Annals of Work Exposures and Health, 2017, Vol. 61, No. 1, 56–66
doi: 10.1093/annweh/wxw009
Original Article

Original Article

Evaluating Exposure–Response Associations for Non-Hodgkin Lymphoma with Varying Methods of Assigning Cumulative Benzene Exposure in the Shanghai Women's Health Study

Melissa C. Friesen[1*#], Bryan A. Bassig[1#], Roel Vermeulen[2], Xiao-Ou Shu[3], Mark P. Purdue[1], Patricia A. Stewart[1,4], Yong-Bing Xiang[5], Wong-Ho Chow[6], Bu-Tian Ji[1], Gong Yang[3], Martha S. Linet[7], Wei Hu[1], Yu-Tang Gao[5], Wei Zheng[3], Nathaniel Rothman[1†] and Qing Lan[1†]

[1]Occupational and Environmental Epidemiology Branch, Division of Cancer Epidemiology and Genetics, National Cancer Institute, 9609 Medical Center Drive, Rm. 6E634, Rockville, MD 20850, USA; [2]Institute for Risk Assessment Sciences, Utrecht University, Yalelaan 2, Utrecht 3508 TD, The Netherlands; [3]Division of Epidemiology, Department of Medicine, Vanderbilt Epidemiology Center, Vanderbilt-Ingram Cancer Center, Vanderbilt University Medical Center, 2525 West End Avenue, Nashville, TN 37203, USA; [4]Stewart Exposure Assessments, LLC, 6045 N 27th St, Arlington, VA 22207, USA; [5]Department of Epidemiology, Shanghai Cancer Institute, 2200 Xietu Road, Xuhui, Shanghai 200032, China; [6]Department of Epidemiology, The University of Texas MD Anderson Cancer Center, 1155 Pressler Street, Houston, TX 77030, USA; [7]Radiation Epidemiology Branch, Division of Cancer Epidemiology and Genetics, National Cancer Institute, 9609 Medical Center Drive, Rockville, MD 20850, USA

*Author to whom correspondence should be addressed. Tel: +1-240-276-7278; e-mail: friesenmc@mail.nih.gov
[#]These authors equally contributed to this work.
[†]These authors co-supervised this work.

Submitted 28 July 2016; revised 14 September 2016; editiorial decision 16 September 2016; revised version accepted 8 November 2016.

Abstract

Objectives: To provide insight into the contributions of exposure measurements to job exposure matrices (JEMs), we examined the robustness of an association between occupational benzene exposure and non-Hodgkin lymphoma (NHL) to varying exposure assessment methods.
Methods: NHL risk was examined in a prospective population-based cohort of 73 087 women in Shanghai. A mixed-effects model that combined a benzene JEM with >60 000 short-term, area benzene inspection measurements was used to derive two sets of measurement-based benzene estimates: 'job/industry-specific' estimates (our presumed best approach) were derived from the model's fixed effects (year, JEM intensity rating) and random effects (occupation, industry); 'calibrated JEM'

estimates were derived using only the fixed effects. 'Uncalibrated JEM' (using the ordinal JEM ratings) and exposure duration estimates were also calculated. Cumulative exposure for each subject was calculated for each approach based on varying exposure definitions defined using the JEM's probability ratings. We examined the agreement between the cumulative metrics and evaluated changes in the benzene–NHL associations.

Results: For our primary exposure definition, the job/industry-specific estimates were moderately to highly correlated with all other approaches (Pearson correlation 0.61–0.89; Spearman correlation > 0.99). All these metrics resulted in statistically significant exposure–response associations for NHL, with negligible gain in model fit from using measurement-based estimates. Using more sensitive or specific exposure definitions resulted in elevated but non-significant associations.

Conclusions: The robust associations observed here with varying benzene assessment methods provide support for a benzene–NHL association. While incorporating exposure measurements did not improve model fit, the measurements allowed us to derive quantitative exposure–response curves.

Keywords: benzene; exposure assessment methodology

Introduction

In population-based epidemiologic studies, retrospective estimates for occupational exposures are often obtained using generic, or population-level, job exposure matrices (JEMs) that link the participants' work histories to exposure decisions assigned based on standardized occupation and industry classification systems. The JEM's exposure estimates are typically derived using expert judgment (Kromhout and Vermeulen, 2001), with exposure measurements becoming more commonly used to anchor the experts' estimates (Vincent and Jeandel, 2001; Pukkala *et al.*, 2005; Kauppinen *et al.*, 2009).

To develop a more systematic, transparent, rigorous, and reproducible use of measurements with JEM estimates, a novel mixed-effects statistical model framework was previously developed to systematically combine ordinal JEM intensity estimates with measurement data (Peters *et al.*, 2011b; Friesen *et al.*, 2012; Koh *et al.*, 2014). The framework incorporated the ordinal JEM intensity estimates and calendar year as fixed-effects terms in the model and incorporated subgroups such as job, industry, and/or country as random-effects terms. The fixed-effects terms were used to assign measurement levels to the JEM ordinal ratings across time. The random-effects terms were used to obtain the 'best linear unbiased predictors' (BLUPs) for each subgroup, which were used to modify the JEM estimate to calculate subgroup-specific estimates. The BLUPs work as a shrinkage estimator that pulls the subgroup estimate (i.e. job, industry, and/or country-specific estimate) towards the group estimate (i.e. JEM rating estimate) when data are sparse or when within-subgroup variability is large and towards the subgroup estimate when data for that subgroup are plentiful and/or within-subgroup variability is small. Previous evaluations have reported Pearson correlation coefficients between cumulative subgroup-specific estimates and calibrated JEM estimates (that exclude the random effects of job, industry, and country) of 0.88 for benzene (14.8% exposed, model based on 63 221 measurements) (Friesen *et al.*, 2012), 0.79 for lead fume (8% exposed, 20 084 measurements) (Koh *et al.*, 2014), 0.84 for lead dust (4% exposed, 5383 measurements) (Koh *et al.*, 2014), and 0.74 for silica (24.8% exposed, 23 640 measurements) (Peters *et al.*, 2011a), indicating at least some differences in exposure are identified using this framework. The subgroup-specific estimates are expected to capture greater discrimination of exposure levels across subgroups than would be captured using the calibrated JEM estimates alone and are expected to reduce exposure misclassification and improve our ability to detect exposure-disease associations. In the absence of a gold standard of exposure, the improvements obtained by using these presumably more refined estimates can be examined indirectly through sensitivity analyses (Kromhout *et al.*, 1999; Loomis *et al.*, 1999; Heederik and Attfield, 2000; Peters *et al.*, 2011a, 2012).

In this article, our objective was to obtain insight into the contribution of these JEM refinements by examining the robustness of a benzene–non-Hodgkin lymphoma (NHL) association to varying JEM-based methods of assigning cumulative benzene exposure. For a population-based prospective cohort of Shanghai women, the above-mentioned mixed-effects model framework (Friesen *et al.*, 2012) was used to obtain job/industry-specific estimates of historical occupational benzene exposure previously used by Bassig *et al.* (2015) in analyses examining the cohort's risk of NHL. Bassig *et al.* reported significant trends in NHL risk with

increasing cumulative benzene exposure levels (trend *P*-value = 0.005), with a hazard ratio (HR) of 2.16 [95% confidence interval (CI): 1.17, 3.98] for the highest exposure tertile. In the sensitivity analyses reported here, we compared alternative exposure assessment approaches to our previously used and presumed best measure, the job/industry-specific estimates. These additional approaches ranged from using exposure duration without regard to exposure intensity, using the JEM alone without incorporating measurements, and using the mixed-effects model framework to calculate calibrated JEM estimates. We also examined the robustness of associations with NHL for more sensitive and specific definitions of benzene exposure based on the probability ratings from the JEM. In this article, we focus on what we can learn about the exposure assessment process and the contributions of the measurements. We refer the reader to Bassig *et al.* (2015) for discussions on the plausibility of the benzene–NHL association and the strengths and limitations of the epidemiologic study design.

Methods

Study population

The study population comprised 73 087 women from the SWHS prospective cohort (*n* = 74 942) who had no prevalent cancer at baseline and who had a valid occupational history (Zheng *et al.*, 2005; Bassig *et al.*, 2015). These women were aged 40–70 years who were living in urban Shanghai during study enrollment between 1996 and 2000 (participation rate 92.7%). The participants completed interviewer-and self-administered questionnaires with components for demographic characteristics, lifestyle habits, medical history, and residential history. In addition, lifetime occupational histories were collected using the self-administered questionnaire for participants up until their year of entry into the cohort. Occupational history data were not available during the follow-up period; however, only ~2% of the overall cohort still had a benzene-exposed job at baseline. The reported jobs and industries were coded to the 1982 Standard Chinese Classification of Industries and Occupations for the third national census. Incident cases of NHL [102 incident cases, International Classification of Diseases, Ninth Revision (ICD-9) codes: 200, 202, 204.0–204.1] and other cancers as well as vital status were identified through annual linkage to Shanghai cancer and vital statistics registries and from in-person follow-up questionnaires conducted every 2–3 years and were coded based on the ICD-9. Follow-up for this study was completed through 2009. The mean age at baseline for women in the benzene analysis was 52 (± 9) years, and the mean attained age at the end of the follow-up period was 64 (± 9) years. All participants provided informed consent, and the study protocols were approved by each participating institution's institutional review board.

Exposure assessment

Annual benzene exposure estimates corresponding to each subject's employment year obtained from the occupational histories were calculated using several approaches. The foundation for all estimates was a benzene JEM that comprised two separate JEMs, one based on occupation (Occ-JEM) and one based on industry (Ind-JEM), that were developed for this cohort by study industrial hygienists (Friesen *et al.*, 2012). For each standardized occupation and standardized industry code in the respective matrices, industrial hygienists assigned a probability (*P*), intensity (*I*), and frequency (*F*) estimate using an ordinal scale from 0 to 3. The probability ratings for both occupation and industry (P_{occ} and P_{ind}, respectively) were estimated as ~0, 1 = >0 to <5%, 2 = 5 to <50%, and 3 = ≥50% of workers exposed. The intensity ratings for occupation and industry (I_{occ} and I_{ind}, respectively) were estimated as the overall long-term average: 0, 1 = >0 to <10%, 2 = 10 to <100%, and 3 = ≥100% of the maximum allowable benzene concentration in 1980 in China of 40 mg m^{-3} (Yin *et al.*, 1987). The frequency ratings for occupation and industry (F_{occ} and F_{ind}, respectively) were estimated based on the average percent of the work shift exposed and were classified as 0, 1 = >0 to 10%, 2 = 10 to <50%, and 3 = ≥50% of work shift. These estimates did not account for changes over time.

Two sets of measurement-based benzene estimates that accounted for changes in exposure across time were derived from a previously described mixed-effects model (Friesen *et al.*, 2012). The model combined the intensity ratings from the independently developed Occ- and Ind-JEMs with over 60 000 short-term area benzene inspection measurements collected between 1954 and 2000 during inspections of Shanghai factories. The model terms included natural log-transformed benzene concentrations as the response variable, I_{occ}, I_{ind}, and a b-spline time trend as fixed-effects terms, and occupation group and industry group nested within occupation as random-effects terms. From the model, we calculated annual 'job/industry-specific estimates', which were used in Bassig *et al.* (2015) to calculate cumulative exposure, using the model parameter estimates from the fixed-effects terms and the BLUP estimates from the occupation and industry random-effects terms. We also calculated annual 'calibrated JEM estimates' ($JEM_{Calibrated}$) using only the fixed-effects terms.

A more typical approach is to use the JEM ratings alone, without incorporating the measurement data; we refer to these estimates as 'uncalibrated JEM' estimates. Because the above-mentioned benzene JEM was based on separate Occ- and Ind-JEMs, we first combined the two sets of estimates for intensity [equation (1)] and frequency [equation (2)]. These equations modified those originally described by Dosemeci *et al.* (1989), who proposed a variety of calculations that depended on the occupation and the probability ratings, ranging from using the occupation rating only to using the occupation and industry rating as independent contributors of exposure. Here, for simplicity and a single treatment for all jobs, we assumed the primary driver of benzene exposure was the occupation intensity estimate and that industry was a modifier of the occupation estimate rather than an independent contributor, which was based on both expert judgment and our observations in developing the calibrated JEM. We then calculated two 'uncalibrated JEM estimates'. 'Uncalibrated JEM without frequency estimates' (JEM_{NoFreq}) were equivalent to the combined intensity estimates [equation (1)]. 'Uncalibrated JEM with frequency estimates' (JEM_{Freq}) were calculated by multiplying together the combined frequency and intensity estimates to derive a time-weighted average [equation (3)]. P_{occ} and P_{ind} were used only in the exposure definitions (described below).

$$I_{combined} = I_{occ}\hat{}(I_{ind}/3) = JEM_{NoFreq} \quad (1)$$

$$F_{combined} = F_{occ}\hat{}(F_{ind}/3) \quad (2)$$

$$JEM_{Freq} = F_{combined} \times I_{combined} / 3 \quad (3)$$

The above approaches were used to assign annual benzene estimates only when the subject's job met a specified exposure definition that was based on the JEM probability ratings P_{occ} and P_{ind}; when the criteria were not met, the benzene estimate for that job was assigned '0'. We examined three exposure definitions—primary, lenient, and strict—that were based on the combinations of P_{occ} and P_{ind} shown in Fig. 1. The primary definition was used in our prior evaluations (Friesen *et al.*, 2012; Bassig *et al.*, 2015). Exposure distributions for this metric according to the most prevalent exposed industries are presented in Supplemental Table 1 at *Annals of Work Exposures and Health* online.

Cumulative exposure estimates for each participant up to study enrollment were calculated for the primary exposure definition for exposure duration, the uncalibrated and calibrated JEM estimates, and the job/industry-specific estimates by summing the annual exposure estimates across the full occupational history. For the lenient and strict definitions, cumulative exposure estimates were calculated only for exposure duration and the job/industry-specific estimates.

Statistical analyses

Cox proportional hazard models with age as the time scale were used to estimate the HR and 95% CI for NHL risk for each cumulative benzene metric using SAS version 9.3 (Cary, NC, USA). Entry was age at enrollment; exit was the earliest of age of NHL diagnosis, age at death, or age at last follow-up. All models were adjusted for education, body mass index, ever smoking, and ever use of alcohol, and were stratified on birth cohort (in 5-year intervals) to adjust for calendar effects.

Occupation probability rating	Industry probability rating			
	0: 0% workers exposed	1: >0-<5% of workers exposed	2: 5-<50% of workers exposed	3: ≥50% of workers exposed
0: 0% workers exposed	Unexposed (44.2%)			
1: >0-<5% of workers exposed	Unexposed (35.9%)			Lenient, Primary (1.3%)
2: 5-<50% of workers exposed	Unexposed (2.9%)		Lenient (8.3%)	Lenient, Primary, Strict (1.8%)
3: ≥50% of workers exposed	Lenient, Primary (0.1%)	Lenient, Primary (1.4%)	Lenient, Primary, Strict (3.2%)	Lenient, Primary, Strict (0.7%)

Figure 1. Three exposure definitions used to apply non-zero benzene estimates to the participants' jobs based on the probability ratings from the occupation- and industry-exposure matrices, with proportion of employed person-years. Lenient definition: all shaded cells; primary definition: medium and dark gray cells; strict definition: dark gray cells only.

Each cumulative benzene metric was examined in three ways: (i) as a categorical variable based on exposure tertiles (cut points 0, >0 to 33rd, >33rd to ≤67th, and >67th percentile); (ii) as a continuous metric incorporated as a linear term; and (iii) as a continuous term using penalized splines, where the optimal degree of smoothing was chosen based on the Akaike information criterion in R (R, version 3.0.1; R Development Core Team, Vienna, Austria) (Eisen *et al.*, 2004). Trends across exposure categories were examined by testing a linear continuous variable with each category assigned its median cumulative exposure. We present only unlagged cumulative exposure metrics because previous analyses showed that the 5- and 10-year lags provided similar magnitudes of risks and significant trends (Bassig *et al.*, 2015).

We calculated the agreement between the job/industry-specific estimate based on the primary exposure definition as our reference metric and alternative metrics with varying exposure definitions using Stata/SE version 11.2 (StataCorp LP, College Station, TX, USA). The agreement between continuous metrics was examined using Pearson correlation (ρ), partial Pearson correlation (which accounted for the common component of exposure duration in each metric, ρ_p), and Spearman correlation (rho). The agreement between cumulative metrics categorized based on an ordinal four-category scale (0–3) was examined using the proportion agreement, kappa (κ), and weighted kappa based on quadratic weights (κ_w).

The impact of each cumulative metric on exposure–response relationships was evaluated by examining the improvement in the −2 log likelihood model fit statistic between a Cox regression model with only covariates and the same model that also included the exposure metric. For categorical metrics, the *P*-value for trends across categories was also compared. To facilitate comparisons across metrics on different exposure scales, we report the HR and CIs for the interquartile range (IQR; 75th percentile–25th percentile) from analyses of continuous metrics.

Results

Correlation of exposure metrics

The exposure definition varied the proportion of benzene-exposed participants from 10.2% for the stricter definition (14 exposed cases) to 14.8% for the primary definition (24 exposed cases) and 27.8% for the more lenient definition (33 exposed cases) (Table 1).

Compared to the cumulative job/industry-specific estimates using the primary exposure definition, the $JEM_{Calibrated}$ estimates were the most similar (ρ = 0.89, rho_{all} = 0.999; $rho_{exposed}$ = 0.87, κ_w = 0.96) (Table 1). However, moderately high to very high measures of agreement were also observed for the two uncalibrated JEM metrics (ρ = 0.76–0.80, $rho_{all} \geq 0.998$, $rho_{exposed}$ = 0.73–0.74, κ = 0.79; κ_w = 0.94). Similarly, partial correlations adjusted for exposure duration remained high for the $JEM_{Calibrated}$ estimates (ρ_p = 0.84) and moderately high for the uncalibrated JEM_{Freq} (ρ_p = 0.63) and JEM_{NoFreq} estimates (ρ_p = 0.67). Exposure duration was also moderately correlated with the job/industry-specific estimates (ρ = 0.61, rho_{all} = 0.997; $rho_{exposed}$ = 0.64, κ_w = 0.93). Agreement was moderate to moderately high for the metrics based on the strict (rho = 0.81, κ_w = 0.72–0.78) and lenient exposure definitions (rho = 0.67–0.70, κ_w = 0.59–0.67). Moderately high to excellent agreement was also observed when the cumulative metrics were categorized for all subjects and for exposed subjects (Table 1).

Benzene–NHL risk

The mean age at diagnosis for the NHL cases was 65 years and the length of time between first year of exposure to benzene and NHL diagnosis was on average 38 years (range, 13–59 years). Statistically significant exposure–response trends with NHL risk were observed for all categorical and continuous metrics based on the primary exposure definition (Table 2). For the primary definition, HRs for the highest tertile ranged from 1.8 to 2.5, trend *P*-values ranged from 0.002 to 0.02, and HRs ranged from 1.1 to 1.5 for a continuous exposure equal to the IQR. For the categorical metrics, the best model fit occurred for the job/industry-specific estimates (change in model fit: 8.5 units) and the strongest trend was observed for the JEM_{Freq} estimates (trend *P*-value = 0.002). For the continuous metrics, the best model fit occurred for the JEM_{NoFreq} estimates (change: 7.9 units), the highest HR occurred with exposure duration (HR = 1.52), and the narrowest CI occurred for the JEM_{NoFreq} estimates. When one considers that the categorical metrics required two additional degrees of freedom (i.e. approximately subtracting two from the model fit), the exposure–response relationships were stronger using the continuous metrics for the JEM_{Freq}, JEM_{NoFreq}, and duration metrics and stronger using the categorical metrics for the job/industry-specific estimates and the $JEM_{Calibrated}$ estimates. After this crude adjustment, the model fit was highest for the continuous JEM_{NoFreq} estimates. The same patterns were observed in comparisons based on the Akaike information criterion (data not shown).

Using penalized smoothing splines to examine the shape of the exposure–response relationship for the

Table 1. Comparison of the job/industry-specific cumulative benzene estimates using our *a priori* exposure definition and alternative cumulative metrics and exposure definitions.

	Continuous cumulative metrics[a]						Categorical cumulative metrics[b]			
	Distribution of cumulative estimates for exposed participants		Comparison to *a priori* job/industry-specific estimates				Comparison to *a priori* job/industry-specific estimates, all subjects			Comparison to job/industry estimates, exposed subjects
Exposure definition/assessment method	Median (10th–90th), all exposed	Median (10th–90th), exposed cases	Pearson correlation	Partial correlation, adjusted for duration	Spearman correlation	Spearman correlation, exposed subjects	% agreement	Kappa	Weighted kappa	Weighted kappa
Primary (14.8% ever exposed)										
Job/industry estimates	59 (12–273)	103 (21–425)	Ref				Ref.			Ref.[c]
Calibrated EM estimates	70 (17–220)	102 (32–418)	0.89	0.84	0.999	0.87	96.0	0.85	0.96	0.77
Uncalibrated EM without frequency[d]	26 (6–54)	30 (14–90)	0.76	0.67	0.998	0.74	94.4	0.79	0.94	0.62
Uncalibrated EM with frequency[e]	15 (3–42)	20 (7–90)	0.80	0.63	0.998	0.73	94.4	0.79	0.94	0.64
Duration	17 (4–30)	19 (7–32)	0.61	n/a	0.997	0.64	93.5	0.76	0.93	0.56
Lenient (27.8% ever exposed)										
Job/industry estimates	54 (11–221)	82 (24–344)	0.93	0.89	0.70	n/a	85.2	0.60	0.67	Ref.[c]
Duration	18 (5–30)	19 (6–31)	0.44	n/a	0.67	n/a	80.1	0.46	0.59	0.57
Strict (10.2% ever exposed)										
Job/industry estimates	53 (10–279)	93 (12–657)	0.88	0.84	0.81	n/a	94.5	0.76	0.78	Ref.[c]
Duration	15 (4–30)	17 (6–34)	0.54	n/a	0.81	n/a	90.5	0.59	0.72	0.52

EM, exposure matrix (combines occupation- and industry-specific exposure matrices, see the text for details on the calculation); n/a, not applicable; ref., reference.

[a]The job/industry and calibrated EM estimates units were mg m^{-3} years; uncalibrated EM estimates were EM unit years; and the duration units were years.

[b]Categorical metrics are defined as 0 (ref.), >0 to ≤33rd percentile, >33 to ≤67th percentile, and >67th percentile of specified cumulative exposure metric.

[c]Comparison group is the job/industry-specific estimates for the specified exposure definition group.

[d]Uncalibrated EM estimate incorporates the intensity metrics, but not the frequency metric. The probability metric was used in the exposure definition.

[e]Uncalibrated EM estimate incorporates the intensity and frequency metrics. The probability metric was used in the exposure definition.

Table 2. Exposure–response relationships between NHL and various metrics of cumulative benzene exposure.

	Categorical[a]						Continuous		
	Change in −2 log likelihood model fit[b]	HRs (95% CI)					Change in −2 log likelihood model fit[b]	IQR (75th–25th percentile)	HRs per IQR (95% CI)
		Ref.	Exposure T1	Exposure T2	Exposure T3	Trend *P*-values			
Primary exposure definition									
Job/industry estimates	8.5	1	0.92 (0.29–2.94)	2.20 (1.10–4.41)	2.16 (1.17–4.00)	0.006	2.9	110 mg m^{-3} year	1.11 (1.002–1.24)
N (person-years)		78 (686 929)	3 (39 150)	9 (40 337)	12 (38 771)				
Calibrated EM estimates	7.5	1	1.26 (0.46–3.47)	2.48 (1.24–4.99)	1.81 (0.95–3.43)	0.02	5.2	100 mg m^{-3} year	1.25 (1.06–1.48)
N (person-years)		78 (686 929)	4 (39 194)	9 (40 347)	11 (38 717)				
Uncalibrated EM without frequency	7.6	1	1.25 (0.50–3.09)	1.93 (0.93–4.02)	2.33 (1.23–4.41)	0.003	7.9	26 JEM units year	1.47 (1.16–1.85)
N (person-years)		78 (686 929)	5 (39 626)	8 (40 257)	11 (38 374)				
Uncalibrated EM with frequency	7.9	1	1.44 (0.63–3.31)	1.51 (0.65–3.48)	2.54 (1.37–4.70)	0.002	7.1	18 JEM units year	1.33 (1.11–1.59)
N (person-years)		78 (686 929)	6 (40 299)	6 (39 365)	12 (38 593)				
Exposure duration	6.8	1	1.45 (0.63–3.32)	2.09 (1.00–4.36)	2.04 (1.05–3.97)	0.007	6.3	15 years	1.52 (1.13–2.05)
N (person-years)		78 (686 929)	6 (40 816)	8 (38 899)	10 (38 543)				
Lenient exposure definition									
Job/industry estimates	5.0	1	0.63 (0.23–1.75)	1.75 (0.97–3.14)	1.32 (0.74–2.33)	0.22	1.4	88 mg m^{-3} year	1.07 (0.97–1.18)
N (person-years)		69 (582 815)	4 (73 655)	14 (75 891)	15 (72 826)				
Exposure duration	1.5	1	1.12 (0.56–2.26)	1.38 (0.72–2.65)	1.33 (0.73–2.45)	0.24	1.0	15 years	1.16 (0.88–1.54)
N (person-years)		69 (582 815)	9 (76 844)	11 (73 630)	13 (71 898)				
Strict exposure definition									
Job/industry estimates	2.5	1	0.84 (0.21–3.44)	1.89 (0.82–4.34)	1.42 (0.62–3.28)	0.28	2.2	110 mg m^{-3} year	1.11 (0.99–1.25)
N (person-years)		88 (723 312)	2 (27 186)	6 (27 884)	6 (26 805)				
Exposure duration	5.5	1	1.34 (0.49–3.66)	1.22 (0.45–3.35)	1.71 (0.74–3.93)	0.19	1.8	15 years	1.33 (0.91–1.96)
N (person-years)		88 (723 312)	4 (27 485)	4 (29 851)	6 (24 540)				

EM, exposure matrix; ref., reference.

[a]Categorical metrics are defined as 0 (ref.), >0 to ≤33rd percentile (exposure T1), >33 to ≤67th percentile (exposure T2), and >67th percentile (exposure T3) of specified cumulative exposure metric.

[b]Larger values suggest a stronger exposure–disease relationship.

primary exposure definition, we found that a log HR-linear exposure relationship (i.e. equivalent to the analyses using the continuous metrics described above) was an appropriate characterization of the shape for the $JEM_{Calibrated}$, JEM_{NoFreq}, and JEM_{Freq} estimates (Fig. 2). In contrast, a log-linear relationship was rejected for the job/industry-specific estimates, where the rate of increase per unit exposure began to decrease at cumulative estimates >300 mg m^{-3} years. Based on this model, the HR for a continuous exposure equal to the IQR (110 mg m^{-3} years) was 1.44 (95% CI: 1.02, 2.01).

In analyses based on stricter and more lenient exposure definitions, exposure–response relationships were suggestive with HRs in the highest tertile of 1.3–1.7; however, no category or trend was statistically significant (Table 2). In addition, for the job/industry-specific estimates, the HRs for the highest tertiles were lower than that for the second highest tertile for both cumulative exposure and duration. We examined the differences between the strict and primary exposure definitions further by refining the reference group in these analyses by treating the participants who met the primary but not the strict definition as a separate group in the analyses, rather than including them in the unexposed group. This reduced the strict definition trend *P*-value for the job/industry-specific estimates from 0.28 to 0.21 and increased the HR for the highest tertile from 1.4 to 1.5, but no category or trend was statistically significant (not shown). These participants represented 54 115 of the 181 562 (29.8%) exposed person-years meeting the primary definition and they had a statistically significant overall increased NHL risk of 2.58 (95% CI: 1.33, 4.99). For these participants, the most prevalent industries were the rubber products and organic chemicals

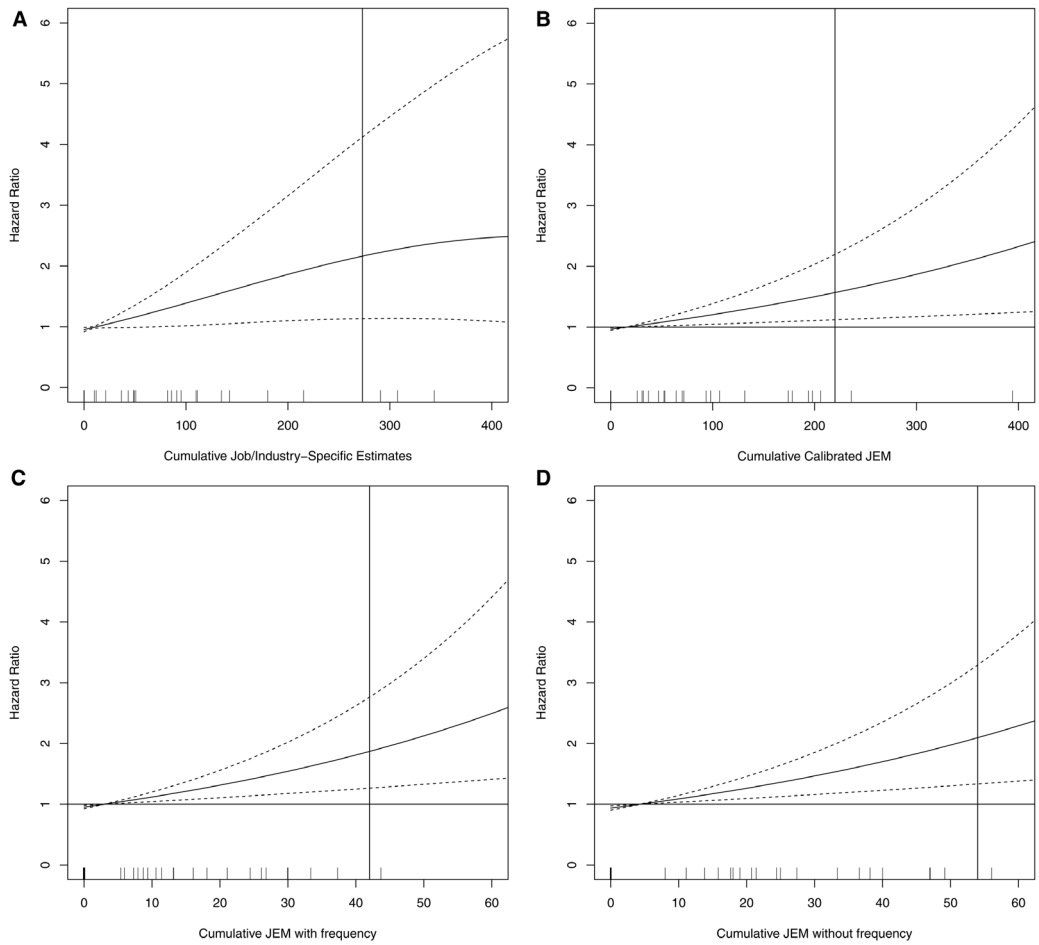

Figure 2. Shape of exposure–response curves using penalized splines for different cumulative benzene metrics based on the primary exposure definition. Solid line = HR; dashed lines = 95% CIs. Vertical line represents the 90th percentile of exposed subjects. Rug lines represent cumulative exposure of NHL cases.

manufacturing industries, and the most prevalent occupation was installers or assemblers of electric and electronic equipment (Table 3).

Discussion

Here, the benzene–NHL associations were robust to varying benzene JEM-based exposure assessment methods, with statistically significant associations observed with all metrics when our primary exposure definition was used. The similarities between approaches suggest that there was little gain in model fit from incorporating exposure measurements and capturing within-JEM differences in exposure. However, incorporating measurements provided quantitative exposure–response curves on a concentration scale, whereas the uncalibrated JEM metrics were dimensionless.

Table 3. Industries and occupations that met the primary, but not the strict, exposure definition.

Industry or occupation	Person-years
Industries with Ind-EM probability = 3, but Occ-EM probability < 2	
341: Rubber products manufacturing	5649
314: Organic chemicals industry	5269
472: Motor vehicles	4707
481: Measuring tools/instruments	2442
252: Leather goods manufacturing	2195
430: Blanks for the forge and foundry	1728
270: Furniture manufacturing	688
243: Shoe manufacturing	610
361: Crude oil processing	351
070: Petroleum and natural gas extraction	127
362: Artificial crude oil production	9
Occupations with Occ-EM probability = 3, but Ind-EM probability < 2	
863: Install/assembly electric/electronic equipment	15 817
964: Lab technicians/analyzers	6295
901: Painters	4003
853: Clocks, watches, precision instrument makers	1655
823: Printing workers	700
869: Other electric/electronic equipment installation	616
861: Install/maintenance electric equipment	457
773: Shoemakers, hat-makers	339
851: Machinery installation and assembly	303
909: Other painters	61
918: Engravers of artistic works	58
743: Tire production/vulcanization workers	19
736: Oil refinery workers	17

Ind-EM, industry-exposure matrix; Occ-EM, occupation-exposure matrix.

The high agreement between metrics and the similarities in the exposure–response associations based on the primary exposure definition was not surprising. These metrics were not independent because each approach incorporated exposure duration and all JEM-based metrics incorporated the same ordinal intensity ratings. As a result, moderately high to high correlations were observed between all estimation approaches. These findings were similar to the strong correlations observed between various silica exposure metrics derived using a similar model framework (Peters *et al.*, 2011a) and consistent with the negligible to modest gains observed in previous sensitivity analyses conducted for moderately highly correlated metrics, such as the various silica estimates in a population-based study of lung cancer (Peters *et al.*, 2012), total chlorophenol and pentachlorophenol metrics in the sawmill industry (Friesen *et al.*, 2007a), and two measures of coal tar pitch volatiles in the aluminum smelter industry (Friesen *et al.*, 2007b). Here, the two measurement-based cumulative metrics were highly correlated with each other. The $JEM_{Calibrated}$ estimates had better model fit in linear continuous analyses, and the job/industry-specific estimates had better model fit in categorical analyses and higher HRs but wider CIs in analyses based on penalized splines. These differences suggest that the additional refinement to calculate job/industry-specific estimates may have introduced some exposure misclassification, likely because of the limited number of measurements for most jobs and industries. Similarly, the uncalibrated JEM_{NoFreq} and JEM_{Freq} estimates were highly correlated with each other ($\rho_p = 0.94$, $\kappa_w = 0.98$) and demonstrated negligible gain from incorporating frequency. The JEM_{Freq} estimates had slightly better model fit and exposure trends than the JEM_{NoFreq} estimates in categorical analyses (trend *P*-value 0.002 versus 0.003, respectively); the opposite was found in continuous analyses.

There are two features of the exposure assessment process that may explain these striking similarities in model fit for measurement-based estimates and the uncalibrated JEM estimates, despite the fact the latter were unitless and did not capture across-rating or time-specific differences in exposure concentration. First, the ordinal Occ-JEM intensity ratings had weights of 0.32, 0.65, and 1.0 predicted by a mixed-effects model (Friesen *et al.*, 2012), which corresponds to weights of 1.0, 2.0, and 3.1 when divided by the lowest value, indicating that the ordinal JEM ratings of 1, 2, and 3 nearly perfectly represented the measured differences between

the rating categories in this study. Greater differences between the cumulative estimates would likely have been observed if the measurement-based weights between ratings had not paralleled the weights applied to the uncalibrated JEM ratings. Second, participants in the highest tertile of any cumulative metric were generally those with the longest exposure duration, which would include the time periods with the highest exposure concentrations. As a result, accounting for the exposure time trend in the measurement-based estimates—and ignoring it in the uncalibrated JEM estimates—had less influence on the cumulative exposure estimates than might be expected for a 13-fold exposure decline over four decades. This also accounts for the strong benzene–NHL associations observed here with duration of exposure.

The benzene–NHL associations varied by exposure definition. Associations were inconclusive using a more lenient exposure definition. This is consistent with previous findings that sensitive, rather than specific, exposure definitions can result in attenuated exposure–response associations when unexposed participants are characterized as exposed (Dosemeci and Stewart, 1996; Kromhout and Vermeulen, 2001). Contrary to expectations, associations were also weak when using a strict definition. This suggests that either the power to detect a difference was too limited compared to the primary definition (14 versus 24 exposed cases) or, more likely, that the JEM assignments of probability may have been incorrect for some occupations or industries. The latter is supported by the finding that the participants meeting the primary but not strict definition had a statistically significantly elevated NHL risk of 2.6 (reported in Results section), which suggests that important benzene exposure scenarios were likely missed in the strict definition. While our previous efforts used the exposure database to calibrate the JEM intensity estimates, it may be possible in future work to use the database to also improve the JEM probability estimates.

The strengths of this study included the large database of benzene measurements spanning over four decades that was used to calibrate the JEM and develop job/industry-specific estimates, combined with over 100 incident NHL cases from a prospective population-based cohort study, which allowed us to evaluate the robustness of exposure–response relationships to varying methods of assigning exposure. Limitations included the inability to examine associations by NHL subtype, the lack of a gold standard with which to compare the varying exposure assessment approaches, and lack of subject-specific information on work activities beyond broad occupation and industry group. As a result, we were unable to capture the within-job exposure variability experienced by subjects performing different work tasks or under different workplace conditions within the same occupation group. The measurement data used here has many limitations, which were previously reported (Friesen *et al.*, 2012; Koh *et al.*, 2014). Briefly, the measurements were short-term, area measurements, which can both overestimate and underestimate personal exposure concentrations depending on the proximity of the sampler to the worker and the variability in work activities. The monitored facilities included in the Shanghai exposure database may not be representative of the facilities in which the study subjects worked. The predicted concentrations were lower than reported in a multi-industry benzene cohort study (Dosemeci *et al.*, 1994) or in other studies where measurements were collected during surveys investigating benzene-poisoning events or in workplaces targeted because of known high exposure levels (Friesen *et al.*, 2012). We were also unable to account for differences in exposure due to changes in sampling and analytical methods; however, prior sensitivity analyses found nearly identical rank order estimates when adjustments for sampling and analytical methods were made (Friesen *et al.*, 2012). Because of the lack of gold standards, the risk per unit dose estimates should be interpreted cautiously. In addition, there is no standard approach for combining occupation and industry intensity ratings and thus used both expert judgment and empirical evidence from the available benzene measurements to derive our conceptual framework for calculating the uncalibrated JEM estimates. Finally, as has been discussed previously (Bassig *et al.*, 2015), we lacked data on potential co-exposures in the workplace such as other types of solvents that have been suggestively associated with NHL risk. However, this limitation may be mitigated by the fact that the SWHS includes participants with an employment history consisting of diverse industries and job tasks. Therefore, the presence of a common confounder across industries that would influence the NHL association to the extent of the observed magnitude in our population-based cohort may be less likely compared to an occupational cohort that includes limited diversity in employment histories.

In this study, the gains from including measurements were small, and the benzene–NHL associations were robust to varying benzene metrics, providing support for the ability of JEM-based approaches to identify subjects with broad contrasts in exposure that can be useful in identifying exposure–disease associations. However, using the measurements provided a mechanism with which to obtain quantitative exposure–response curves. In this study, the lack of improvement in the exposure–response relationship by using measurements rather

than using the JEM on its own may be an artifact of the study-specific features discussed above. Thus, the influence of incorporating exposure measurements into JEMs on exposure–response associations warrants additional evaluations in other settings and for other exposures.

Supplementary Data

Supplementary data are available at *Annals of Work Exposures and Health* online.

Acknowledgements

This study was supported by the US National Institutes of Health (grant R37 CA70867) and the Intramural Research Program of the Division of Cancer Epidemiology and Genetics of the National Cancer Institute within the National Institutes of Health (contract N02 CP1101066; Z01 CP10122-19). The authors have no competing financial interests to declare.

References

Bassig BA, Friesen MC, Vermeulen R *et al*. (2015) Occupational exposure to benzene and risk of non-Hodgkin lymphoma in a population-based cohort study of Chinese women in Shanghai. *Environ Health Perspect*; **123**: 971–7.

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**: 256–9.

Eisen EA, Agalliu I, Thurston SW *et al*. (2004) Smoothing in occupational cohort studies: an illustration based on penalised splines. *Occup Environ Med*; **61**: 854–60.

Friesen MC, Coble JB, Lu W *et al*. (2012) Combining a job-exposure matrix with exposure measurements to assess occupational exposure to benzene in a population cohort in shanghai, china. *Ann Occup Hyg*; **56**: 80–91.

Friesen MC, Davies HW, Teschke K *et al*. (2007a) Impact of the specificity of the exposure metric on exposure-response relationships. *Epidemiology*; **18**: 88–94.

Friesen MC, Demers PA, Spinelli JJ *et al*. (2007b) Comparison of two indices of exposure to polycyclic aromatic hydrocarbons in a retrospective aluminium smelter cohort. *Occup Environ Med*; **64**: 273–8.

Heederik D, Attfield M. (2000) Characterization of dust exposure for the study of chronic occupational lung disease: a comparison of different exposure assessment strategies. *Am J Epidemiol*; **151**: 982–90.

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

Koh DH, Bhatti P, Coble JB *et al*. (2014) Calibrating a population-based job-exposure matrix using inspection measurements to estimate historical occupational exposure to lead for a population-based cohort in Shanghai, China. *J Expo Sci Environ Epidemiol*; **24**: 9–16.

Kromhout H, Loomis DP, Kleckner RC. (1999) Uncertainty in the relation between exposure to magnetic fields and brain cancer due to assessment and assignment of exposure and analytical methods in dose-response modeling. *Ann N Y Acad Sci*; **895**: 141–55.

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

Loomis D, Salvan A, Kromhout H *et al*. (1999) Selecting indices of occupational exposure for epidemiologic studies. *Occup Hyg*; **5**: 73–91.

Peters S, Vermeulen R, Cassidy A *et al*. (2011a) Comparison of exposure assessment methods for occupational carcinogens in a multi-centre lung cancer case-control study. *Occup Environ Med*; **68**: 148–53.

Peters S, Vermeulen R, Portengen L *et al*. (2011b) Modelling of occupational respirable crystalline silica exposure for quantitative exposure assessment in community-based case-control studies. *J Environ Monit*; **13**: 3262–8.

Peters S, Vermeulen R, Portengen L *et al*. (2012) Quantitative exposure assessment in community-based studies: the case for respirable crystalline silica and lung cancer. In Peters S, editor. *Quantitative exposure assessment in community-based studies*. Utrecht, The Netherlands: Thesis, Utrecht University. pp. 101–114.

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.

Vincent R, Jeandel B. (2001) COLCHIC-occupational exposure to chemical agents database: current content and development perspectives. *Appl Occup Environ Hyg*; **16**: 115–21.

Yin SN, Li Q, Liu Y *et al*. (1987) Occupational exposure to benzene in China. *Br J Ind Med*; **44**: 192–5.

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.