Title

Bivariate Left-Censored Bayesian Model for Predicting Exposure: Preliminary Analysis of Worker Exposure during the *Deepwater Horizon* Oil Spill

Authors

Groth, Caroline
Banerjee, Sudipto

Publication Date

2015-11-23

Peer reviewed

Bivariate Left-Censored Bayesian Model for Predicting Exposure: Preliminary Analysis of Worker Exposure during the *Deepwater Horizon* Oil Spill

Caroline Groth[1], Sudipto Banerjee[2*], Gurumurthy Ramachandran[3], Mark R. Stenzel[4], Dale P. Sandler[5], Aaron Blair[6], Lawrence S. Engel[7], Richard K. Kwok[5] and Patricia A. Stewart[8]

[1] Division of Biostatistics, University of Minnesota, Minneapolis, Minnesota 55455.

[2] Department of Biostatistics, University of California-Los Angeles, Los Angeles, California 90095.

[3] Division of Environmental Health Sciences, University of Minnesota, Minneapolis, Minnesota 55455.

[4] Exposure Assessments Applications, LLC, Arlington, Virginia 22207

[5] National Institute of Environmental Health Sciences, Research Triangle Park, North Carolina 27709

[6] National Cancer Institute, Bethesda, Maryland 20892

[7] Department of Epidemiology, University of North Carolina at Chapel Hill, Chapel Hill, North Carolina 27599

[8] Stewart Exposure Assessments, LLC, Arlington, Virginia 22207

[*] *email:* sudipto@ucla.edu

Abstract

In April 2010, the *Deepwater Horizon* oil rig caught fire and exploded, releasing almost 5 million barrels of oil into the Gulf of Mexico over the ensuing 3 months. Thousands of workers participated in the spill cleanup and response efforts. The Gulf Longterm Follow-up Study (GuLF STUDY) being conducted by the National Institute of Environmental Health Sciences (NIEHS) is an epidemiological study to investigate potential adverse health effects among these response workers. Many volatile chemicals were released from the oil into the air, including total hydrocarbons (THC) that include benzene, toluene, ethylbenzene, xylene (BTEX), and hexane. Our goal is to estimate exposure levels to these toxic chemicals for groups of workers in the study (hereafter called exposure groups) with likely comparable exposure distributions. Although a large number of air measurements was collected, many exposure groups are characterized by a large percentage of censored measurements (below the analytic methods' limit of detection) or small sample sizes. Here we use THC, which is a composite of the volatile components of oil, the measurements of which have a low degree of censoring, as a predictor to develop linear models for estimating BTEX and hexane air exposure with higher degrees of censoring. We present a novel Bayesian hierarchical linear model that allows us to model, for different exposure groups simultaneously, exposure levels of a second chemical while accounting for censoring in both THC and the chemical of interest. We illustrate the methodology by estimating exposure levels for exposure groups on the *Development Driller III*, a rig vessel charged with drilling one of the relief wells. The model provided credible estimates in this example for geometric means, arithmetic means, variances, correlations, and regression coefficients for each group. This approach should be considered when estimating exposures in situations when multiple chemicals are correlated and have varying degrees of censoring.

KEYWORDS Bayesian Statistics, Bivariate Left-Censoring, Correlation, *Deepwater Horizon* Oil Spill, Exposure Assessment, total hydrocarbons (THC)

Introduction

On April 20, 2010, the *Deepwater Horizon* oil rig caught fire, exploded, and sank two days later, resulting in the release of almost 5 million barrels of oil into the Gulf of Mexico over the ensuing 3 months. Thousands of workers were involved in the cleanup response (hereafter called response).

Oil releases many harmful chemicals into the air reported as total hydrocarbons (THC). Total hydrocarbons is a composite of the volatile components of oil. Some of the more volatile components

include benzene, toluene, ethylbenzene, and xylene (BTEX) as well as hexane. A review of the literature by D'Andrea and Reddy (2014) found that increased exposure to these chemicals in prior spills was associated with a variety of detrimental health effects. The National Institute for Environmental Health Sciences (NIEHS) is conducting the GuLF STUDY to investigate potential adverse health exposure to these chemicals. Once estimates of exposure levels have been developed, epidemiologists will use these estimates to assess the relationships between various health outcomes and these exposures.

Exposures were measured using passive dosimeters worn by workers. The over 24,000 personal air samples (simultaneously measuring multiple chemicals resulting in about 140,000 measurements) included a number of results at or below the analytic limits of detection(LOD). Even these low levels of exposure are of interest because of possible additive or synergistic effects with other exposures experienced at the time (e.g. other oil components, dispersants, particulates, heat, and long working hours). Furthermore, the proportion of measurements below the LOD are a function of the sensitivity of the method and duration of measurement. Therefore values may be below the LOD because the test is insufficiently sensitive, not because they are necessarily negligible. In addition, a low exposed group is often used in epidemiological studies as the reference population to which higher exposed groups are compared. Thus, it is of interest to assess even these exposures below the LOD.

We, therefore, sought a statistical model that allows us to predict exposure levels to oil-related chemicals using the corresponding THC measurements because censoring is much more prevalent among these other chemical measurements. Equally important, the statistical model must properly account for uncertainty in the estimates. Specifically, we seek to quantify the uncertainty in our estimate of each censored observation and account for that uncertainty in our estimate of the relationship between oil-related chemicals and THC. This will allow us to gain insight over and above approaches that substitute a single value for all values below the LOD. Therefore, we constructed a hierarchical Bayesian regression modeling framework that accounts for censored observations in both THC and the chemical of interest (including cases when both the chemical and THC are censored). In this paper, we model THC as the predictor (X) and consider the response (Y) to be one of the following: benzene, toluene, ethylbenzene, xylene, or hexane. After establishing a relationship between THC and each of these chemicals of interest, we can use this relationship to generate relatively unbiased exposure estimates for each of these chemicals for groups of workers expected to have similar distributions of exposures (hereafter called exposure groups, EGs).

In the next section, we briefly describe the underlying chemistry for the relationship between THC and the chemicals of interest and review the statistical methodology currently in use to account for censored data. This is followed by a discussion of the importance of including censored data when estimating exposures by presenting particularly striking data from one of the rig ships in our study (the *Discoverer Enterprise*). Then we provide a description of the Bayesian modeling framework and evaluate our methodology with a simulation study conducted at a various censoring levels in X and Y, and conclude with an example of xylene exposure estimates, derived from this method, for a subset of EGs on another rig ship (the *Development Driller III*).

Background on the Relationships of Oil-related Chemicals

THC is a composite of the volatile components of crude oil that make up approximately 20% of the crude oil released in the *Deepwater Horizon* spill. Only about 10% of the volatile components (or 2% of the composition of the crude oil) combined is attributed to BTEX and hexane. This means for a particular exposure measurement, the concentration of the specific BTEX chemical or hexane will be a small fraction of the THC concentration. The analytical method sensitivity of the BTEX chemicals and hexane is more sensitive than that of THC (about 3 ppb for the former vs. 100 ppb for THC). Considering that the BTEX and hexane components comprised less than 10% of the THC by weight, the measurement data of the individual chemicals were more highly censored than the THC measurements.

The bases for the relationships of the chemical components in oil or in other mixtures are the Ideal Gas Law and Raoults Law (Stenzel and Arnold, 2015). The vapor concentration (VC) of a pure chemical in the air above a chemical's liquid surface, at a specific liquid temperature, is the ratio of the chemical's vapor pressure (VP) divided by the atmospheric pressure. With mixtures such as crude oil, the VP of each chemical component in the mixture is lower than that of the pure chemical. The degree of lowering is related to the chemical's percent composition in the mixture and its molecular weight. This lower VP is referred to as the chemical's adjusted vapor pressure (AVP). Once the AVP is determined, it can be divided by atmospheric pressure to estimate the VC of the chemical component in the air above the mixture surface. If the composition (in mass percent) of the mixture generating the vapor is constant, then the VC of each component will be constant, resulting in the relative VCs of the components of the mixture being constant (Stenzel and Arnold, 2015). This relative relationship of two components in the air can therefore be estimated

from the correlation of the components in a group of two corresponding sets of air measurements. The correlation approach is typically used in assessing exposure when analogous data from one chemical in a mixture are used to estimate exposures to another chemical in the mixture.

The exposures to THC, BTEX and hexane, as experienced by workers who participated in the *Deepwater Horizon* oil spill response, could have come from a number of sources including: the spilled crude oil, various solvents used in cleaning agents or in paints, chemicals associated with specific activities such as the drilling mud used in the drilling of the relief wells or chemicals associated with operating the equipment and vessels, fuels, or engine exhausts. The chemical composition, and concentration of those chemicals, varies substantially in these mixtures. If the predominant source of the THC was crude oil, and the composition of the oil did not change, a strong correlation should be observed between the measurement results for each of the chemical components and THC across the various EGs. In contrast, if the predominant source of exposure was the solvents and other chemicals in the workplace, weaker correlations would be expected because the composition of the solvents and other chemicals likely varied across the products used at the worksites and thus across the EGs. If the primary source was fuel exhausts, and the composition of the fuel was relatively constant, which is not expected, a correlation would be observed for those groups exposed to fuel or emissions, but the correlation would be different from that observed related to crude oil. Only a small portion of the personal measurements were collected on workers handling fuel or working near engine exhausts, and therefore this correlation would impact a very limited number of EGs.

A complicating factor related to oil spills is that the composition of released crude oil does not remain constant. The components in the oil can evaporate, dissolve into water, or be broken down by sunlight or bacteria. These processes are referred to as oil weathering and will result in a change in crude oil composition.

When the *Deepwater Horizon* sank it caused the oil to be released 5000 feet below the surface. Some of the weathering occurred to the oil as it rose to the surface, but once oil reached the surface, significant oil weathering occurred within a few days. During the period when the well was leaking, fresh leaking oil replaced the weathered oil on the surface, leading to a pseudo steady state that should result in an approximately constant crude oil composition and should lead to observable and strong correlations over the period the oil was being released. In cases where the integrity of the surface barrier of the oil or tar (arising from the weathered oil) was maintained, such as is the case with undissolved submerged oil, the volatile components of the crude should remain within

the crude oil plume or tar and only be released when the surface barrier is broken. Thus, release of crude oil volatile components could occur after the well was capped at significant distances in space or time from the original source of the spill. This phenomenon is also why the THC concentration can vary over time and space. From the perspective of this paper, it is not necessary to definitively identify the reason for the correlations between THC and the chemicals of interest, but rather, if, by empirical observation, correlations are constant over a defined time period, then these correlations can be used to estimate exposures to the THC components for the workers involved in the response.

Since the relationship is known to be linear in nature, we further assessed whether an intercept was necessary using a cross-validation approach. From this analysis (not shown), we concluded that including an intercept allowed us to predict exposure levels while minimizing the influence of outlying observations. Additionally, chemical concentrations are often depicted on a ln scale to better meet normality assumptions. This proved to be appropriate for our data. Therefore, the final model uses an intercept and the natural log scale for both the predictor and response.

From the occupational health literature, it is known that exposures may differ across a variety of determinants including job title, activity, and location. EGs are identified for a unique set of determinants and therefore within an EG, exposure distributions are assumed to be similar. The intensity of exposure can vary among EGs because of variations in tasks performed and the duration of these tasks, but the correlations between the specific chemicals and THC concentrations should be constant because the dominant source of the exposure is the crude oil. Since the ultimate goal in exposure assessment is providing exposure estimates for each of these EGs, we simultaneously modeled multiple EGs, each having its own intercept and slope coefficient to allow us to distinguish different levels of exposure among these different EGs. To draw unbiased inferences, we allowed the slope and intercept estimates to be influenced by other EGs in the model when censoring was high or sample size was low. This makes the model estimates for particular EGs more stable because the model formulates a global intercept and global slope estimate to use when information is limited.

Left Censoring Statistical Methods

Limited research exists on a statistical methodology that can account for left censoring in occupational health data. Methods commonly used in occupational health studies with censored data in one variable include classical methods where we impute the censored values. Some examples of these methods include β-substitution, LOD/$\sqrt{2}$, and LOD/2. These methods are fully discussed in Ganser and Hewett (2010). Although methods such as β-substitution were developed to find

the best imputed single value, such imputative strategies may not properly account for the inherent uncertainty in the imputed estimates and may introduce bias if the estimate is based on a small sample size or a high percentage of censored observations. A discussion of such substitution methods, including β-substitution, maximum likelihood and simple Bayesian models using non-regression approaches has been presented in Huynh et al. (2014, 2015).

Most statistical methods related to censoring focus on modeling survival data with a censored response and/or censored predictor. Statisticians often use these methods to study informative missingness, a situation where a particular known mechanism leads to missing or censored data. For example, Herring et al (2004) studied non-ignorable missingness in covariates. Their methods, however, only considered missing predictors and did not account for LOD left-censored responses. May et al (2011) analyzed LOD-censoring in multiple covariates in a generalized linear model framework using a Monte Carlo version of the Expectation Maximization (EM) algorithm. The EM algorithm is a classical iterative procedure used to estimate parameters of a model using expectations and by maximizing the likelihood. Likewise, Chen et al (2013) developed a framework to account for left censoring due to LOD in a Bayesian longitudinal study of overall pesticide concentrations measured over time, which were then used to predict pesticide poisoning symptoms. While both these methods considered censoring in more than one variable, these methods did not consider LOD-based censoring in multiple chemicals with possibly different LODs, and these methods did not model multiple EGs. Classical approaches for accounting for censoring in both response and predictors have been developed in Chu et al (2005), who used a likelihood based two-component mixture model to assess the correlation between two immunodeficiency (HIV) viral load measurements. Specifically, Chu et al assumed that X or Y could each be modeled by a mixture of two distribution components rather than a single normal distribution. Our work looks to expand this idea to a Bayesian context where we assess inference using a single component (compared to multiple components). We use a bivariate normal distribution framework to represent linear regressions and infer on both components X and Y.

Previous research by Huynh et al (2015) suggests that Bayesian models have advantages over frequentist methods such as β-substitution for censored analyses in occupational health studies. Comparing the frequentist methods of maximum likelihood estimation, β-substitution, and reverse Kaplan Meier methods, Huynh et al (2014) concluded that the β-substitution method was less biased and provided lower root mean squared error estimates than the other two methods under conditions of high censoring (>50%) and small (n=5-10) sample sizes. After assessing these fre-

quentist methods, Huynh et al (2015) compared Bayesian models to the β-substitution method. At various censoring levels, Bayesian models performed similarly (for bias and root mean squared error) to the β-substitution method for generating exposure estimates. The Bayesian models, however, also provided variance estimates, i.e. credible intervals, whereas the β-substitution method did not provide equations for calculating confidence intervals for many statistics of interest to us (i.e. geometric means, geometric standard deviations (GSDs) defined as the exponentiated standard deviation of a variable on the ln scale, and 95^{th} percentiles). Our contribution here is to expand upon the earlier work of Huynh et al. (2014, 2015) and propose a method for regression settings where either the dependent variable or the independent variable, or both, may be censored in one or more EG(s).

Importance of Censored Data

Because of the high level of censoring in our data, we determined whether the presence of censored data influences the relationship between oil-related chemicals. To do so, we analyzed data from various EGs on one of the drilling rigs charged with stopping the oil release, the *Discoverer Enterprise*. While our primary analysis in this paper focuses on analysis of the *Development Driller III*, the *Discoverer Enterprise* provided a particularly vivid example of the importance of censored data. We considered the relationship of THC to the BTEX chemicals and hexane because we assumed they all came from the source (i.e. crude oil) and THC had the lowest level of censoring. An overall regression of all outside measurements of workers who spent most of the time on the outside of the rig's living areas, such as the galley and sleeping quarters was performed accounting for censoring and not accounting for censoring, and 95% credible intervals were obtained for the slope and intercept. We generally found that there was a difference in the intercepts and slopes between groups of measurements with censored data and those without. For example, for benzene, approximately 86 percent of the measurements were censored as were 11 percent of the THC observations (Figure 1). Using inverse-gamma priors on the variance components, we found that the 95% credible interval for the slope was significantly positive when including censored data, while the credible interval for the slope included 0 when censored data were excluded.

Although we recognize that bias cannot be formally assessed using the comparison of these two datasets due to not knowing the true mean estimate, we believe, nevertheless, that these findings cogently demonstrate that additional censored data apparently provide important information that

could result in biasing the results if we omit them. If the slope was truly 0, we would expect the benzene exposure levels to be constant for all levels of THC. This is contrary to expectations arising from the oil composition. After accounting for censored information, we clearly saw that lower THC levels were associated with lower benzene exposure levels, as expected from physical and chemical laws. Similar discrepancies in slope and intercept estimates were found for the other four chemicals when comparing non-censored models to models including censored data (not shown). Thus, including the censored observations allows us to utilize more information over a wider range of values and yields more statistical power (because of larger sample sizes) to detect significant relationships between the chemicals.

Statistical Methods

First, consider a fairly standard hierarchical linear regression framework assuming that all measurements on Y and X are above their respective LODs. Instead of focusing solely upon the conditional distribution of $Y \,|\, X$, which proves restrictive when extending to censored or partially observed measurement pairs (X and Y), we prefer to work with a joint distribution for Y and X. We build this joint distribution by first modeling $X \sim N(\mu, \sigma_X^2)$ and then modeling the conditional distribution $Y \,|\, X \sim N(\beta_0 + \beta_1 X, \sigma_{Y|X}^2)$. A Bayesian hierarchical model is formulated by assigning prior distributions on these parameters. We use a customary univariate normal prior for μ, the mean of X with mean θ_μ and variance σ_μ^2, a bivariate normal prior for $\boldsymbol{\beta} = (\beta_0, \beta_1)^\top$ with mean vector $\boldsymbol{\mu}_\beta$ and variance-covariance matrix $\mathbf{V}_\beta$, and inverse-gamma priors for the variances σ_X^2 and $\sigma_{Y|X}^2$. This yields the joint distribution

$$\begin{aligned} &IG(\sigma_{Y|X}^2 \,|\, a, b) \times IG(\sigma_X^2 \,|\, c, d) \times N(\mu \,|\, \theta_\mu, \sigma_\mu^2) \times N(\boldsymbol{\beta} \,|\, \boldsymbol{\mu}_\beta, \mathbf{V}_\beta) \\ &\quad \times \prod_{j=1}^{m} N(X_j \,|\, \mu, \sigma_X^2) \times \prod_{j=1}^{m} N(Y_j \,|\, \beta_0 + \beta_1 X_j, \sigma_{Y|X}^2) \,, \end{aligned} \tag{1}$$

where we use the standard parametrizations for the normal $N(\cdot, \cdot)$ and inverse-gamma $IG(\cdot, \cdot)$ distributions, as given in, e.g., the text by Gelman et al. (2013). The shape parameters (a and c) and scale parameters (b and d) in the IG densities stipulate the extent of prior information on the variance components. For example, the *a priori* means for $\sigma_{Y|X}^2$ and σ_X^2 are $\dfrac{b}{a-1}$ and $\dfrac{d}{c-1}$, respectively, while the variances are $\dfrac{b^2}{(a-1)^2(a-2)}$ and $\dfrac{d^2}{(c-1)^2(c-2)}$, respectively. Now consider the situation we face: some measurements on X and Y are below LOD and, hence, not known

exactly. Let $LOD_j(X)$ and $LOD_j(Y)$ be the limits of detection on a ln scale for the j-th observation on X and Y, respectively. Let $C_X = \{j : X_j \leq LOD_j(X)\}$ and $C_Y = \{j : Y_j \leq LOD_j(Y)\}$ be the sets of indices for which Xs and Ys are censored, and let O_X and O_Y denote the complements of C_X and C_Y, respectively. The Bayesian hierarchical model is the joint distribution

$$\begin{aligned} & IG(\sigma^2_{Y|X} \,|\, a, b) \times IG(\sigma^2_X \,|\, c, d) \times N(\mu \,|\, \theta_\mu, \sigma^2_\mu) \times N(\boldsymbol{\beta} \,|\, \boldsymbol{\mu}_\beta, \mathbf{V}_\beta) \\ & \quad \times \prod_{j \in O_X} N(X_j \,|\, \mu, \sigma_X) \times \prod_{j \in C_X} TN(X_j \,|\, \mu, \sigma_X; -\infty, LOD_j(X)) \\ & \quad \times \prod_{j \in O_Y} N(Y_j \,|\, \beta_0 + \beta_1 X_j, \sigma^2_{Y|X}) \times \prod_{j \in C_Y} TN(Y_j \,|\, \beta_0 + \beta_1 X_j, \sigma^2_{Y|X}; -\infty, LOD_j(Y)) \,, \end{aligned} \tag{2}$$

where $TN(Z \,|\, \mu, \sigma^2; a, b)$ denotes the truncated normal distribution with mean μ, variance σ^2, and truncated between a and b with $-\infty \leq a < Z < b \leq \infty$.

The above assumes that the relationship between Y and X remains the same across EGs. We now extend (2) to multiple EGs by allowing the slope and intercept to vary across EGs. We call this model the hierarchical Bayesian EG model. Let Y_{ij} and X_{ij} be the j-th measurement on Y and X, respectively, in exposure group i, where $i = 1, 2, \ldots, N_{EG}$, and $j = 1, 2, \ldots, m_i$. With analogous definitions of O_X, O_Y, C_X and C_Y, the joint distribution is

$$\begin{aligned} & IW(\mathbf{V}_\beta \,|\, \mathbf{S}, \omega) \times N(\boldsymbol{\mu}_\beta \,|\, \boldsymbol{\theta}_\mu, \mathbf{V}_\mu) \times \prod_{i=1}^{N_{EG}} N(\mu_i \,|\, \nu, \gamma^2) \times N(\boldsymbol{\beta}_i \,|\, \boldsymbol{\mu}_\beta, \mathbf{V}_\beta) \\ & \quad \times \prod_{i=1}^{N_{EG}} IG(\sigma^2_{Y|X_i} \,|\, a_i, b_i) \times IG(\sigma^2_{X_i} \,|\, c_i, d_i) \times \prod_{(i,j) \in O_X} N(X_{ij} \,|\, \mu_i, \sigma^2_{X_i}) \\ & \quad \times \prod_{(i,j) \in C_X} TN(X_{ij} \,|\, \mu_i, \sigma^2_{X_i}; -\infty, LOD_{ij}(X)) \times \prod_{(i,j) \in O_Y} N(Y_{ij} \,|\, \beta_{0i} + \beta_{1i} X_{ij}, \sigma^2_{Y|X_i}) \\ & \quad \times \prod_{(i,j) \in C_Y} TN(Y_{ij} \,|\, \beta_{0i} + \beta_{1i} X_{ij}, \sigma^2_{Y|X_i}; -\infty, LOD_j(Y)) \,, \end{aligned} \tag{3}$$

where β_{0i} and β_{1i} are the intercept and slope parameters for exposure group i, μ_i is the mean of X's for each EG i, $\sigma^2_{Y|X_i}$ is the conditional variance of $Y \,|\, X_i$ for exposure group i and $\sigma^2_{X_i}$ is the variance of X_i for EG i. These two variances are assumed to be distributed independently across the EGs as inverse-gamma distributions. While the shape and scale of these inverse-gamma distributions are allowed to vary across the EGs in (3), in practice it is difficult to have strong prior information regarding these distributions, so we will assume that $a_i = a$, $b_i = b$, $c_i = c$ and $d_i = d$ and specify values for a, b, c and d. The μ_i's are also modeled *a priori* as normal

distributions, independent across EGs. For prior distributions on the regression coefficients, we define $\boldsymbol{\beta}_i = (\beta_{0i}, \beta_{1i})^\top$ as the 2×1 vector containing the intercept and the slope for EG i, which is distributed as a bivariate normal distribution with mean $\boldsymbol{\mu}_\beta$ and a 2×2 variance-covariance matrix $\mathbf{V}_\beta$. Again, these coefficients are assumed to be independent across EGs, but they borrow strength by shrinking the EG means to $\boldsymbol{\mu}_\beta$. Finally, $\boldsymbol{\mu}_\beta$ is assigned a Gaussian prior and $\mathbf{V}_\beta$ is modeled *a priori* with an inverse-Wishart (IW) distribution with parameters $\mathbf{S}$ and ω (see, e.g., Gelman et al., 2013).

The posterior distribution for the model parameters is proportional to the corresponding joint distribution in the respective models (1), (2) and (3). The posterior distribution is evaluated using numerical methods, arguably the most popular being Markov chain Monte Carlo (MCMC) algorithms such as the Gibbs sampler and Metropolis-Hastings algorithms (see, e.g., Gilks, Richardson, and Spiegelhalter, 1996; Marin and Robert, 2007; Carlin and Louis, 2008; Gelman et al. 2013; Brooks, Gelman, Jones and Meng, 2011). MCMC algorithms produce samples from the marginal posterior distribution of each unknown parameter in (1). All subsequent inference proceeds from these samples. Models (1), (2) and (3) are easily implemented in both Openbugs and RJAGS and easily evaluated. The code for these programs are provided in the supplementary materials.

Posterior Predictive Model Comparisons: D-statistics

Once the posterior distribution has been evaluated, e.g., using MCMC, Bayesian model assessment often proceeds from simulating replicates of the observed data (e.g., Gelman et al., 2013). To be specific, for (1), the joint posterior predictive distribution of the replicates for the i-th observation, $Y_{rep,i}$ and $X_{rep,i}$, is given by

$$p(Y_{rep,i}, X_{rep,i} \,|\, y_{obs}, x_{obs}) = \int N(Y_{rep,i} \,|\, \beta_0 + \beta_1 X_{rep,i}, \sigma^2_{Y|X}) \times N(X_{rep,i} \,|\, \mu, \sigma^2_X) \times p(\theta \,|\, y_{obs}, x_{obs}) d\theta \ , \quad (4)$$

where y_{obs} and x_{obs} are the observed Y's and X's, respectively, and $\theta = \{\beta_0, \beta_1, \mu, \sigma^2_X, \sigma^2_{Y|X}\}$. We draw samples from (4) by first sampling $\theta^{(l)}$ from the posterior distribution $p(\theta \,|\, y_{obs}, x_{obs})$, then sampling $X^{(l)}_{rep,i} \sim N(\mu^{(l)}, \sigma^{2(l)}_X)$, and finally sampling $Y^{(l)}_{rep,i} \sim N(\beta^{(l)}_0 + \beta^{(l)}_1 X^{(l)}_{rep,i}, \sigma^{2(l)}_{Y|X})$. This is repeated for $i = 1, 2, \ldots, n$.

For the censored model (2), how the replicates will be generated depends upon how X and Y have been measured. If both X and Y are above their respective LODs for the i-th observation, the posterior predictive distribution is the same as in (4), and we generate the replicates $Y_{rep,i}$ and $X_{rep,i}$ as described above. When X is above the LOD for the i-th observation, we draw the

replicates $X_{rep,i}^{(l)} \sim N(\mu^{(l)}, \sigma_X^{2(l)})$ for each posterior sample $\theta^{(l)}$. This is done irrespective of whether Y is above its LOD for that observation.

Now suppose that for the i-th observation, Y is above its LOD, but X is below its LOD. Bayesian inference treats unmeasured variables as unknown parameters, and any imputation of unmeasured variables must be carried out by sampling from the posterior distribution of the unmeasured variable. To be specific, let $X_{c,i}$ be the random variable denoting the unmeasured, or censored, X for the i-th observation. Note that replicates are defined only for observed data, so $X_{c,i}$ is not a replicate and will not be used in model assessment. However, it will be sampled in order to correctly sample the replicate $Y_{rep,i}$. Using the posterior samples $\mu_X^{(l)}$ and $\sigma_X^{2(l)}$, we will draw $X_{c,i}^{(l)} \sim TN(\mu_X^{(l)}, \sigma_X^{2(l)}; -\infty, LOD_j(X))$. Then, for each $X_{c,i}^{(l)}$, we draw $Y_{rep,i}^{(l)} \sim N(\beta_0^{(l)} + \beta_1^{(l)} X_{c,i}^{(l)}, \sigma_{Y|X}^{2(l)})$.

Finally, consider the model in (3). This extends (2) by allowing the parameters to vary by EG. Sampling the replicates will be the same as for (2) with θ now being the collection of all model parameters in (3).

For model comparisons, we use the replicated data to construct a "D-statistic" developed by Gelfand and Ghosh (1998) as an option for Bayesian predictive model assessment. The "D-statistic" can be computed for each model and can be used to compare different models fitted to the same dataset. Specifically, we compare the replicated data to the observed data by computing a goodness-of-fit measure G and a predictive variance P that penalizes more complex models. For (3), we compute

$$G = \sum_{(i,j)\in O_Y} \left(y_{ij} - \mu_{Y_{rep},ij}\right)^2 + \sum_{(i,j)\in O_X} \left(x_{ij} - \mu_{X_{rep},ij}\right)^2 \text{ and } P = \sum_{(i,j)\in O_Y} \sigma^2_{Y_{rep,ij}} + \sum_{(i,j)\in O_Y} \sigma^2_{X_{rep,ij}} \,, \tag{5}$$

where y_{ij} and x_{ij} are the observed measurements on Y and X, respectively, $\mu_{Y_{rep},ij}$ and $\mu_{X_{rep},ij}$ are the means, and $\sigma^2_{Y_{rep,ij}}$ and $\sigma^2_{X_{rep,ij}}$ are variances of $Y_{rep,ij}$ and $X_{rep,ij}$, respectively. The means and variances of $Y_{rep,ij}$ and $X_{rep,ij}$ used in (5) are computed from their samples. We then calculate $D = G + P$ as a metric for comparing models. Lower D-statistics are preferred. See Gelfand and Ghosh (1998) for theoretical details.

Simulation Studies

Methods

We performed three simulation experiments under different levels of LOD censoring. Table 1 describes the parameters we set for all three scenarios. We started off by dividing a set of 300 observations into 10 groups of various sizes ranging from 9 to 92, which is similar to what we see in the GuLF STUDY measurement data.

We set the true parameters using common characteristics of the data (not shown). Since the model uses a natural logged response and predictor, the parameters listed reflect what the parameters would be on the natural log scale for both X and Y. Specifically, we selected for each group a slope parameter between 0.6 and 1, which, for example, corresponds to the slopes generally found between ln(THC) (X) and ln(xylene) (Y). Based on previous regression models, we set intercept values between -2.5 and -0.5 (in ln(ppb) units), where an intercept can be interpreted as the mean estimate of ln(xylene) when ln(THC) is 0.

Then, also using common characteristics of the data, we set the mean of X to be between 5 and 7.25 ln(ppb) (148.4 ppb and 1408.1 ppb). Next, we set the variances to be between 1.44 and 5.29 for X and 0.49 and 3.24 for $Y \mid X$, corresponding to GSDs ranging from 3.3 to 10 for X, and from 2 to 6 for a second chemical Y. We then generated Xs from $N(\mu_x, \sigma_X^2)$ and, for each generated X, we drew a Y from $N(\beta_{0i} + \beta_{1i}X, \sigma^2_{Y|X_i})$, where μ_i, $\sigma^2_{X_i}$, β_{0i}, β_{1i}, and $\sigma^2_{Y|X_i}$ are as defined earlier.

The parameters described above were kept for all simulation studies. After assigning the parameters, three scenarios were defined. For the first scenario, the censoring levels were below 31% in both X and Y, corresponding to lower levels of LOD censoring. In the second scenario, the censoring on X remained the same as scenario 1, but we increased the censoring on Y to 25-50%. Finally, in the third scenario, censoring on X remained as in scenario 1, but the censoring ranged from 25-70% in Y, to demonstrate a scenario with highly censoring (censoring $> 50\%$) in the predicted variable of some groups. Censoring levels among the groups varied within a scenario, allowing for similar sample sizes to have different censoring levels. To be consistent, the percent censored in Y was always greater than or equal to the percent censored in X (as is generally seen in our GuLF STUDY data).

In order to create censoring, we determined the quantiles in each scenario corresponding to above or below each percentage censoring. All values below the quantile were censored or became missing. Following this, a set of LODs was assigned for each group in a uniform distribution just

below the quantile chosen. This allowed for multiple LODs for each group, due to, in our data, different durations of sampling (i.e. 4-18 hr). We implemented our Bayesian models by running an additional 10,000 MCMC iterations after 5,000 initial iterations for burn-in. In our model, we used inverse-gamma priors on the variance components. We also conducted simulation studies using informative uniform priors on the standard deviations with GSDs ranging from 1.01 to 12. Estimates of the intercept and slope parameters were similar, but the variance estimates varied more under the inverse-gamma priors as expected. The results with uniform priors on the standard deviations are included in the supplementary materials.

In our model, we used an inverse-Wishart prior on $\mathbf{V}_\beta$ with 2 degrees of freedom. The 2 by 2 scale matrix of this prior had upper left element 200, lower right element 0.2, and 0 otherwise. A normal prior was placed on μ_β with a mean vector $\mathbf{0}$ and variance-covariance matrix $\mathbf{V}_\mu$. The variance-covariance matrix $\mathbf{V}_\mu$ had variances of 1,000,000 and covariances of 0. We used a normal prior on each μ_i with mean 0 and variance 100,000 for all 10 groups. Then, finally, we used an inverse-gamma distribution on the $\sigma^2_{X_i}$ and $\sigma^2_{Y|X_i}$ for each group with shape parameter 0.01 and scale parameter 0.01.

We also compared our hierarchical Bayesian EG model in (3) with three simpler models for each of the three scenarios described earlier. For model comparisons, we replicated the observed Xs and observed Ys from the respective models. In the first model, only an intercept was included for prediction of X and Y; X was not used in the estimation of Y, and each group was modeled separately. This assumed different variances for each group where we simply modeled means, not accounting for additional information. The second model had a global intercept and global slope, where groups were not modeled separately but as one group. The third model used varying intercepts for groups but assumed that all groups had the same slope estimate. In all of the above models, we account for censoring in X and Y. D-statistics were used to compare models.

Results

The results of the model comparison for all three scenarios indicate that the hierarchical Bayesian EG model was preferred according to the D-statistic (Table 2). The D-statistic, in all scenarios, was lower for the hierarchical Bayesian EG model than for the other three model types. This finding demonstrates that if groups really did have their own intercept and slope estimates, the hierarchical Bayesian EG model would be preferred over the simpler models. Across all three scenarios, modeling the groups separately was meaningful. The D-statistics should not be compared across scenarios

since the datasets between the three scenarios were fundamentally different due to the different levels of censoring. G-statistics were consistently higher for the common intercept and common slope model, indicating that there were great deviations between the replicates based on this model and the real values. The real values were not generated based on a single regression line, and were, as described above, based on individual regression lines per group. Thus, this finding was expected.

The credible intervals (CI) are provided for the intercept (β_0), slope (β_1), variance of X (σ^2_X), and variance of $Y \mid X$ ($\sigma^2_{Y|X}$) to see if they contained the true value of the parameter (Table 3). In all scenarios, all parameters were contained within the credible intervals. In scenario 1, all slopes were significantly positive, although group 6 was barely so. This particular group had a small number of non-censored samples below 10 that likely led to the wide credible interval. Thus, slope estimation was reasonable in this scenario and followed what we would expect based on the values we provided. The upper bounds on the variance of X in groups 3 and 7 were quite high. However, in both of these cases, we had set the highest variances for these parameters of the groups, so this result was expected.

In scenario 2 with moderate censoring in Y (25-50% censoring), the credible intervals tended to slightly wider for the slopes compared to scenario 1. With increased censoring, there was less certainty and smaller non-censored sample sizes to estimate the true parameters. For group 6, the slope was insignificant as seen from the 95% credible interval, which marginally includes 0. We note that the 90% credible interval (not shown) did not include 0, indicating significance at this level. For most groups, the upper bounds of the variance of $Y \mid X$ increased from scenario 1 to scenario 2. As censoring increased in Y, there may have been more variability that went into estimation of Y at lower values of X, increasing the variance of $Y \mid X$ in some cases.

In scenario 3 containing some high levels of censoring in Y, the group with the highest censoring, group 2, had an increased median posterior intercept and a decreased median posterior slope compared to scenarios 1 and 2 (medians not shown). Since censoring was relatively high in this scenario, our model began to use inference from other groups to model this group. Overall, the slopes for the other groups were lower than this group. Therefore, this group's slope estimate at very high levels of censoring closely reflected the slopes of other groups.

To summarize, these results highlight that the model performed well under a variety of levels of censoring and that the 95% credible intervals contained the true parameters. It is expected that as censoring increases the relationships will change, but the model clearly was able to generate reasonable estimates and model the data adequately at levels $< 70\%$ censoring.

Illustrative Example of the *Deepwater Horizon* Oil Spill

In this preliminary analysis, we focused on seven EGs who worked between May 15 and July 15, 2010 on the *Development Driller III (DD3)*, a rig ship charged with drilling one of the relief wells. During this time period, efforts by other vessels were being made to stop the oil release. In addition, dispersants were being injected at the Gulf floor and on the surface of the water to break up the oil and reduce atmospheric concentrations of oil-derived substances.

The EGs considered are summarized in Table 4. Censoring in the THC measurements ranged from 0-25% and in the xylene measurements from 0-32.3%. Censoring was higher for xylene than for THC in most of the groups. Sample sizes ranged from 6-96. A total of 169 observations was considered in this analysis.

Inverse-gamma priors on the variance components were used for modeling. The prior parameters used for modeling were the same as described in the simulation study, except with seven groups instead of ten. We also used informative uniform priors on the standard deviation of THC and standard deviation of xylene | THC. The GSDs for these informative priors were set to be between 1.01 and 12 for these parameters as these are the GSDs that have been observed in our GuLF STUDY datasets. To test how influential the priors would be, both models were compared. The GSD estimates tended to be higher in the inverse-gamma prior case, which influenced the upper bounds of the parameter estimates. However, inference was not substantially changed between models (most patterns remained consistent). Figures of the intercepts, slopes, correlation coefficients, geometric means (GM), GSDs, and arithmetic means (AM) using the uniform priors are provided in the supplementary materials (Figures 8-14).

Convergence diagnostics, as assessed by Gelman Rubin statistics and trace plots, indicated that convergence was almost immediate. The Gelman Rubin diagnostics were less than 1.2 for all parameters of interest for the first 5,000 iterations of the model. Therefore, to ensure all parameters had converged adequately, we used 25,000 iterations after 5,000 iterations of burn-in.

Figure 2 displays a plot of the non-censored datapoints and separate linear regression lines for each EG. As demonstrated by this plot, particular EGs may have had slightly different linear relationships. In general, most points tended to follow a linear trend that could be summarized by a single regression line. However, the censored information must be included in order to know how the relationships differ among EGs. The plot also indicates that a few of the observations may be outliers. Nevertheless, since every point is considered important in our dataset, outliers were not

excluded.

A model comparison like that performed for the simulation study was repeated for this dataset. We compared our hierarchical Bayesian EG model to an intercept only model, a global slope and intercept model, and an EG-specific intercept with a global slope model. As earlier, we used D-statistics. In this case, the single intercept and slope model accounting for censoring in X and Y had the lowest D-statistic of the models tested (Table 5). This is related to the relatively low degree of censoring in this dataset and the highly linear trend among most of the non-censored data points as shown in Figure 2. In addition, the P statistic was elevated in the hierarchical Bayesian EG model because many additional parameters were estimated compared to other models. However, we still argue that our hierarchical model provides additional inference that may be useful. While the D-statistic indicates that the single slope and intercept was a good fit for the data, that model doesn't allow us to fully grasp the differences in the chemical relationship between ln(THC) and ln(xylene) for these different EGs, since all groups were modeled together.

As previously discussed, limited work has been done to incorporate a linear relationship in estimation while accounting for censoring in X and Y. From this model comparison, we can see that a common slope with a common intercept model and common slope with varying intercept models were superior to the intercept only model. Therefore, accounting for the additional information from the linear relationship was useful.

The global parameter estimates from the hierarchical Bayesian EG model are displayed in Table 6. The overall intercept posterior median estimate was -1.49 in natural log units, but this estimate was not significant. The lack of significance suggests that when ln(THC)=0 (or THC=1 ppb), ln(xylene)=0 (or xylene=1 ppb). The global slope estimate was significantly positive with a median posterior estimate of 0.70. This indicates that for every one unit increase in ln(THC), there is a corresponding 0.70 ln unit increase in xylene. The large amount of variance in the intercept (33.31) is likely due to the low accuracy and precision of the analytical method near the chemical's LOD. The relatively low variability in the slope estimate, however, suggests that there is likely to be one major source generating these exposures.

Median and 95% credible intervals for the intercept, slope, correlation, GSD of xylene, GSD of THC, GSD of xylene | THC, AM of THC, and AM of xylene are reported in Table 7. The corresponding figures, including figures for the GMs, are included in the supplementary materials (Figures 1-7).

The floorhand/roughneck and the operations technician/operator had the highest median es-

timates for both THC and for xylene for the slope, GMs, GSDs and AMs, as well as the highest credible intervals for these statistics. These groups were characterized by having some very high and some very low measurements for each of the two chemicals (e.g., for the floorhand/roughneck, the measurements ranged for THC >5000 ppb to <LOD of 100 ppb). These groups were directly involved in the drilling, tasks that had the highest exposures, but the work also varied from day to day and hour to hour, resulting in the high variability. The roustabout and crane operator (who often worked together) and the ROV technician generally had low GMs and AMs for both THC and xylene.

The correlation estimates between ln(THC) and ln(xylene) were all quite strong; all were significantly positive with median posterior estimates above 0.6. The correlations were strongest for the ROV EG (median correlation posterior estimate: 0.94) and IH-Safety (0.89). The floorhand/roughneck and operations technician/operator EGs had the lowest median correlation posterior estimates but the credible intervals were once again very wide for both these EGs. Given the variability of the tasks, these jobs may have more than one source of THC or xylene exposure.

Discussion

In general, these results demonstrated that this new method is promising for estimating exposures under a variety of levels of censoring of multiple EGs. These findings suggest that we could be biasing the exposure levels if we simply use the non-censored data to predict the mean exposure in the study since we would be omitting known, and apparently important, information. Therefore, it is important to include censored data information. Our model provides a potential avenue for accounting for these censored data when left-censoring occurs in both X and Y in a linear regression context. Our work highlights that the chemical relationship between two chemicals, such as the BTEX or hexane and THC, can be used to estimate exposures when one chemical is less highly censored.

Our simulation study showed that the model will perform well under low, moderate, and moderate to high levels of censoring in Y while having censoring levels low in X. D-statistics indicated that our model was consistently the best model in each scenario of the simulation study. All 95% credible intervals contained the true simulated parameters. Estimation was relatively robust for small sized EGs or with groups with higher levels of censoring.

Our application of the methodology to the GuLF STUDY data indicated patterns consistent

with the measurement data and tasks being performed. Expected high exposure jobs that involved a variety of tasks had higher levels of exposure, higher variability and wider credible intervals than jobs that were expected to have lower levels of variability. Higher censoring and smaller sample sizes increased the width of the credible intervals.

These results suggest that the correlation between THC and its volatile components may be a powerful tool to use for generating exposure estimates, particularly when censoring is lower in one chemical than in another chemical. The correlations were surprisingly strong (median posterior estimates from 0.64 to 0.94) in all EGs allowing for this estimation to be robust. The strong correlations indicate that we may be able to use them as Bayesian priors for the prediction of xylene exposure when xylene exposure was not detected for EGs on a variety of other vessels in the study.

This method also appears to more properly account for values below a LOD than other methods. Instead of simply ignoring censored data or substituting a single value, we are able to fully estimate the distribution of each chemical. We are able to use the known information that the chemical was censored to generate estimates. This helps us avoid having potentially biased estimates.

Our results were dependent on the level of censoring. Although we used a range of censoring for X and Y in the simulation study, for some EGs in the real study we see even higher levels of censoring or even smaller sample sizes. This is likely to result in increased uncertainty, and therefore, it is recommended that our model be used with caution at higher levels and smaller sample sizes than evaluated here.

These results depended on the relationship between THC and its volatile components being linear in nature. If the relationships were not linear, this methodology would not work sufficiently. We also assumed that each volatile component on a ln scale was normally distributed. Other distributions were not investigated.

Additionally, here one chemical was used for predicting each single BTEX chemical or hexane. Future work should explore expanding this work to a scenario with multiple X censored variables. It is possible that by including more than one chemical as predictors, we can obtain even stronger estimates of exposure. Likewise, researchers should further investigate whether modeling the observed values with truncated normal distributions above the LOD is statistically worthwhile over that of modeling observed data as normally distributed.

Acknowledgments

We would like to thank Wendy McDowell, Tran Huynh, and other members of the GuLF STUDY for their continued efforts. This research was supported by the Intramural Research Program of the NIH, National Institute of Environmental Health Sciences (Z01 ES102945) and the NIH Common Fund.

References

Brooks S, Gelman A, Jones G, Meng X. (2011) 1st ed. Handbook of Markov Chain Monte Carlo. Boca Raton, FL: Chapman & Hall/CRC. ISBN 1420079417.

Carlin BP, Louis TA. (2008) 3rd ed. Bayesian Methods for Data Analysis. Boca Raton, FL: Chapman & Hall/CRC. ISBN 1584886978.

Chen H, Quandt SA, Grzywacz JG, Arcury TA. (2013) A Bayesian multiple imputation method for handling longitudinal pesticide data with values below the limit of detection. Environmentrics; 24(2):132-42.

Chu H, Moulton LH, Mack WJ, Passaro DJ, Barraso PF, Muñoz A. (2005) Correlating two continuous variables subject to detection limits in the context of mixture distributions. Appl Stat; 54(5): 831-45.

D'Andrea M., Reddy, K. (2014) Crude oil spill exposure and human health risks. JOEM; 56(1): 1029-1041.

Ganser GH, Hewitt P. (2010) An accurate substitution method for analyzing censored data. J of Occup Environ Hyg; 7(4): 233-44.

Gelfand AE, Ghosh SK. (1998) Model choice: A minimum posterior predictive loss approach. Biometrika; 85(1): 1-11.

Gelman A, Carlin B, Stern HS, Dunson DB, Vehtari A, Rubin DB. (2013) 3rd ed. Bayesian Data Analysis. Boca Raton, FL:CRC Press. ISBN 1439840954.

Gilks WR, Richardson S, Spiegelhalter, D. (1996) Markov Chain Monte Carlo in Practice. Boca Raton, FL:Chapman & Hall/CRC. ISBN 0412055511.

Herring AH, Ibrahim JG, Lipsitz SR. (2004) Non-ignorable missing covariate data in the survival analysis: A case study of international breast cancer study group trial. J R Stat Ser C Appl Stat; 53(2): 293-310.

Huynh T, Ramachandran G, Banerjee S, Monteiro J, Stenzel G, Sandler DP, et al. (2014) Comparison of methods for analyzing left-censored occupational exposure data. Ann Occup Hyg.

Huynh T, Quick H, Ramachandran G, Banerjee S, Stenzel G, Sandler DP, et al. (2015) A comparison of the β substitution and Bayesian approach for handling left-censored data. Accepted with revision to Ann Occup Hyg.

Marin J, Robert C. (2007) 1st ed. Bayesian Core: A Practical Approach to Computational Bayesian Statistics. New York, NY: Springer. ISBN 0387389792.

May RC, Ibrahim JG, Chu H. (2011) Maximum likelihood estimation in generalized linear models with multiple covariates subject to detection limits. Stat Med; 30:2511-61.

Stenzel MR, Arnold SF. Rules and Guidelines to Facilitate Professional Judgments. In Jahn DJ, Bullock WH, Ignacia JS, editors. A Strategy for Assessing and Managing Occupational Exposures, 4th Ed. AIHA: 2015. p. 335-348. ISBN 978-1-935082-46-0

Tables and Figures

		True Parameter Values					Scenario 1		Scenario 2		Scenario 3	
Group	N	β_0	β_1	σ^2_X	$\sigma^2_{Y\|X}$	μ_x	$X < LOD$ (%)	$Y < LOD$ (%)	$X < LOD$ (%)	$Y < LOD$ (%)	$X < LOD$ (%)	$Y < LOD$ (%)
1	9	-1.50	0.70	2.25	1.00	5.50	22.2	22.2	22.2	33.3	22.2	33.3
2	50	-2.50	1.00	2.25	1.44	6.75	10.0	10.0	10.0	50.0	10.0	70.0
3	10	-2.00	0.80	4.00	3.24	6.50	20.0	30.0	20.0	30.0	20.0	40.0
4	92	-2.00	0.80	2.25	1.00	6.50	15.2	25.0	15.2	41.3	15.2	59.8
5	20	-1.50	0.70	1.44	0.49	6.80	5.0	10.0	5.0	35.0	5.0	35.0
6	12	-1.20	0.65	2.56	1.44	5.50	25.0	25.0	25.0	25.0	25.0	25.0
7	15	-2.00	0.80	5.29	3.24	6.20	13.3	13.3	13.3	26.7	13.3	33.3
8	14	-2.50	1.00	1.44	1.00	7.25	14.3	14.3	14.3	35.7	14.3	50.0
9	16	-0.50	0.60	2.25	1.69	5.00	18.8	25.0	18.8	43.8	18.8	43.8
10	62	-1.50	0.70	2.89	1.96	6.25	21.0	30.6	21.0	30.6	21.0	59.7

Table 1: Simulation study scenarios for assessing oil-related chemical exposure. β_0 is the intercept, β_1 is the slope, σ^2_X is the variance of X, $\sigma^2_{Y|X}$ is the variance of $Y \mid X$, and μ_x is the mean of X.

Scenario	Model	D-Statistic	P	G
1	Intercept Only Model	2710.0	1681.7	1028.4
	Common Intercept and Common Slope	2775.0	1591.8	1183.3
	Common Slope and Varying Intercepts	2748.5	1706.6	1041.8
	Hierarchical Bayesian EG Model	2690.8	1664.1	1026.6
2	Intercept Only Model	2656.6	1632.0	1024.6
	Common Intercept and Common Slope	2731.8	1535.3	1196.5
	Common Slope and Varying Intercepts	2635.5	1617.9	1017.6
	Hierarchical Bayesian EG Model	2602.9	1595.9	1007.0
3	Intercept Only Model	2517.0	1476.5	1040.5
	Common Intercept and Common Slope	2555.1	1362.1	1193.0
	Common Slope and Varying Intercepts	2415.1	1431.5	983.6
	Hierarchical Bayesian EG Model	2404.2	1417.8	986.4

Table 2: Model Comparison: Simulation study of models for assessing exposure to oil-related chemicals

	Scenario1				Scenario 2				Scenario 3			
	β_0	β_1	σ^2_X	$\sigma^2_{Y\|X}$	β_0	β_1	σ^2_X	$\sigma^2_{Y\|X}$	β_0	β_1	σ^2_X	$\sigma^2_{Y\|X}$
Group	CI	CI	CI	CI	CI	CI	CI	CI	CI	CI	CI	CI
1	(-4.89, -0.56)	(0.46, 1.23)	(0.72, 12.29)	(0.20, 2.30)	(-5.30, -0.50)	(0.44, 1.28)	(0.67, 9.96)	(0.21, 3.23)	(-5.17, -0.43)	(0.44, 1.25)	(0.67, 9.02)	(0.21, 3.40)
2	(-3.19, -0.80)	(0.74, 1.11)	(1.93, 4.51)	(0.73, 1.71)	(-4.29, -0.96)	(0.75, 1.22)	(1.95, 4.64)	(0.84, 2.80)	(-3.65, 0.14)	(0.66, 1.14)	(1.93, 4.54)	(0.48, 2.26)
3	(-4.63, 1.96)	(0.20, 1.00)	(2.55, 27.37)	(0.77, 9.88)	(-4.98, 1.94)	(0.20, 1.04)	(2.59, 28.67)	(0.77, 10.81)	(-5.64, 2.00)	(0.20, 1.09)	(2.57, 26.46)	(0.87, 14.75)
4	(-3.13, -1.05)	(0.68, 0.96)	(1.78, 3.46)	(0.63, 1.27)	(-3.12, -0.83)	(0.66, 0.96)	(1.77, 3.43)	(0.51, 1.13)	(-3.38, -0.48)	(0.62, 0.98)	(1.76, 3.41)	(0.48, 1.31)
5	(-2.59, 2.00)	(0.19, 0.87)	(0.62, 2.42)	(0.43, 1.85)	(-3.63, 1.42)	(0.28, 1.00)	(0.64, 2.47)	(0.32, 2.00)	(-3.59, 1.37)	(0.29, 0.99)	(0.64, 2.49)	(0.32, 1.97)
6	(-3.33, 2.12)	(0.01, 0.89)	(1.38, 10.95)	(0.69, 8.14)	(-3.60, 2.16)	(-0.01, 0.92)	(1.38, 11.93)	(0.68, 8.40)	(-3.60, 2.03)	(0.01, 0.92)	(1.34, 10.49)	(0.69, 8.02)
7	(-5.40, -1.36)	(0.57, 1.15)	(3.04, 15.76)	(0.83, 4.87)	(-6.43, -1.77)	(0.61, 1.27)	(3.17, 19.30)	(0.83, 5.82)	(-6.63, -1.71)	(0.62, 1.30)	(3.05, 16.65)	(0.62, 4.55)
8	(-4.31, 1.98)	(0.39, 1.21)	(0.60, 3.69)	(0.55, 3.37)	(-5.22, 2.42)	(0.33, 1.30)	(0.61, 3.78)	(0.60, 5.20)	(-5.48, 2.54)	(0.28, 1.28)	(0.60, 3.83)	(0.77, 11.13)
9	(-2.86, 1.30)	(0.22, 0.91)	(1.86, 10.70)	(0.78, 5.07)	(-4.12, 1.25)	(0.22, 1.04)	(1.90, 11.26)	(0.83, 8.20)	(-4.06, 1.22)	(0.22, 1.05)	(1.85, 11.14)	(0.81, 7.42)
10	(-3.03, -0.40)	(0.50, 0.87)	(2.63, 6.07)	(0.99, 2.47)	(-3.11, -0.46)	(0.50, 0.88)	(2.65, 6.02)	(0.97, 2.53)	(-4.26, -0.66)	(0.53, 0.99)	(2.68, 6.16)	(0.92, 3.21)

Table 3: Simulation study credible intervals for parameters in our hierarchical Bayesian EG model. β_0 is the intercept, β_1 is the slope, σ^2_X is the variance of X, and $\sigma^2_{Y|X}$ is the variance of $Y \mid X$. The median and 95% credible intervals (CI) are reported for each parameter.

Exposure Group	N	% Censored THC	% Censored Xylene
Derrick Hand	6	0.0	0.0
Floorhand/Roughneck	10	10.0	10.0
Crane Operator	16	12.5	31.3
Roustabout	96	19.8	32.3
Operations Technician or Operator	10	20.0	20.0
ROV Technician	12	25.0	25.0
IH-Safety	19	10.5	5.3

Table 4: Description of EGs (Number of EGs=7) on the *DD3* May 15-July 15, 2010 assessed in this preliminary analysis as part of the GuLF STUDY.

Model	D-Statistic	P	G
Intercept Only Model	772.0	479.1	292.8
Common Intercept and Common Slope	732.4	406.0	326.4
Common Slope and Varying Intercepts	750.3	440.6	309.8
Hierarchical Bayesian EG Model	790.2	471.0	319.2

Table 5: Model Comparison of models for accessing worker exposure to xylene on the DD3

Model Parameter	Median	95% Credible Interval
μ_{β_0}	-1.49	(-6.47, 3.46)
μ_{β_1}	0.70	(0.45, 0.97)
Σ_{11}	33.31	(13.06, 123.21)
Σ_{22}	0.06	(0.02, 0.27)
$\rho(\mu_{\beta_0}, \mu_{\beta_0})$	-0.16	(-0.75, 0.58)

Table 6: Preliminary Results:*DD3* May 15-July 15, 2010 hierarchical Bayesian EG model global parameter estimates

	Derrick Hand		Floorhand/Roughneck		Crane Operator		Roustabout		Operations Technician or Operator		ROV Technician		IH-Safety	
Parameter	Median	95% CI	Median	95% CI	Median	95% CI	Median	95% CI	Median	95% CI	Median	95% CI	Median	95% CI
Intercept	-1.98	(-4.76, 0.86)	-1.65	(-4.67, 1.24)	-1.34	(-3.51, 0.61)	-2.84	(-4.29, -1.51)	-0.96	(-3.86, 1.65)	-1.03	(-2.49, 0.21)	-0.49	(-1.73, 0.58)
Slope	0.76	(0.36, 1.17)	0.72	(0.33, 1.13)	0.62	(0.32, 0.95)	0.91	(0.69, 1.14)	0.68	(0.28, 1.09)	0.64	(0.44, 0.86)	0.59	(0.43, 0.78)
Correlation	0.78	(0.32, 0.97)	0.64	(0.27, 0.89)	0.69	(0.35, 0.88)	0.70	(0.57, 0.80)	0.67	(0.27, 0.91)	0.94	(0.77, 0.99)	0.89	(0.70, 0.96)
GSD of THC	2.55	(1.73, 8.50)	4.63	(2.72, 16.11)	2.90	(2.12, 5.21)	3.09	(2.63, 3.83)	5.45	(2.88, 25.05)	3.24	(2.08, 9.05)	2.68	(2.00, 4.57)
GSD of Xylene | THC	1.76	(1.38, 3.70)	3.76	(2.32,11.79)	2.00	(1.57, 3.39)	2.82	(2.40, 3.50)	3.48	(2.12, 12.38)	1.31	(1.19, 1.66)	1.36	(1.24, 1.61)
GSD of Xylene	2.64	(1.80, 7.55)	6.04	(3.31, 20.47)	2.68	(1.95, 4.86)	4.32	(3.46, 5.80)	5.87	(3.02, 25.49)	2.24	(1.65, 4.55)	1.95	(1.63, 2.67)
AM of THC (ppb)	1472	(701, 12168)	3454	(1170, 60690)	871	(503, 2321)	675	(517, 944)	2050	(601, 98761)	852	(420, 5030)	889	(554, 2016)
AM of Xylene (ppb)	42	(20, 277)	144	(39, 3487)	20	(12, 50)	35	(24, 59)	121	(33, 4747)	23	(15, 57)	32	(24, 50)

Table 7: Preliminary Results: *DD3* May 15-July 15, 2010 hierarchical Bayesian EG model parameter estimates

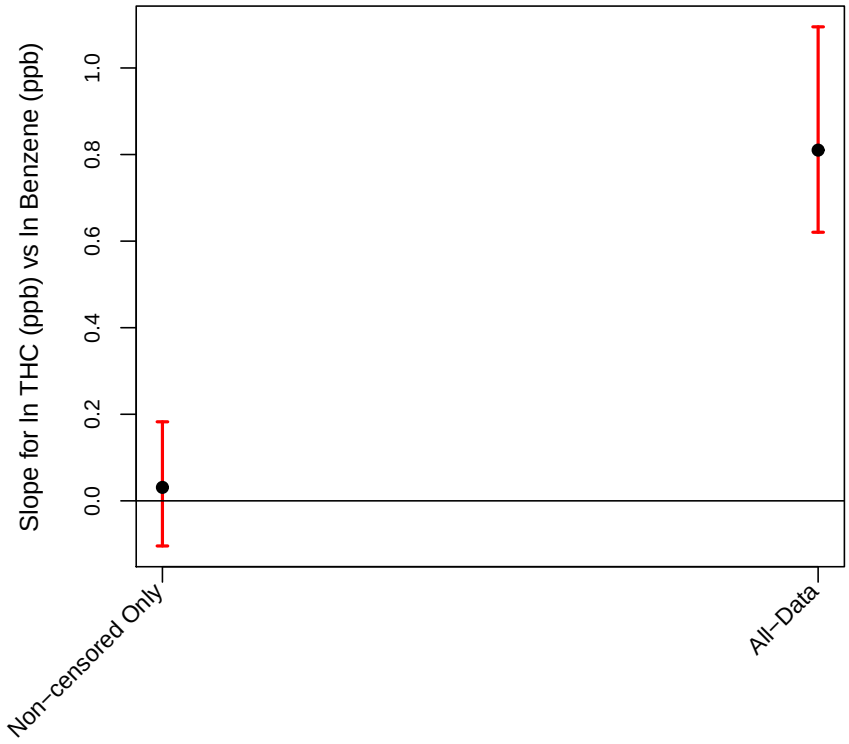

Figure 1: Comparing slope 95% credible intervals for only non-censored observations and for all outside observations (censored and non-censored) on the *Enterprise* in the THC and benzene datasets. The dots in each bar represent the median posterior samples.

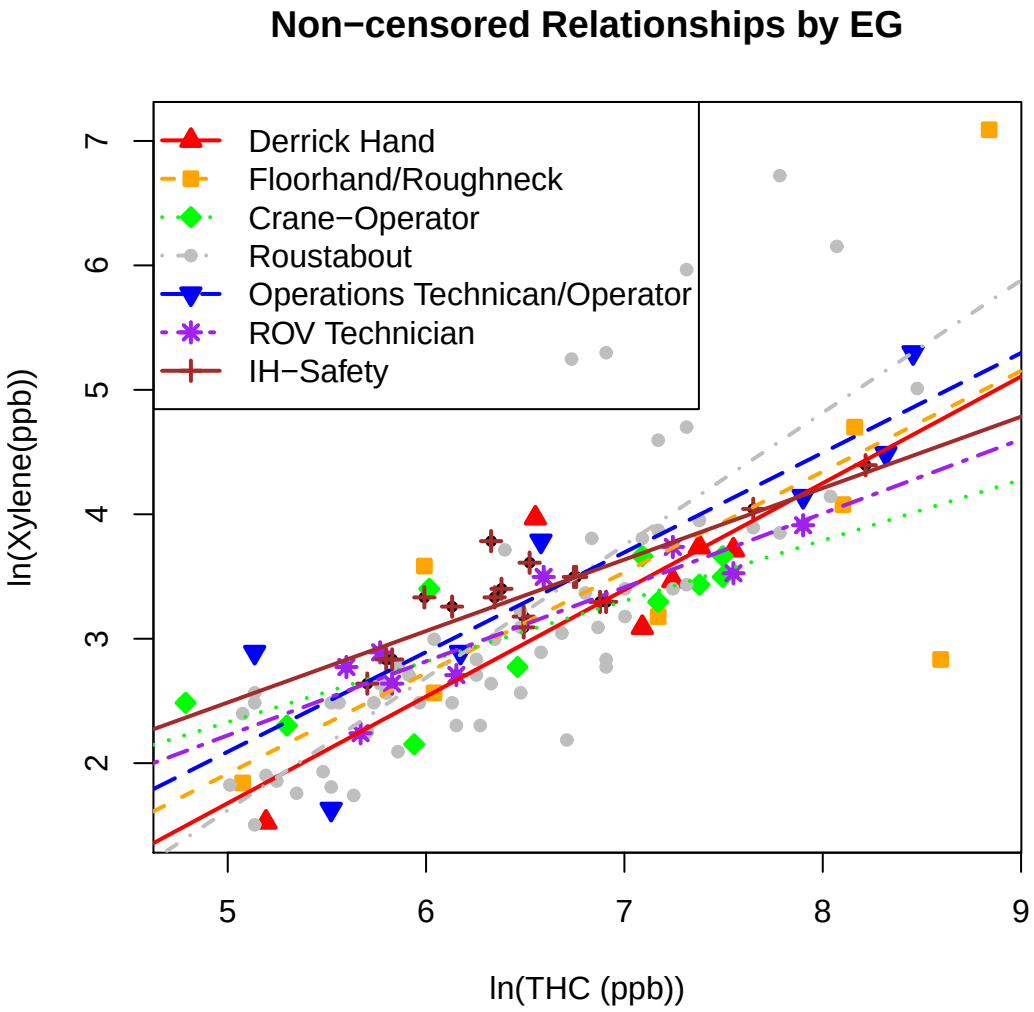

Figure 2: Non-censored relationships by EG on the *DD3* from May 15-July 15, 2010. The plot displays all the non-censored datapoints for each EG and the corresponding linear relationship for each of those non-censored EG datasets.