Nordic Occupational Cancer Study (NOCCA). Acta Oncol; 48: 791–800.

Kauppinen T, Teschke K, Savela A *et al*. (1997) International data base of exposure measurements in the pulp, paper and paper product industries. Int Arch Occup Environ Health; 70: 119–27.

Kennedy SM, Le Moual N, Choudat D *et al*. (2000) Development of an asthma specific job exposure matrix and its application in the epidemiological study of genetics and environment in asthma (EGEA). Occup Environ Med; 57: 635–41.

Kromhout H, Heederik D. (1995) Occupational epidemiology in the rubber industry: implications of exposure variability. Am J Ind Med; 27: 171–85.

Kromhout H, Symanski E, Rappaport SM. (1993) A comprehensive evaluation of within- and between-worker components of occupational exposure to chemical agents. Ann Occup Hyg; 37: 253–70.

Kromhout H, Vermeulen R. (2000) Long-term trends in occupational exposure: Are they real? What causes them? What shall we do with them? Ann Occup Hyg; 44: 325–7.

Kromhout H, Vermeulen R. (2001) Application of job-exposure matrices in studies of the general population: some clues to their performance. Eur Respir Rev; 11: 80.

Lavoue J, Vincent R, Gerin M. (2006) Statistical modelling of formaldehyde occupational exposure levels in French industries, 1986–2003. Ann Occup Hyg; 50: 305–21.

Lavoue J, Vincent R, Gerin M. (2008) Formaldehyde exposure in U.S. industries from OSHA air sampling data. J Occup Environ Hyg; 5: 575–87.

Liang YX, Wong O, Armstrong T *et al*. (2005) An overview of published benzene exposure data by industry in China, 1960–2003. Chem Biol Interact; 153–154: 55–64.

Liu H, Liang Y, Bowes S *et al*. (2009) Benzene exposure in industries using or manufacturing paint in China—a literature review, 1956–2005. J Occup Environ Hyg; 6: 659–70.

Lubin JH, Colt JS, Camann D *et al*. (2004) Epidemiologic evaluation of measurement data in the presence of detection limits. Environ Health Perspect; 112: 1691–6.

McCracken JP, Schwartz J, Bruce N *et al*. (2009) Combining individual- and group-level exposure information: child carbon monoxide in the Guatemala woodstove randomized control trial. Epidemiology; 20: 127–36.

Peplonska B, Stewart P, Szeszenia-Dabrowska N *et al*. (2010) Occupational exposure to organic solvents and breast cancer in women. Occup Environ Med; 67: 722–9.

Pukkala E, Guo J, Kyyronen P *et al*. (2005) National job-exposure matrix in analyses of census-based estimates of occupational cancer risk. Scand J Work Environ Health; 31: 97–107.

Ramachandran G. (2001) Retrospective exposure assessment using Bayesian methods. Ann Occup Hyg; 45: 651–67.

Ramachandran G, Banerjee S, Vincent JH. (2003) Expert judgment and occupational hygiene: application to aerosol speciation in the nickel primary production industry. Ann Occup Hyg; 47: 461–75.

Ramachandran G, Vincent JH. (1999) A Bayesian approach to retrospective exposure assessment. Appl Occup Environ Hyg; 14: 547–57.

Rothman N, Li GL, Dosemeci M *et al*. (1996) Hematotoxicity among Chinese workers heavily exposed to benzene. Am J Ind Med; 29: 236–46.

Seidler A, Mohner M, Berger J *et al*. (2007) Solvent exposure and malignant lymphoma: a population-based case-control study in Germany. J Occup Med Toxicol; 2: 2.

Seixas NS, Sheppard L. (1996) Maximizing accuracy and precision using individual and grouped exposure assessments. Scand J Work Environ Health; 22: 94–101.

Sottas PE, Lavoue J, Bruzzi R *et al*. (2009) An empirical hierarchical Bayesian unification of occupational exposure assessment methods. Stat Med; 28: 75–93.

Stewart PA, Lees PS, Francis M. (1996) Quantification of historical exposures in occupational cohort studies. Scand J Work Environ Health; 22: 405–14.

Symanski E, Kupper LL, Hertz-Picciotto I *et al*. (1998a) Comprehensive evaluation of long-term trends in occupational exposure: Part 2. Predictive models for declining exposures. Occup Environ Med; 55: 310–6.

Symanski E, Kupper LL, Rappaport SM. (1998b) Comprehensive evaluation of long-term trends in occupational exposure: Part 1. Description of the database. Occup Environ Med; 55: 300–9.

Teschke K, Marion SA, Vaughan TL *et al*. (1999) Exposures to wood dust in U.S. industries and occupations, 1979 to 1997. Am J Ind Med; 35: 581–9.

Teschke K, Olshan AF, Daniels JL *et al*. (2002) Occupational exposure assessment in case-control studies: opportunities for improvement. Occup Environ Med; 59: 575–93; discussion 94.

van Tongeren M, Galea KS, Ticker J *et al*. (2009) Temporal trends of flour dust exposure in the United Kingdom, 1985–2003. J Environ Monit; 11: 1492–7.

Verbeke G, Molenberghs G. (2000) Linear mixed models for longitudinal data. New York, NY: Springer-Verlag.

Vermeulen R, Bos RP, Pertijs J *et al*. (2003) Exposure related mutagens in urine of rubber workers associated with inhalable particulate and dermal exposure. Occup Environ Med; 60: 97–103.

Vermeulen R, Coble JB, Lubin JH *et al*. (2010) 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, de Hartog J, Swuste P *et al*. (2000) Trends in exposure to inhalable particulate and dermal contamination in the rubber manufacturing industry: effectiveness of control measures implemented over a nine-year period. Ann Occup Hyg; 44: 343–54.

Vermeulen R, Li G, Lan Q *et al*. (2004) Detailed exposure assessment for a molecular epidemiology study of benzene in two shoe factories in China. Ann Occup Hyg; 48: 105–16.

Vlaanderen J, Fransman W, Miller B *et al*. (2010) A graphical tool to evaluate temporal coverage of occupational history by exposure measurements. Occup Environ Med; 67: 636–8.

Wacholder S, Armstrong B, Hartge P. (1993) Validation studies using an alloyed gold standard. Am J Epidemiol; 137: 1251–8.

Wang L, Zhou Y, Liang Y *et al*. (2006) Benzene exposure in the shoemaking industry in China, a literature survey, 1978–2004. Regul Toxicol Pharmacol; 46: 149–56.

Wang R, Zhang Y, Lan Q *et al*. (2009) Occupational exposure to solvents and risk of non-Hodgkin lymphoma in Connecticut women. Am J Epidemiol; 169: 176–85.

Wild P, Sauleau EA, Bourgkard E *et al*. (2002) Combining expert ratings and exposure measurements: a random effect paradigm. Ann Occup Hyg; 46: 479–87.

Wong O, Harris F, Armstrong TW *et al*. (2010) A hospital-based case-control study of acute myeloid leukemia in Shanghai: analysis of environmental and occupational risk factors by subtypes of the WHO classification. Chem Biol Interact; 184: 112–28.

Zheng W, Chow WH, Yang G *et al*. (2005) The Shanghai Women's Health Study: rationale, study design, and baseline characteristics. Am J Epidemiol; 162: 1123–31.