



ENVIRONMENTAL HEALTH PERSPECTIVES

<http://www.ehponline.org>

Analysis of Environmental Chemical Mixtures and Non-Hodgkin Lymphoma Risk in the NCI-SEER NHL Study

**Jenna Czarnota, Chris Gennings, Joanne S. Colt,
Anneclaire J. De Roos, James R. Cerhan,
Richard K. Severson, Patricia Hartge, Mary H. Ward,
and David C. Wheeler**

<http://dx.doi.org/10.1289/ehp.1408630>

Received: 30 April 2014

Accepted: 4 March 2015

Advance Publication: 6 March 2015

This article will be available in its final, 508-conformant form 2–4 months after Advance Publication. If you need assistance accessing this article before then, please contact ehp508@niehs.nih.gov. Our staff will work with you to assess and meet your accessibility needs within 3 working days.



Analysis of Environmental Chemical Mixtures and Non-Hodgkin Lymphoma Risk in the NCI-SEER NHL Study

Jenna Czarnota,¹ Chris Gennings,¹ Joanne S. Colt,² Anneclaire J. De Roos,³ James R. Cerhan,⁴
Richard K. Severson,⁵ Patricia Hartge,⁶ Mary H. Ward,² and David C. Wheeler¹

¹Department of Biostatistics, School of Medicine, Virginia Commonwealth University,
Richmond, Virginia, USA; ²Occupational and Environmental Epidemiology Branch, Division of
Cancer Epidemiology and Genetics (DCEG), National Cancer Institute (NCI), National Institutes
of Health (NIH), Department of Health and Human Services (DHHS), Bethesda, Maryland,
USA; ³Department of Environmental and Occupational Health, Drexel University School of
Public Health, Philadelphia, Pennsylvania, USA; ⁴Division of Epidemiology, Mayo Clinic,
Rochester, Minnesota, USA; ⁵Department of Family Medicine and Public Health Sciences,
Wayne State University, Detroit, Michigan, USA; ⁶Epidemiology and Biostatistics Program,
DCEG, NCI, NIH, DHHS, Bethesda, Maryland, USA

Address correspondence to David C. Wheeler, Department of Biostatistics, School of
Medicine, Virginia Commonwealth University, One Capitol Square, 830 East Main Street,
Richmond, Virginia 23298-0032 USA. Telephone: (804) 828-9827. E-mail: dcwheeler@vcu.edu

Running title: Environmental chemical mixtures and NHL risk

Acknowledgments: The authors (JC) gratefully acknowledge support from the National Institute
of Environmental Health Sciences grant (#T32 ES0007334). This research was partially
supported by the Intramural Research Program of the National Cancer Institute, National

Institutes of Health, and Public Health Service (contracts N01-PC-65064, N01-PC-67008, N01-PC-67009, N01-PC-67010, N02-PC-71105).

Competing financial interests: The authors declare they have no actual or potential competing financial interests.

Abstract

Background: There are several suspected environmental risk factors for non-Hodgkin lymphoma (NHL). The associations between NHL and environmental chemical exposures have typically been evaluated for individual chemicals (i.e., one-by-one).

Objectives: We determined the association between a mixture of 27 correlated chemicals measured in house dust and NHL risk.

Methods: We conducted a population-based case-control study of NHL in four National Cancer Institute - Surveillance, Epidemiology, and End Results centers: Detroit, Iowa, Los Angeles, and Seattle from 1998-2000. We used weighted quantile sum (WQS) regression to model the association of a mixture of chemicals and risk of NHL. The WQS index was a sum of weighted quartiles for 5 polychlorinated biphenyls (PCBs), 7 polycyclic aromatic hydrocarbons (PAHs), and 15 pesticides. We estimated chemical mixture weights and effects for study sites combined and for each site individually, and also for histologic subtypes of NHL.

Results: The WQS index was statistically significantly associated with NHL overall (odds ratio [OR]=1.30; 95% CI: 1.08, 1.56; p=0.006; for one quartile increase) and in the study sites of Detroit (OR=1.71; 95% CI: 1.02, 2.92; p=0.045), Los Angeles (OR=1.44; 95% CI: 1.00, 2.08; p=0.049), and Iowa (OR=1.76; 95% CI: 1.23, 2.53; p=0.002). The index was marginally statistically significant in Seattle (OR=1.39; 95% CI: 0.97, 1.99; p=0.071). The most highly-weighted chemicals for predicting risk overall were PCB congener 180 and propoxur. Highly-weighted chemicals varied by study site; PCBs were more highly-weighted in Detroit and pesticides were more highly-weighted in Iowa.

Conclusions: An index of chemical mixtures was significantly associated with NHL. Our results show the importance of evaluating chemical mixtures when studying cancer risk.

Introduction

Risk of non-Hodgkin lymphoma (NHL) is suspected to be associated with several chemicals through occupational or environmental routes of exposure; geographic variation in NHL rates further suggests the importance of environmental risk factors (Hartge et al. 2006). Positive associations have been found with persistent organochlorine chemicals, including polychlorinated biphenyls (PCBs) (Engel et al. 2007a), particularly PCB congener 180 (Colt et al. 2005; De Roos et al. 2005; Morton et al. 2008), and dichlorodiphenyldichloroethylene (DDE) (Colt et al. 2005; Engel et al. 2007a). An association between NHL overall (Colt et al. 2006) and certain NHL subtypes (Morton et al. 2008) has also been found for residential termite treatment before 1988 (a surrogate for the insecticide chlordane). Several studies have found higher risk of NHL among persons living in areas with industrial emissions to air or industrial waste exposure (Bithell et al. 1994; De Roos et al. 2010; Dreiherr et al. 2005; Floret et al. 2003; Franchini et al. 2004; Goldberg et al. 1999; Pronk et al. 2013).

Existing studies of environmental chemical exposures and NHL generally considered only single-chemical risk or total exposure within specific chemical groups, such as PCBs (Colt et al. 2005; De Roos et al. 2005), but did not consider the effects of simultaneous exposure to multiple diverse chemicals or environmental risk factors. Because individuals are exposed to many chemicals simultaneously, it is of particular importance to examine the relationship between chemical mixtures and disease risk. Additionally, the analysis of multiple chemical exposures must also consider the inherent correlations among co-occurring environmental chemicals. The complex correlation pattern among chemical exposures and subsequent issue of collinearity has not been directly addressed in studies of NHL or other diseases.

Some studies of environmental factors and disease risk consider many exposures (Everett et al. 2008), sometimes controlling for multiple comparisons in so-called environment-wide association studies (Patel et al. 2010; Patel et al. 2012), but they use separate regression models for each environmental exposure. This type of analysis ignores that environmental exposures may interact (Engel et al. 2007b; Porta et al. 2012). Studies also examine pairwise correlation coefficients between environmental factors (Ioannidis et al. 2009; Patel et al. 2012), but most do not account for the correlation among factors in statistical models. The lack of statistical independence observed among exposures presents challenges to assessing many exposure effects simultaneously in one traditional regression model.

Here, we present an application of the weighted quantile sum (WQS) regression method (Carrico et al. 2014) to estimate an index for 27 correlated environmental chemicals measured in residential carpet dust in a case-control study of NHL. Estimation of chemical weights and the resulting WQS index while considering the correlation between compounds allows us to make generalized inference about the mixture effect and identify the individual chemicals ('bad actors') most strongly associated with NHL. Due to the design of our study in four geographic regions, the analysis took a site-specific approach in a preliminary effort to consider the effects of spatially varying levels of exposures among chemical mixtures.

Methods

Study population

We conducted a population-based case-control study of NHL in four National Cancer Institute-Surveillance Epidemiology and End Results Program (NCI-SEER) study sites (<http://seer.cancer.gov/>). The study design has been previously described (Colt et al. 2004; Wheeler et al. 2011). Briefly, the study was conducted in Iowa, Los Angeles County, and the

metropolitan areas of Detroit (Macomb, Oakland, and Wayne counties) and Seattle (King and Snohomish counties). Eligible cases were aged 20-74 years, diagnosed with a first primary NHL between July 1998 and June 2000, and uninfected with HIV. In Seattle and Iowa, all consecutive cases were chosen. In Detroit and Los Angeles, all African American cases and a random sample of white (regardless of Hispanic ethnicity) cases were eligible for study, allowing for oversampling of African American cases. Of the 2,248 potentially eligible cases, 320 (14%) died before they could be interviewed, 127 (6%) were not located, 16 (1%) had moved away, and 57 (3%) had physician refusals. Of the 1,728 remaining cases, 1,321 (76%) participated. Controls (\geq 65 years) were selected from Center for Medicare and Medicaid Services files (<http://dnav.cms.gov/>) or the general population using random digit dialing ($<$ 65 years) and were frequency matched to cases by sex, age (within 5-year groups), race, and study site. Of the 2,409 potentially eligible controls, 2,046 were able to be located and contacted, and 1,057 (52%) of these subjects participated. The study was approved by the human subjects review boards at all participating institutions. Written informed consent was obtained from each participant.

Computer-assisted personal interviews were conducted in the home of each participant. Interviewers asked about demographics including race and education, age of the home, housing type, the presence of oriental rugs, pesticide use in the home and garden, residential and occupational histories, and other factors.

Dust samples and laboratory analysis

As described in detail (Colt et al. 2004; Colt et al. 2005), dust was collected between February 1999 and May 2001 from vacuum cleaners of participants who gave permission (93% of cases, 95% of controls) and who had used their vacuum cleaner within the past year and owned at least half their carpets or rugs for 5 years or more (695 cases (57%), 521 controls (52%). Dust samples

from 682 cases (98%) and 513 controls (98%) were successfully analyzed between September 1999 and September 2001.

Exposure to a mixture of 27 chemicals measured in house dust (5 PCBs, 7 polycyclic aromatic hydrocarbons (PAHs), and 15 pesticides) was of interest. The PCBs were congeners 105, 138, 153, 170, and 180. The PAHs were benz(a)anthracene, benzo(a)pyrene, benzo(b)fluoranthene, benzo(k)fluoranthene, chrysene, dibenz(ah)anthracene, and indeno(1,2,3-*cd*)pyrene. The pesticides were α -chlordane, γ -chlordane, carbaryl, chlorpyrifos, *cis*-permethrin, *trans*-permethrin, 2,4-dichlorophenoxyacetic acid (2,4-D), DDE, dichlorodiphenyltrichloroethane (DDT), diazinon, dicamba, methoxychlor, *o*-phenylphenol, pentachlorophenol, and propoxur. Extraction and analysis were performed on 2-g aliquots of dust samples using gas chromatography/mass spectrometry (GC/MS) in selected ion monitoring mode. Concentrations were quantified using the internal standard method. Usual detection limits were: 20.8 ng/g of dust for α -chlordane, γ -chlordane, DDE, DDT, propoxur, *o*-phenylphenol, PAHs, and PCBs; 42-84 ng/g for chlorpyrifos, diazinon, *cis*-permethrin, dicamba, pentachlorophenol, and 2,4-D; and 121-123 ng/g for carbaryl and *trans*-permethrin. Changes in analytic procedures during the study resulted in increased detection limits for methoxychlor (from 20.7 to 62.5 ng/g). A small proportion of samples weighing less than 2 g had detection limits that were higher than the usual detection limits.

The laboratory measurements for the 27 analytes contained various types of “missing data,” primarily when the concentration was below the minimum detection level. To a lesser extent, missing data occurred when there was co-elution between the target chemical and interfering compounds. Chemical concentrations were assumed to follow a log-normal distribution, and data were imputed using a “fill-in” approach to create ten complete datasets for each of the 27

analytes. Details about the imputation of analyte values have been published previously (Colt et al. 2004; Lubin et al. 2004).

A total of $N = 1180$ subjects with complete dust analysis results and covariate values were included in this analysis. The sample included 508 (43%) controls and 672 (57%) cases, and was predominantly white (88%) with an average age of 60 years (standard deviation = 11.2). Of these 1180 subjects, 202 (17%) were from the Detroit study site, 340 (29%) from Iowa, 292 (25%) from Los Angeles, and 346 (29%) from Seattle.

Statistical analysis

In previous analyses of individual chemicals in the study population overall, we evaluated NHL risk comparing tertiles or other groupings of levels above the detection limit to those with no detectable level of the chemical (Colt et al. 2005; Colt et al. 2006; Hartge et al. 2005). Study site specific risk estimates were not presented in these publications. Here, we used a weighted quartile sum approach in conjunction with non-linear logistic regression to evaluate the effect of several chemical exposures together on the risk of NHL. Exposure to a mixture of 27 chemicals measured in house dust was evaluated overall and in study site-specific models. All models were adjusted for gender, age at diagnosis (cases)/selection date (controls), race, and level of education. Age was treated as continuous, race was dichotomized as white or non-white, and education was treated as ordinal (grouped as <12, 12-15, and ≥ 16 years). In the overall model, we also adjusted for study site.

The WQS method (Carrico et al. 2014) is constrained to have associations in the same direction for chemical exposures and risk, and is designed for variable selection over prediction. WQS regression estimates a weighted linear index in which the weights are empirically determined

through the use of bootstrap sampling. The approach considers data with c correlated components scored as ordinal variables into quantiles (here, quartiles) that are reasonable to combine (i.e., all chemicals) into an index and potentially have a common adverse outcome. The weights are constrained to sum to 1 and be between 0 and 1, thereby reducing dimensionality and addressing issues associated with collinearity. For this analysis, the $c = 27$ chemical concentrations were scored into quartiles based on the case and control data combined and denoted by q_i , where $q_i = 0, 1, 2, \text{ or } 3$ for $i = 1$ to c . A total of $B = 100$ bootstrap samples (of the same size as the total sample, $N = 1180$) were generated from the full dataset and used to estimate the unknown weights, \mathbf{w} , that maximized the likelihood for $b = 1$ to B for the following model

$$g(\mu) = \beta_0 + \beta_1 \left(\sum_{i=1}^c w_i q_i \right) + \mathbf{z}' \boldsymbol{\phi} \Big|_b \quad [1]$$

subject to the constraints $\sum_{i=1}^c w_i \Big|_b = 1$ and $0 \leq w_i \leq 1$ for $i = 1$ to c . In the above equation, w_i

represents the weight for the i^{th} chemical component q_i and the term $\sum_{i=1}^c w_i q_i$ represents a

weighted index for the set of c chemicals of interest. Furthermore, \mathbf{z} denotes a vector of covariates determined prior to estimation of the weights, $\boldsymbol{\phi}$ are the coefficients for the covariates in \mathbf{z} , and $g(\cdot)$ is any monotonic and differentiable link function that relates the mean, μ , to the predictor variables in the right hand side of the equation. Because the outcome variable of interest in this analysis is binary (case status), a logit link was assumed for g .

For each bootstrap sample, the p -value of β_1 , the parameter estimate for the weighted index, was used to evaluate the statistical significance of the estimated vector of weights ($\alpha = 0.10$). The

weighted quantile score was then estimated as $WQS = \sum_{i=1}^c \bar{w}_i q_i$, where $\bar{w}_i = \frac{1}{n_B} \sum_{j=1}^{n_B} w_{ij}$, and n_B is the number of bootstrap samples in which β_1 was significant. Finally, the significance of the WQS index was determined using the original data set and the model

$$g(\mu) = \beta_0 + \beta_1 WQS + \mathbf{z}'\phi, \quad [2]$$

where $\exp(\beta_1)$ is the odds ratio (OR) associated with a unit (quartile) increase in the weighted sum of exposure quartiles (WQS index).

Weights estimated from the full data set were used to create a WQS index denoted as WQSF. In addition to WQSF, four site-specific indices (denoted as WQSD, WQSI, WQSL, and WQSS) were estimated using data from each site. Differences in the distributions of the chemical concentrations across sites prohibited the use of quantiles based on the full data set in the estimation of site-specific weights; therefore, we used site-specific quartiles based on the combined case control distribution to estimate site-specific indices. The association between the WQS indices and NHL was examined by testing each index within its respective data set, with statistical significance set at $\alpha = 0.05$. The primary statistical analysis was done with one randomly selected imputation dataset. A secondary analysis estimated WQS indices for all ten imputed datasets to assess sensitivity of the results to the data imputation.

We conducted further analyses of major subtypes of NHL: diffuse large B-cell (DLBCL), follicular, small lymphocytic/chronic lymphocytic leukemia/ (SLL/CLL), marginal zone, other lymphomas, and lymphomas where subtype was not specified/unknown (not otherwise specified [NOS]). Our study primarily included SLL rather than CLL (Morton et al. 2008). Other lymphomas consisted of mantle cell lymphoma, lymphoplasmacytic lymphoma, Burkitt lymphoma/leukemia, mycosis fungoides/Sézary syndrome, and peripheral T-cell lymphoma. We

fitted WQS regression models separately for each of these groups to determine if the mixture effect varied by subtype using all 508 controls in each model.

As a comparison to the WQS regression results, we also conducted single chemical analyses (one-by-one) for all of the data (adjusted for study site) and separately within each study site using study-site specific cutpoints based on the distributions among cases and controls combined. Models were adjusted for gender, age, race, and level of education. Odds ratios comparing each of the three highest quartiles to the first quartile of exposure were estimated for each individual chemical. Given the exploratory nature of these analyses, no adjustments were made for multiple comparisons.

Results

Characteristics of the study population are summarized overall and by study site in Table 1. The demographics were similar across the four sites, with the exceptions of race (varying from 73% white in Los Angeles to 99% white in Iowa) and education (individuals with ≥ 16 years of education ranged from 19% in Iowa to 36% in Seattle).

The distribution of the pairwise Spearman correlations of the chemical concentrations was complex (see Supplemental Material, Figure S1), with pairwise correlations ranging from slightly negative ($r = -0.15$) to nearly perfect correlation ($r = 0.99$ for *cis*- and *trans*-permethrin). Of the 351 unique pairwise correlations, 289 were significant ($p < 0.05$). Correlation among the PAHs ranged from 0.87 to 0.96 (all significant), and correlation among the PCBs ranged from 0.69 to 0.91 (all significant). The correlation among pesticides was generally weaker (84% significant) with an interquartile range of 0.06 to 0.26. Median concentrations of PCBs were generally similar across the four sites, although the concentration distributions were more

positively skewed in Detroit and Seattle than in Iowa and Los Angeles (see Supplemental Material, Figure S2). Chemical concentrations for the PAHs and pesticides varied considerably by site. Concentrations for all seven PAHs were elevated (higher than in other locations) in Detroit, while pesticide concentrations were elevated in Iowa (e.g., carbaryl, 2,4-D, methoxychlor, and dicamba) and Los Angeles (e.g., γ -chlordane, *trans*-permethrin, diazinon, and propoxur) (data not shown).

The WQS index for the overall study population was significantly associated with NHL ($p = 0.006$, Table 2). More specifically, a quartile increase in the WQS index resulted in an increase of 1.30 (95% CI: 1.08, 1.56) in the odds of NHL in the overall study population. In the site-specific analyses, odds ratios for a quartile increase in the Detroit (OR = 1.71; 95% CI = 1.02, 2.92), Iowa (OR = 1.76; 95% CI = 1.23, 2.53) and Los Angeles (OR = 1.44; 95% CI = 1.00, 2.08) indices were significantly associated with NHL. The ORs for each of the five WQS indices were generally robust to the analyte imputation (see Supplemental Material, Figure S3). Over the ten imputations, the average OR for a quartile increase in the WQS index for the study population overall was 1.25, while the average site-specific ORs were 1.38 for Detroit, 1.67 for Iowa, 1.61 for Seattle, and 1.45 for Los Angeles. ORs were significant for only one imputation in Detroit, and four of the ten imputations in Seattle.

The estimated chemical weights for each WQS index are shown in Table 3. Note that each weight would be 0.037 if all chemicals in the index received equal weight. The most heavily weighted chemicals in the index for the overall data set were PCB 180 (weight $w = 0.32$), propoxur ($w = 0.17$), DDE ($w = 0.08$), γ -chlordane ($w = 0.08$), and benzo(k)fluoranthene ($w = 0.07$). The weight for PCB 180 was more than eight times the weight expected if all chemicals were equal.

The chemicals most heavily weighted in the index varied by site (Table 3). PCBs were more heavily weighted in the urban study sites of Detroit, Los Angeles, and Seattle, and pesticides were more heavily weighted in Iowa, an agricultural state. PCB 180 was the most heavily weighted chemical in Detroit ($w = 0.18$), followed by PCB 170 ($w = 0.17$) and the organochlorine pesticide methoxychlor ($w = 0.12$). In Los Angeles, the PAH benzo(k)fluoranthene had the highest weight ($w = 0.30$), followed by PCB 153 ($w = 0.12$) and the herbicide 2,4-D ($w = 0.11$). In Seattle, propoxur had the highest weight ($w = 0.16$), followed by PCB 180 ($w = 0.14$) and DDE ($w = 0.14$). The pesticides propoxur ($w = 0.30$), γ -chlordane ($w = 0.12$), DDE ($w = 0.11$), and *o*-phenylphenol ($w = 0.11$) had the highest weights in Iowa. Chemicals that were highly weighted in more than one site included PCB 180 (Detroit and Seattle), propoxur (Iowa and Seattle), and DDE (Iowa and Seattle).

The distributions of the weights for PCB 180, propoxur, and benzo(k)fluoranthene, three highly weighted chemicals, differed greatly across the sites (see Supplemental Material, Figure S4). Although PCB 180 was weighted prominently overall, and in Detroit and Seattle, the distribution of its weights was centered near 0 in both Iowa and Los Angeles. Benzo(k)fluoranthene had a distribution of weights with a median above 0 only in Los Angeles. The weights for propoxur were mostly distributed above 0 in Iowa, but had a median of 0 for Los Angeles. The distributions of weights for propoxur were similar in Seattle and the full study population.

Of the 672 cases of NHL, 31% were classified as DLBCL, 23% as follicular, 10% as SLL/CLL, 9% as marginal zone, 14% as other, and 13% as NOS. The distribution of cases across subtypes was similar for each site (data not shown). WQS regression results by subtype are shown in Table 4. ORs for a one quartile increase in the index were statistically significant for follicular

lymphomas (OR = 1.47; 95% CI = 1.08, 2.00), marginal zone lymphomas (OR = 2.06; 95% CI = 1.25, 3.47), and the subtype other (OR = 2.26; 95% CI = 1.55, 3.34).

For the individual chemical analyses, the OR for the fourth vs. first quartile of exposure for each chemical is listed in Supplemental Material, Table S1 for the overall analyses and in Supplemental Material, Table S2 for the site-specific analyses. For the overall study population, PCB 180 was significantly associated with NHL (OR = 1.55; 95% CI: 1.11, 2.17 for the fourth vs. first quartile). Associations between NHL and the remaining PCBs were positive (ORs \geq 1.20) but not significant (p -values \leq 0.29). Additionally, although non-significant, the ORs for each of the PAHs were less than one. With respect to the pesticides, the highest quartiles of α -chlordane (OR = 1.40; 95% CI: 0.99, 1.98) and γ -chlordane (OR = 1.35; 95% CI: 0.95, 1.92) were positively associated with NHL, while fourth quartile levels of chlorpyrifos (OR = 0.73; 95% CI: 0.52, 1.02), 2,4-D (OR = 0.70; 95% CI: 0.48, 1.03), and dicamba (OR = 0.74; 95% CI: 0.53, 1.04) were inversely associated with NHL.

In the site-specific analyses of individual chemicals (highest vs. lowest quartiles), PCB 180 was significantly associated with NHL in Detroit (OR = 2.87; 95% CI: 1.19, 6.91). In Iowa, associations were positive and significant for α -chlordane (OR = 2.18; 95% CI: 1.15, 4.14), γ -chlordane (OR = 2.25; 95% CI: 1.20, 4.24), DDE (OR = 1.96; 95% CI: 1.05, 3.68), and propoxur (OR = 2.02; 95% CI: 1.09, 3.78); whereas associations were negative and significant for 2,4-D (OR = 0.36; 95% CI: 0.19, 0.68) and dicamba (OR = 0.48; 95% CI: 0.26, 0.90). In Los Angeles, benzo(k)fluoranthene was significantly associated with NHL (OR = 2.05; 95% CI: 1.04, 4.04). In Seattle, no significant positive associations were found, however propoxur (OR = 1.53; 95% CI: 0.82, 2.85), DDE (OR = 1.53; 95% CI: 0.83, 2.84), and PCB 180 (OR = 1.53; 95% CI: 0.82, 2.85) were nominally associated with NHL. Finally, 2,4-D (OR = 0.53; 95% CI: 0.29, 0.97) and

dicamba (OR = 0.41; 95% CI: 0.22, 0.76) were both significantly inversely associated with NHL in Seattle.

Discussion

We used weighted quantile sum regression to model the association of a mix of 27 correlated environmental chemicals measured in house dust and risk of NHL in a case-control study in four study centers. We fitted site-specific WQS models and an overall WQS model. We found evidence of an increased risk of NHL associated with an increase in the quantile of the weighted chemical index in the overall study population and in each of the four study sites. These associations were statistically significant for the study population overall and for three of the four study centers (Iowa, Detroit, and Los Angeles), and marginally significant in Seattle.

The most highly weighted chemicals in the overall WQS index were PCB congener 180, propoxur, γ -chlordane, and DDE. The chemicals most heavily weighted in the site-specific mixture indexes varied by site. Additional chemicals that were relatively highly weighted in site-specific models and associated with an increased risk of NHL included PCB 153 (Los Angeles), PCB 170 (Detroit), 2,4-D (Los Angeles), benzo(k)fluoranthene (Los Angeles), methoxychlor (Detroit), and *o*-phenylphenol (Iowa).

By comparison, in single chemical analyses, only PCB 180 was found to be significantly associated with NHL in the study population overall, although γ -chlordane was marginally significant. In prior analyses of these data using a slightly different approach, γ -chlordane and DDE were both associated with significantly increased risks of NHL (Colt et al. 2005; 2006). PCB 180 was the only chemical significantly associated with NHL in analyses of individual chemicals in Detroit. In Iowa, α - and γ -chlordane, DDE, and propoxur were associated with

significantly increased risk of NHL. Benzo(k)fluoranthene was the only chemical significantly associated with NHL risk in Los Angeles. There were no chemicals significantly associated with increased risk in Seattle.

WQS regression highlighted some chemicals that were not significantly associated with NHL in prior analyses or by our single chemical analysis here. These included propoxur in the overall study population, and methoxychlor, *o*-phenylphenol, 2,4-D, PCB 153, and PCB 170 in individual study sites. These chemicals had weights that were several times greater than the value associated with equal weight for all chemicals. In addition, chemicals with non-negligible weights factor positively into the weighted quantile sum index and hence were part of the exposure term found to be significantly associated with NHL in the study population and in three of the four study sites.

Generally, the results from our individual chemical analyses supported the WQS regression findings. The chemicals associated with a significant increase in risk of NHL by individual chemical analyses were also selected as potential risk factors by WQS regression (received non-negligible weights). These include PCB 180 (Detroit), propoxur (Iowa), benzo(k)fluoranthene (Los Angeles), and γ -chlordane (Iowa). However, WQS was able to place non-negligible weights on several additional chemicals that were not significantly associated with increased risk in individual chemical analyses. When considered individually, many of these chemicals had elevated ORs that may be potentially meaningful, but were not found to be significantly associated with NHL, likely due to a lack of power. For example, single chemical analysis for propoxur in the full study population resulted in an OR of 1.27 (95% CI = 0.90, 1.79; p-value = 0.18) overall. WQS regression selected this chemical as a strong risk factor, giving it a substantial weight (0.17). In Seattle, the single chemical ORs for DDE (OR = 1.53; 95% CI =

0.83, 2.84) and propoxur (OR = 1.53; 95% CI = 0.82, 2.85) were not significant, but these chemicals received WQS weights of 0.14 and 0.16, respectively. Additionally, WQS placed zero or negligible index weights on chemicals that were identified as having inverse associations with NHL in the one-by-one analyses, (e.g., diazinon and dicamba).

Our findings also show that chemicals identified as important based on a site-specific WQS index may not be identified as important in an index derived from the full dataset. Similarly, chemicals identified as important in the index developed from the full dataset may not be identified as important in all site-specific indices. These differences are due in part to different concentration ranges across sites and overall (see Supplemental Material, Figure S2). The differences may also be due to different sources of these chemicals across study sites or differences in correlations with unmeasured exposures or other factors. For example, PAHs in Los Angeles may be correlated with benzene exposure from traffic and sources in Detroit may be different. The results illustrate the importance of estimating site-specific weights when developing the chemical mixture index when multiple sites are simultaneously under study.

The main strengths of this analysis are the evaluation of a broad range of environmental chemicals together in one modeling approach, the estimation of the association of a mixture of chemicals with NHL risk, and the estimation of mixture effects by study site. Previous studies have focused on effects of individual chemical exposures. We have used a statistical method that has been shown to have high specificity and adequate sensitivity in identifying important chemicals in regression models in simulation studies (Carrico et al. 2014). In simulation studies based on pairwise correlations of 11 phthalates in NHANES (2005-2008), WQS regression had greater accuracy in identifying the 7 of 11 truly important chemicals (i.e., chemicals set to be related to the outcome) correctly as the correlation of the exposures and the outcome increased

from that observed (range of 0.03 to 0.08) to 3 times that observed (Carrico et al. 2014). It also showed an improvement in specificity over traditional ordinary regression and popular shrinkage methods (lasso, adaptive lasso and elastic net). WQS regression tends to place negligible weight on components with no correlation with the outcome.

Single chemical analyses are subject to confounding because of the high degree of correlation among chemical exposure concentrations, but WQS regression allows one to consider exposure to several chemicals simultaneously while accounting for collinearity. Estimation of individual chemical weights enables the identification of potentially harmful chemicals while accommodating the complex correlation observed among exposures. Additionally, WQS regression has the advantage of the estimation of a mixture effect and its association with NHL. Because all potentially harmful chemicals (i.e., those receiving non-zero weights) contribute to the estimation of the WQS index, the OR corresponding to the index is interpretable as the increase in risk associated with a quartile increase in the index of the mixture of chemicals, allowing for further inference and insight regarding the potentially harmful effects of exposure to these environmental chemicals.

A limitation of WQS regression is that it cannot identify associations in different directions for the components of the index. However, chemicals in our analysis with inverse associations with NHL in the single chemical analyses were estimated to have negligible weight in the WQS index that had a positive association with NHL. A limitation of the risk analysis is the potential for exposure misclassification from the use of chemical concentrations in house dust as a measure of past exposures. Levels of chemicals in carpet dust do not indicate the source of the chemicals or when they entered the home, and do not reflect dietary ingestion. However, dust sampling has important advantages over questionnaire- and biologically-based approaches. Levels in dust are

unaffected by difficulties or biases in recall of past activities, and by factors that may influence body burdens of chemicals such as age, body mass index, and disease status or treatment (Colt et al. 2005). Further, with the exception of the persistent organochlorine chemicals (e.g., PCBs, chlordane, and DDT), which are no longer in use, biological measures of many chemicals may reflect only recent exposures due to short half-lives in the body. Carpets act as long-term chemical repositories, and hence chemical concentrations in carpet dust may reflect integrated chemical exposure over the time the carpet was in the home, providing potentially more relevant exposure indicators than a biologic measure of recent exposure. Moreover, studies have found positive associations between chemical concentrations in dust samples and biomarkers of exposure including serum and house dust levels of PCBs (Knobeloch et al. 2012; Rudel et al. 2008) and lead (Lanphear et al. 1998), and urinary and house dust levels of chromium (Stern et al. 1998).

Another potential limitation was the large number of nonparticipants. However, previous analyses of spatial variation in NHL risk based on eligible nonparticipants and participants in this study population did not substantially change results based on study participants only (Wheeler et al. 2011). Additionally, analyses of potential bias in NHL risk associated with census-tract educational level among eligible nonparticipants and participants found a negligible bias between 1% and 8% (Shen et al. 2008). Also, dust samples were collected only from study participants who owned most of their carpets for at least five years; therefore, concentrations may not be representative of cases and controls who moved or replaced their carpets within five years of the interview.

Conclusions

We applied weighted quantile sum regression to estimate the association between NHL and an index for 27 correlated environmental chemicals measured in residential carpet dust. The WQS method allowed us to make generalized inference about the chemical mixture effect and identify the individual chemicals most strongly associated with NHL while considering the correlation between compounds. Using WQS regression, we found a positive association between the chemical index and NHL in the overall study population and in each of the four study sites. The WQS analysis also implicated several chemicals as NHL risk factors that were not associated with NHL when evaluated individually. Our results demonstrate the importance of evaluating chemical mixtures when studying cancer risk.

References

- Bithell JF, Dutton SJ, Draper GJ, Neary NM. 1994. Distribution of childhood leukaemias and non-Hodgkin's lymphomas near nuclear installations in England and Wales. *BMJ* 309(6953):501-505.
- Carrico C, Gennings C, Wheeler DC, Factor-Litvak P. 2014. Characterization of a weighted quantile sum regression for highly correlated data in a risk analysis setting. *J Agr Biol Envir St.*
- Colt JS, Davis S, Severson RK, Lynch CF, Cozen W, Camann D, et al. 2006. Residential Insecticide Use and Risk of Non-Hodgkin's Lymphoma. *Cancer Epidemiol Biomarkers Prev* 15(2):251–257.
- Colt JS, Lubin J, Camann D, Davis S, Cerhan J, Severson RK, et al. 2004. Comparison of pesticide levels in carpet dust and self-reported pest treatment practices in four US sites. *J Expo Anal Env Epid* 14:74–83.
- Colt JS, Severson RK, Lubin J, Rothman N, Camann D, Davis S, et al. 2005. Organochlorines in carpet dust and non-Hodgkin lymphoma. *Epidemiology* 16(4):516–525.
- De Roos AJ, Davis S, Colt JS, Blair A, Airola M, Severson RK, et al. 2010. Residential proximity to industrial facilities and risk of non-Hodgkin lymphoma. *Environ Res* 110(1):70–78.
- De Roos AJ, Hartge P, Lubin JH, Colt JS, Davis S, Cerhan JR, et al. 2005. Persistent organochlorine chemicals in plasma and risk of non-Hodgkin's lymphoma. *Cancer Res* 65(23):11214–11226.
- Dreiherr J, Novack V, Barachana M, Yerushalmi R, Lugassy G, Shpilberg O. 2005. Non-Hodgkin's lymphoma and residential proximity to toxic industrial waste in southern Israel. *Haematologica* 90(12):1709–1710.
- Engel LS, Laden F, Andersen A, Strickland PT, Blair A, Needham LL, et al. 2007a. Polychlorinated biphenyl levels in peripheral blood and non-Hodgkin's lymphoma: A report from three cohorts. *Cancer Res* 67(11):5545–5552.
- Engel LS, Lan Q, Rothman N. 2007b. Polychlorinated biphenyls and non-Hodgkin lymphoma. *Cancer Epidemiol Biomarkers Prev* 16(3):373–376.

- Everett CJ, Mainous AG III, Frithsen IL, Player MS, Matheson EM. 2008. Association of polychlorinated biphenyls with hypertension in the 1999–2002 National Health and Nutrition Examination Survey. *Environ Res* 108(1):94–97.
- Floret N, Mauny F, Challier B, Arveux P, Cahn JY, Viel JF. 2003. Dioxin emissions from a solid waste incinerator and risk of non-Hodgkin lymphoma. *Epidemiology* 14(4):392–398.
- Franchini M, Rial M, Buiatti E, Bianchi F. 2004. Health effects of exposure to waste incinerator emissions: a review of epidemiological studies. *Ann Ist Super Sanità* 40(1):101–115.
- Goldberg MS, Siemiatyck J, DeWar R, Désy M, Riberdy H. 1999. Risk of developing cancer relative to living near a municipal solid waste landfill site in Montreal, Quebec, Canada. *Arch Environ Health* 54(4):291–296.
- Hartge P, Colt JS, Severson RK, Cerhan JR, Cozen W, Camann D, et al. 2005. Residential herbicide use and risk of non-Hodgkin lymphoma. *Cancer Epidemiol Biomarkers Prev* 14:934–937.
- Hartge P, Wang S, Bracci P, Devesa S, Holly E. 2006. Non-Hodgkin lymphoma. In: *Cancer Epidemiology and Prevention*, Vol. 3rd edition, (Schottenfeld D, Fraumeni J, Jr., eds). New York:Oxford University Press, 898–918.
- Ioannidis JP, Loy EY, Poulton R, Chia KS. 2009. Researching genetic versus nongenetic determinants of disease: A comparison and proposed unification. *Sci Transl Med* 1(7):7ps8.
- Knobeloch L, Turyk M, Imm P, Anderson H. 2012. Polychlorinated biphenyls in vacuum dust and blood of residents in 20 Wisconsin households. *Chemosphere* 86(7):735–740.
- Lanphear BP, Matte TD, Rogers J, Clickner RP, Dietz B, Bornschein RL, et al. 1998. The contribution of lead-contaminated house dust and residential soil to children's blood lead levels. *Environmental Research, Section A* 79(1):51–68.
- Lubin JH, Colt JS, Camann D, Davis S, Cerhan JR, Severson R, et al. 2004. Epidemiologic evaluation of measurement data in the presence of detection limits. *Environ Health Perspect* 112(17):1691–1696.
- Morton LM, Wang SS, Cozen W, Linet MS, Chatterjee N, Davis S, et al. 2008. Etiologic heterogeneity among non-Hodgkin lymphoma subtypes. *Blood* 112(13):5150–5160.
- Patel CJ, Bhattacharya J, Butte AJ. 2010. An Environment-Wide Association Study (EWAS) on Type 2 Diabetes Mellitus. *PLoS ONE* 5(5):e10746.

- Patel CJ, Cullen MR, Ioannidis JP, Butte AJ. 2012. Systematic evaluation of environmental factors: persistent pollutants and nutrients correlated with serum lipid levels. *Int J Epidemiol* 41(3):828–843.
- Porta M, Gasull M, Pumarega J. 2012. Commentary: A step towards more comprehensive analyses of life course effects of mixtures of environmental factors. *Int J Epidemiol* 41(3):843–846.
- Pronk A, Nuckols JR, De Roos AJ, Airola M, Colt JS, Cerhan JR, et al. 2013. Residential proximity to industrial combustion facilities and risk of non-Hodgkin lymphoma: A case–control study. *Environ Health* 12(20).
- Rudel RA, Seryak LM, Brody JG. 2008. PCB-containing wood floor finish is a likely source of elevated PCBs in residents' blood, household air and dust: A case study of exposure. *Environ Health* 7(2).
- Shen M, Cozen W, Huang L, Colt J, De Roos AJ, Severson RK, et al. 2008. Census and geographic differences between respondents and nonrespondents in a case-control study of non-Hodgkin lymphoma. *Am J Epidemiol* 167(3):350–361.
- Stern AH, Fagliano JA, Savrin JE, Freeman NC, Liroy PJ. 1998. The association of chromium in household dust with urinary chromium in residences adjacent to chromate production waste sites. *Environ Health Perspect* 106(12):833–839.
- Wheeler DC, De Roos AJ, Cerhan JR, Morton LM, Severson R, Cozen W, et al. 2011. Spatial-temporal analysis of non-Hodgkin lymphoma in the NCI-SEER NHL case-control study. *Environ Health* 10(63).

Table 1. Characteristics of the study population overall and by study site.

	All Sites (N = 1180)	Detroit (N = 202)	Iowa (N = 340)	Los Angeles (N = 292)	Seattle (N = 346)
Characteristic	n (%)	n (%)	n (%)	n (%)	n (%)
Case Status					
Control	508 (43)	75 (37)	147 (43)	125 (43)	161 (47)
Case	672 (57)	127 (63)	193 (57)	167 (57)	185 (53)
Age ^a (years)	60±11.2	58±11.3	61±11.4	60±11.2	59±10.8
Sex					
Male	631 (54)	114 (56)	181 (53)	163 (56)	173 (50)
Female	549 (47)	88 (44)	159 (47)	129 (44)	173 (50)
Race					
White	1033 (88)	164 (81)	336 (99)	213 (73)	320 (92)
Non-white	147 (12)	38 (19)	4 (1)	79 (27)	26 (8)
Education					
< 12 years	106 (9)	24 (12)	33 (10)	30 (10)	19 (5)
12-15 years	741 (63)	123 (61)	244 (72)	171 (59)	203 (59)
≥ 16 years	333 (28)	55 (27)	63 (19)	91 (31)	124 (36)

^aContinuous variable summarized using mean ± standard deviation.

Table 2. Associations between non-Hodgkin lymphoma and the weighted quantile sum regression index in the study population and in each study site.

Parameter	N	Odds Ratio^a (95% CI)	p-value
WQS _F	1180	1.30 (1.08, 1.56)	0.006
WQS _D	202	1.71 (1.02, 2.92)	0.045
WQS _I	340	1.76 (1.23, 2.53)	0.002
WQS _L	292	1.44 (1.00, 2.08)	0.049
WQS _S	346	1.39 (0.97, 1.99)	0.071

Abbreviations: WQS, weighted quantile sum index; F, full data set; D, Detroit; I, Iowa; L, Los Angeles; S, Seattle.

^aEstimated odds ratios are associated with a unit increase in the WQS index. All models are adjusted for gender, race, education, and age. The model for the study population (i.e., the full data set) was also adjusted for study site.

Table 3. Weighted quantile sum regression index weights^a estimated in the study population and in each study site.

Chemical	Detroit (<i>p</i> = 0.045)^a	Iowa (<i>p</i> = 0.002)^a	Los Angeles (<i>p</i> < 0.049)^a	Seattle (<i>p</i> = 0.071)^a	All Sites (<i>p</i> = 0.006)^a
PCB 105	0.01	0.07	0.03	0.01	0.02
PCB 138	< 0.005	0.01	0.01	< 0.005	< 0.005
PCB 153	< 0.005	0.01	0.12	0.07	0.02
PCB 170	0.17	0.01	0.01	0.01	0.03
PCB 180	0.18	0.02	0.01	0.14	0.32
benz(a)anthracene	< 0.005	< 0.005	< 0.005	0.09	< 0.005
benzo(b)fluoranthene	< 0.005	< 0.005	0.10	< 0.005	< 0.005
benzo(k)fluoranthene	< 0.005	< 0.005	0.30	0.01	0.07
benzo(a)pyrene	0.07	0.04	< 0.005	0.03	0.06
chrysene	< 0.005	< 0.005	< 0.005	< 0.005	< 0.005
dibenz(ah)anthracene	< 0.005	0.03	0.03	0.01	0.01
indeno(1,2,3- <i>cd</i>)pyrene	0.03	< 0.005	< 0.005	< 0.005	< 0.005
α-chlordane	0.02	0.07	0.03	0.01	0.04
γ-chlordane	< 0.005	0.12	0.03	0.01	0.08
carbaryl	0.01	< 0.005	0.03	0.01	0.01
chlorpyrifos	0.01	0.02	< 0.005	< 0.005	< 0.005
<i>cis</i> -permethrin	0.09	0.01	0.03	0.02	< 0.005
<i>trans</i> -permethrin	0.06	< 0.005	0.01	0.09	0.03
2,4-D	0.05	< 0.005	0.11	< 0.005	< 0.005
DDE	< 0.005	0.11	0.07	0.14	0.08
DDT	0.01	< 0.005	0.01	0.01	< 0.005
diazinon	0.01	< 0.005	< 0.005	0.02	< 0.005
dicamba	0.09	< 0.005	0.06	< 0.005	< 0.005
methoxychlor	0.12	0.01	< 0.005	< 0.005	< 0.005
<i>o</i> -phenylphenol	< 0.005	0.11	< 0.005	0.06	0.04
pentachlorophenol	< 0.005	0.06	< 0.005	0.09	0.01
propoxur	0.05	0.30	0.01	0.16	0.17

^a*P*-value associated with the estimated weighted quantile sum regression index parameter as given in

Table 2.

Table 4. Associations between non-Hodgkin lymphoma subtypes and the weighted quantile sum regression index in the study population.

Parameter	N_{cases}	Odds Ratio^a (95% CI)	p-value
WQS _{DLBCL}	207	1.26 (0.94, 1.71)	0.128
WQS _{Follicular}	157	1.47 (1.08, 2.00)	0.014
WQS _{SLL/CLL}	67	1.26 (0.83, 1.93)	0.273
WQS _{MarginalZone}	61	2.06 (1.25, 3.47)	0.006
WQS _{Other}	91	2.26 (1.55, 3.34)	< 0.001
WQS _{NOS}	89	1.32 (0.91, 1.92)	0.144

Abbreviations: N_{cases}, number of cases; WQS, weighted quantile sum index; DLBCL, diffuse large B-cell; SLL/CLL chronic lymphocytic leukemia/small lymphocytic; NOS, not otherwise specified.

^aEstimated odds ratios are associated with a unit increase in the WQS index. Models are adjusted for gender, race, education, age, and study site.