An assessment of occupational inhalation exposures to volatile oil components on four rig vessels for the GuLF STUDY

A DISSERTATION
SUBMITTED TO THE FACULTY OF
UNIVERSITY OF MINNESOTA
BY

Tran Bich Huynh

IN PARTIAL FULFILLMENT OF THE REQUIREMENTS
FOR THE DEGREE OF
DOCTOR OF PHILOSOPHY

Gurumurthy Ramachandran, Ph.D., Adviser
Sudipto Banerjee, Ph.D., Co-adviser
July, 2014

Acknowledgements

I wish to express my deep appreciation to my adviser, Professor Gurumurthy Ramachandran, for his gentle guidance and support, and for inviting me to work on a variety of interesting and challenging research projects during my graduate school years. His relaxing yet effective advising style has made it possible for me to enjoy the academic rigor of graduate school and also grow personally and professionally. I would also like to thank my co-adviser, Professor Sudipto Banerjee, for pushing me so hard to learn Bayesian statistics and programming (intentionally and unintentionally) and with the help from Joao Monterio, Harrison Quick, Caroline Groth, and Bao Le, I survived. Additional thanks are due to committee members – Drs. Bruce Alexander and Jeff Mandel- for their helpful suggestions and review of the dissertation.

I am indebted to Trish Stewart and Mark Stenzel for providing us this wonderful opportunity to work on this GuLF STUDY project. Their technical guidance on the direction of the research project, assistance with the editing of papers and providing the background information for the study as well as support were invaluable. Other research collaborators who made this work possible were Dale Sandler, Richard Kwok, Lawrence Engel, Aaron Blair, Wendy McDowell, and Caitlin Roush. This work was funded by the Intramural Research Program of the National Institute of Health, National Institute of Environmental Health Sciences.

Lastly, I can't thank my family enough for their unconditional love and support of my graduate school endeavors. I wouldn't have become who I am today without such loving family, and grandparents, aunts, uncles, big and small cousins, and little niece who are so much fun to hang out with.

Dedication

To my grandfather, Quang Nguyen, for his love of learning and for his commitment to the education of the family's future generations in America

Abstract

After the *Deepwater Horizon* oil release in the Gulf of Mexico, the National Institute of Environmental Health Sciences initiated an epidemiological study (the GuLF STUDY) to investigate the potential adverse health effects associated with the oil spill response and clean-up work. The exposure assessment is a critical component of the GuLF STUDY because it allows the investigation of the exposure-disease relationship. This involves the analysis of thousands of personal inhalation monitoring measurements that were collected by BP and its contractors during the entire remediation effort. A substantial portion of these data, however, has values below the limits of detection (LOD). This dissertation investigates various statistical methods for handling data with detection limits and presents the methodology and assessment of the inhalation exposures for workers on the four main rig vessels that were responsible for stopping the leak.

The first section of this dissertation evaluates three established classical (or 'frequentist') methods for analyzing data with censored observations to estimate the arithmetic mean (AM), geometric mean (GM), geometric standard deviation (GSD), and the 95th percentile ($X_{0.95}$) of the exposure distribution: the Maximum Likelihood (ML) Estimation, the β-substitution, and the Kaplan-Meier (K-M) methods. Each method was challenged with computer-generated exposure datasets drawn from lognormal and mixed lognormal distributions with sample sizes (N) varying from 5 to 100, GSDs ranging from 2 to 5, and censoring levels ranging from 10% to 90%, with single and multiple LODs. Using relative bias and relative root mean squared error (rMSE) as the evaluation metrics, the β-substitution method was found to generally perform as well or better than the ML and K-M methods in most simulated conditions. The ML method was suitable for large sample sizes (N ≥30) up to 80% censoring for lognormal distributions with small variability (GSD=2-3). The K-M method generally provided accurate estimates of the AM when the censoring was <50% for lognormal and mixed distributions.

The second section describes a Bayesian framework for analyzing censored data. Similar computer simulation was conducted to compare the β-substitution method with a Bayesian method. The Bayesian method using non-informative priors and the β-

substitution method were generally comparable in bias and rMSE when estimating the AM and GM. For the GSD and the 95th percentile, the Bayesian method with non-informative priors was more biased, and had a higher rMSE than the β-substitution method but the use of more informative priors generally improved the Bayesian method's performance, making both the bias and the rMSE more comparable to the β-substitution method. The advantage of the Bayesian method is that it allows the use of prior information and also provides estimates of uncertainty for all parameters (GM, GSD, and 95th percentile) whereas the β-substitution method only provides estimates of uncertainty for the AM.

The third chapter presents a methodology for assessing the occupational exposures and estimates of inhalation exposures for workers on the four main rig vessels (Enterprise, DD2, DD3, and Q4000) that were responsible for stopping the leak in the hot zone closest to the well site. Exposure groups (EGs) were created on based on chemicals, locations, vessels, time periods, and job titles/tasks. Bayesian method were used to analyzed exposures for total hydrocarbons (THCs), benzene, toluene, ethylbenzene, xylene (BTEX chemicals) and hexane. THC measurements were least censored compared other chemicals evaluated. THC exposures changed over time and varied by vessels and exposure groups. Highest exposures were generally observed in the time period before the well was successfully top capped. Exposures gradually decreased over time after top capping in most exposure groups except a few that might be involved in the decontamination effort. BTEX chemicals and hexane exposures were substantially lower than THC. The variability of the EGs for the GuLF STUDY were generally high, reflecting the non-routine, time-dependent nature of spill response efforts as well as the challenges of retrospectively constructing exposures for oil spill study.

Table of Contents

List of Tables

List of Figures

Chapter I

Chapter II

Chapter III

Chapter IV

Chapter I

Introduction

On April 20, 2010, the *Deepwater Horizon* oil drilling rig at the Macondo well exploded, killing 11 people and injuring 19 others. The damaged well subsequently released roughly 4.9 millions of barrels of crude oil into the Gulf of Mexico, the largest oil spill in the US history. Within weeks of the explosion, oil slick quickly reached the coastlines of Florida, Alabama, Mississippi, and Louisiana, covering marshes, river deltas, beaches and wildlife. The fire and subsequent oil release was caused by the rise of high-pressure methane gas into the oil rig where it ignited and a safety device called the blowout preventer (BOP) that failed to shut down the fluid flow. While the oil was gushing out of of the damaged well, BP tried several attempts to stop the leak but failed. BP successfully capped the BOP near the bottom of the sea floor and stopped the majority of the leak by July 15. The well was mostly sealed on September 19 and the majority of clean-up work lasted until December 30, 2010.

Containment efforts and remediation work offshore and on-land

The spill response and clean-up effort covered a vast geographical area which can be divided into five main locations: the hot zone (~ within mile radius from the well site), the source (within 5 mile radius from the well site, other than the hot zone), off shore (other than the source and hot zone), on shore (roughly three miles from shore), and land (Figure 1). The hot zone was a restricted area due high level of hydrocarbons. In the hot zone, four large rig vessels that were responsible for capturing oil and gas and capping the well were the Enterprise, Q4000, DD2, and DD3. There were also many smaller marine vessels (e.g., vessels with remotely operated vehicles (ROVs), oil recovery and research vessels) moving in and out of the hot zone to support the four rig vessels (e.g., mud pumping, oil processing, application of dispersant to the wellhead (~5000 feet below the sea surface), transfer of supplies and personnel to and from the four ships and transfer of crude oil and oily water collected at the source).

The Enterprise was first to arrive in the hot zone in early May and was stationed directly above the wellhead (Figure 2). While the Enterprise was collecting oil and gas, BP tried several attempts to contain the oil using remotely control vehicles (ROVs). First BP tried to activate the blowout preventer using the ROVs but failed. Next they placed a temporary containment dome over the BOP with a plan that the oil would be channeled through a pipe where it could be collected by a tanker. The dome's opening, however, was plugged by methane hydrate crystals. The entire dome had to be removed. They then inserted a riser insertion tube tool (four-inch diameter tube with large rubber diaphragms at the insertion end) into the Horizon's riser (21-inch diameter pipe) near the sea floor. Part of the flow was diverted through the riser and collected by the Enterprise on the surface but much of it continued to be released into the sea. In the mean time, BP attempted to 'top kill' the well by pumping various kinds of materials through the BOP but also failed. The next move was to fit a custom-made cap over the BOP by first removing the damaged riser from the BOP and then placing the cap over the BOP. The top capping stopped the majority of the leak by June 15. The Q4000 was brought in to perform the 'top kill' but from June 16 it was used to collect oil and gas using the equipment that was previously installed for the failed top kill (Figure 2). The Q4000 was also used to flare oil and gas. In the mean time, the DD3 started drilling the first relief well on May 2, and the DD2 on May 16 (Figure 3). On August 3, the Q4000 started the 'static kill' procedure by slowly pumping mud into the well using the equipment from the top kill. After the relief well successfully intercepted the Macondo well, BP pumped cement and permanently closed the well on Sept 19 (Cleveland, 2013).

While the four main rig ships, numerous oil containing vessels and supporting vessels tried to control flow in the hot zone, 835 skimmers and over 9000 vessels (including small boats called Vessels of Opportunities (VoOs)) were deployed to remove the oil on the water (USCG, 2011). Key off shore operations were in-situ burning, oil skimming, aerial and surface application of dispersants to break down the oil. Other activities off shore were large and small vessels and barges supplying fuel, equipment and personnel to the hot zone and carrying oil from the hot zone to land.

Near shore operations included these operations as well as scouting for oil on the beaches, marshes, and bayous, collecting contaminated wildlife, and cleaning rock jetties and other shoreline structures. Generally, dispersants were not applied near shore or on land. Land activities involved transport of workers by land vehicles, hand and mechanical cleaning of the sand, collection of oil and tar, and wildlife rehabilitation. Vessel decontamination occurred both offshore and on land. Most employees worked for contractors or sub-contractors to BP. Individual VoOs were hired to perform the supply and transfer functions, as well as providing assistance to boom operations, skimming and burning.

The GuLF STUDY

During the entire remediation period, more than 55,000 workers were involved in the response and clean-up work (NIOSH, 2011). Within weeks of the spill operations, a large number of response workers reported ill health symptoms including headaches, nausea, lower and upper respiratory irritations, heat stress, mental and physical fatigue, eye and skin irritation (NIOSH, 2010). As part of the comprehensive federal response the incident, the National Institute of Environmental Health Sciences (NIEHS) initated the Gulf Long-Term Follow-up Study for Oil Spill Clean-Up Workers and Volunteers (GuLF STUDY) to investigate the possible short-term (e.g., respiratory symptoms, nausea, headaches, dermatitis, depressive symptoms, anxiety, decreased lung function, and DNA damage) and long-term adverse health effects (e.g., cancer, neurological deficits, cardiovascular injury, reproductive effects) experienced by the workers (Sandler et al, 2012). The study recruited a cohort of almost 33,000 workers. Study participants were asked a series of questions about their health and clean-up jobs. A sub cohort was visited in the home for the administration of detailed questionnaires and collection of biological materials, anthropometric and physiological measurements and household dust.

The characterization of exposures is a critical component of any occupational and environmental epidemiological studies because it allows us to further investigate the exposure-disease relationship (Hill, 1965). The exposure assessment component of the GuLF STUDY is composed of three parts: occupational inhalation exposure, spatial/environmental exposure, and dermal exposure studies. The exposure estimates will then be linked to individual subjects in the epidemiological study to be carried out by NIEHS. Depending on the information provided by the worker (e.g., location of their work, jobs involved, potential skin/hand contact with the oil slick), individual worker's exposure can be the sum of the occupational, environmental and/or dermal exposures. This dissertation focuses on the assessment of occupational exposure via the inhalation route on the four rig vessels in the hot zone.

Health effects associated with oil spill clean-up

Initial health assessments by the National Institute for the Occupational Safety and Health (NIOSH) reported a variety of acute health effects such as eye, skin and upper respiratory tract irritations, headaches, dizziness, vomiting, and coughing experienced by the current oil spill clean-up workers (NIOSH, 2010). Human exposure assessment and epidemiological studies for past oil spills were few in the literature. There have been many major oil spills around world in the 20th century but human health studies have been conducted only for seven major spills (Anguilera et al., 2010). Most of these studies were cross-sectional studies and did not have quantitative estimates of chemical exposures (Palinkas et al., 1992,1993; Campbell et al., 1993, 1994; Crum, 1993; Lyons et al., 1999; Gallacher et al., 2007; Dor et al., 2003; Suarez et al., 2005; Carrasco et al., 2006, 2007;, Zock et al., 2007; Sabucedo et al., 2009; Janjua et al., 2006). In studies that had air monitoring results, few samples were collected for a very small number of contaminants during the spill (Campbell et al., 1993; Morita et al., 1999; Meo et al., 2008, Laffon et al., 2006; Perez-Cadahia et al., 2006, 2007). Nonetheless, these studies generally reported respiratory and dermal symptoms similar to those experienced during the BP oil spill clean up. Evidence of genotoxicity and endocrine toxicity has also been found in-vitro study (Amat-Bronnet et al, 2007) and epidemiological studies (Laffon et al., 2006, Perez-Cadahia et al., 2006, 2007, 2008a, 2008b, 2008c), in addition to bioaccumulation of oil compounds in marine food (Lemiere et al., 2005, Chaty et al, 2008). Psychological and social effects associated with oil spills were also common findings in human studies (Palinkas et al., 1992, 1993, 2004; Gill and Picou, 1998; Sabucedo et al., 2009; Gallacher et al. 2007, Zock et al., 2007; Janjua et al., 2006; Carrasco et al., 2006, 2007; Morita et al., 1999).

In addition to these findings, harmful acute and chronic health effects from BTEX (benzene, toluene, ethylbenzene, xylene) chemicals (components of the crude oil) are also well-known. The health effects range from irritation of the upper respiratory tract and eyes, sore throat, dizziness, and headache to cancer (benzene), developmental defects, kidney and liver damage (IRIS, 2003, 2005). Exposure to components of dispersants (2-

butoxyethanol and propylene glycol) is also known to cause short-term dermal and respirable irritation and long-term health effect such as hemolysis of red blood cells (NIOSH, 2010b).

Challenges in assessing occupational exposures for oil spill clean-up

Assessing exposures for oil spill study is very different from that of most occupational epidemiological studies. Oil spill response activities were often non-routine, dynamic, and highly time-dependent. The location of the work being performed covered a vast geographical area (e.g., off shore, on shore, vessels, and land). No single organization or entity was in charge. Even though BP was charged with the response, it hired many contractors, and contractors hiring subcontractors to perform the work. Governmental agencies were also involved with their own plans. Tasks also varied substantially ranging from maintaining day-to-day operations of rig vessels, drilling of relief wells, collecting oil and gas, operating the ROVs, oil skimming and burning to cleaning up the beaches marshes, rescuing wildlife, decontaminating equipment and vessels, and many more. Therefore, a systematic approach to monitoring the exposure to various chemical hazards for a variety of activities during a fast-paced spill response is often not feasible. During the spill cleanup, thousands of area and personal air monitoring samples were collected by BP and its contractors but those samples did not adequately cover all the activities that took place at various geographical locations and times. In addition, a majority of these samples are less than the analytic limits of detection (LOD), necessitating a special treatment of the data. Another consequence of working in the fast-paced environment is the lack of complete and accurate time-work records needed to classify workers into appropriate exposure groups. While these challenges (i.e., lack of a systematic exposure monitoring plan, incomplete or no monitoring data in many scenarios, highly censored data) are inherent problems in oil spills exposure studies, these problems were not addressed in previous studies. This research aims to address these challenges while developing accurate estimates of exposures for each worker.

Overview of our research approach

The overall strategy for estimating exposure for the GuLF STUDY is summarized in Figure 3. Preliminary exposure groups (EGs) were developed based on professional judgment and qualitative data from a number of sources including questionnaires, public records, site visits, interviews with workers, historical records, and measurements. Many EGs had high percentage of measurements less than the LOD. Computer simulation studies were carried out to identity a statistical method(s) that provided the most accurate estimate of the relevant parameters (i.e. the arithmetic mean, geometric mean, geometric standard deviation, and 95th percentile). If monitoring data had censored observations, the appropriate censored data analysis technique from the simulation study was selected for use. Otherwise, standard method was used to calculate descriptive statistics for completely uncensored datasets. Descriptive statistics and iterative statistical analysis was conducted to refine existing EGs or develop new ones (if needed). This dissertation focuses on the occupational exposure assessment on the four rig vessels which had a lot of monitoring data for many job titles or activities. A separate research effort will be carried out to assess exposures off shore, near shore, and on land. There were much less measurements collected for these areas because the locations were further from the leaking well. Many EGs in these areas had little or no monitoring data and thus occupational and environmental physical models could be used to estimate the exposures. Some of the estimates from the physical models for these EGs may be incorporated in the censored data method (the Bayesian method). The final 24-hour time weighted average will be computed by linking estimates from the monitoring data or exposure modeling data to the time-work history records to provide an exposure estimate for each worker for the GuLF STUDY.

Classification of exposure groups

Due to the lack of measurements for all workers, exposure groups were developed using the American Industrial Hygiene Association (AIHA) strategy to represent individuals with expected similar exposure levels (Mulhausen and Damiano, 2006). The criteria for defining these groups were based on chemical, location, vessel, time period, and job titles/tasks. These criteria were selected such that the EGs were as specific as possible and these groups covered all the tasks/activities evaluated.

The spill effort was divided into five locations: the hot zone (~ 1 mile radius around the wellhead), the source (~ 5 mile radius from the well site, excluding the hotzone) off shore other than the source, near shore (within site of the shore, i.e., ~3 miles), and land. In the hot zone, the Enterprise and the Q4000 collected oil and gas while the DD2 and DD3 drilled the relief wells. There were many other ships or boats in the hot zone that supported the four rig operations (e.g., mud pumping, oil processing, injecting dispersants at wellhead (approximately 5000 feet below the surface) and ROV vessels trying to repaired the well.

There were three significant events that likely affected the workers' exposures over time. On May 15, BP was approved to inject dispersants at the wellhead and applying dispersants on the surface. On the 15 of July, the well was successfully top capped and the majority of leaking stopped. At the beginning of August, the well was most sealed. The entire cleanup period, which lasted from April through September, was divided into four time period (TP) to correspond to these events: TP1a (April 20 - May 14 before dispersants injection), TP1b (May 15 - July 15 after dispersants application and before top capping), TP2 (July 16 – Aug 10 after top capping), TP3 (Aug 11 – December 30, 2010 after bottom capping).

BP provided a database comprising over 26,000 personal air samples that were analyzed for multiple chemicals (e.g., tetrahydrofuran, total hydrocarbons (THCs), benzene, toluene, ethylbenzene, xylene, hexane, heptane, cyclohexane, trimethylbenzenes, and other miscellaneous chemicals), which totaled to over ~155,000 measurements from April 27, 2010 – April 3, 2011. The crude oil is composed of a number of volatile chemicals but only six oil-related chemicals were selected in the study: THCs, benzene, toluene, ethylbenzene, xylene (BTEX chemicals), and hexane. These chemicals were selected because they had been associated with adverse health effects in previous spills studies and BTEX chemicals, in particularly, have been known to be harmful (IRIS, 2003 and 2005). In addition, they are more volatile and more likely to be inhaled by the workers. They also have the largest number of measurements collected over the entire clean-up period (Stewart et al., 2014).

An enormous amount of qualitative data from a number of sources were gathered by the exposure assessment team to so that team members learned about these jobs and processes as much as possible in order to calibrate their professional judgment for the classification process. Field notes with the personal measurements were extracted from BP databases. These time history reports contained worker's names, job title, vessels, sampling information (e.g., time on, off, sampling instrument), personal protective equipment (PPE) worn, and activities. Personals-on-board lists that documented personnel and their job titles on many large vessels were also obtained. Other sources including transportation lists that identified individuals who were transported by helicopter to vessels offshore, site visits to the four rig vessels and vessels of opportunities, numerous reports from governmental agencies, and exposure information from the questionnaire were also carefully reviewed (Stewart et al., 2014). Once preliminary exposure groups were established using these sources of qualitative information and professional judgments, iterative statistical analysis was conducted to refine and finalize the EGs.

As shown in Table 1, hundreds of personal measurements were collected on the four rig vessels. Despite of the large number of measurements, the exposures groups (after being defined by location, vessel, job title, and time period) in many cases the number of available measurements for specific exposure groups is small (e.g., <10) and many exposure groups have a high percentage of censored data (50% -100% for many groups). In addition, these measurements are often marked by high variability which is most likely due to the non-routine nature of activities and changes in these activities over time. The duration of the samples used in the STUDY varied from 4 hours to 18 hours, resulting in multiple LODs. The combination of small sample size, high percentage of censoring, high variability and multiple LODs presents many challenges to the estimation of the study subjects' exposures, including the need to identify a methodology for handling such heavily censored data sets.

Censored data analysis methods

A literature search reveals that there are a few statistical methods for handling data below the detection limits. Previous work by Helsel (2005; 2010), EFSA (2010), Hewett and Ganser (2007), and Ganser and Hewett (2010) together provide excellent discussions on censored data analysis (CDA) methods typically used in occupational and environmental exposure assessments

β-substitution method: This method has its roots in the popular substitution methods where each non-detectable measurement is replaced with LOD/2 or LOD/$\sqrt{2}$ (Hornung and Reed, 1990). Unlike the standard substitution methods where 2 or $\sqrt{2}$ is arbitrarily chosen, Ganser and Hewett (2010) developed an algorithm that computes a b-factor for adjusting the LOD (i.e., β-factor*LOD). The β-factor is derived from the calculation of the uncensored values of the dataset and it is then used to adjust the LOD. The β-factor varies depending on whether the AM or GM is being estimated. The GSD is estimated using the AM and GM, and the 95th percentile using the GM and GSD (See Appendix A for the derivation of the β-factor and R codes). The β-substitution assumes a lognormal distribution of the data. The method is relatively new compared to the ML or the K-M methods. The algorithm can be easily implemented in a simple spreadsheet or statistical software.

Maximum Likelihood (ML) Estimation Method: The ML method can be traced back to work by Fisher (1925) and Cohen (1959; 1961). Exposure data are log-transformed where μ = ln(GM) and σ = ln(GSD). The ML estimates are values of μ and σ that maximize the likelihood function.

$$Likelihood\ function\ \left(x_1, .. x_{n,} x_{n+1}, .. x_{n+m}\ |\mu, \sigma\right) = \prod_{i=1}^{n} PDF(x_i, \mu, \sigma) \prod_{i=n+1}^{m} CDF(LOD_i, \mu, \sigma)\ (1)$$

where n = number of detectable measurements, m = number of nondetectable measurements, x_i values = detects, LOD_i values = detection limits, PDF = normal probability density function, and CDF = normal cumulative distribution function. Most statistical programs have built-in optimization algorithms to solve this equation. There are several variations of ML methods; however, the difference in the performance of these methods was found to be minor (Hewett and Ganser, 2005). The standard ML method was used in this paper (Cohen, 1950, 1959) (See Appendix A for R codes).

Reverse Kaplan-Meier method: The K-M method is a non-parametric method that does not assume any underlying probability distribution of the data (Kaplan and Meier, 1958). Originally developed for analyzing right-censored survival data, the reverse algorithm was adapted to handle left-censored data. The K-M algorithm constructs a curve akin to an empirical CDF while adjusting for censoring (See Appendix A for the formula and R codes). If there are no censored values, the K-M curve is equivalent to the empirical CDF. Most statistical software packages (e.g., Minitab, SAS, or R) have procedures to calculate K-M estimators for right-censored survival analysis. Users can use the same procedures for left-censored data by 'flipping' the data (turning it from left-censored to right-censored) and then return it back to the original scale.

While some statistical methods for handling left-censored data are available, there is no consensus on the best method for any particular situation. For example, Helsel (2005) recommended the Kaplan-Meier (K-M) method, which does not assume any distributional shape of the data, over the Maximum Likelihood (ML) method for sample sizes of less than 50 and censoring less than 50%. Hewett and Ganser (2007), on the other hand, recommended the ML method over the K-M method for lognormal and mixed lognormal distributions based on their computer simulations. More recently, Ganser and Hewett (2010) developed the β-substitution method that was either comparable or superior to ML method in most simulated conditions, even for sample size of 5 to 20. However, the K-M method has not been compared with the β-substitution method. No method has been recommended for data with sample sizes of <5 or percent censoring of >80%. While other CDA methods exist, the literature review suggested that the β-substitution, the ML, and the K-M methods were the most promising candidates for further evaluation.

Bayesian approach

In addition to classical statistical methods, we also consider the Bayesian approach handling censored data. Bayesian inference is based on conditional probabilities through the use of Bayes' Theorem. A likelihood distribution of the data vector, **Y**, given a vector of model parameters, **θ**, is denoted by p(**Y** | **θ**). Bayesian inference combines p(**Y | θ)** with prior information in the form of the prior distribution for **θ**, denoted by p(**θ**). Inference is then made based on the posterior distribution, p(**θ** | **Y**), obtained via Bayes' Theorem:

$$p(\boldsymbol{\theta}|\boldsymbol{Y}) = \frac{p(\boldsymbol{Y}|\boldsymbol{\theta})p(\boldsymbol{\theta})}{p(\boldsymbol{Y})} = \frac{p(\boldsymbol{Y}|\boldsymbol{\theta})p(\boldsymbol{\theta})}{\int p(\boldsymbol{Y}|\boldsymbol{\theta})p(\boldsymbol{\theta})d\boldsymbol{\theta}} = C\ p(\boldsymbol{Y}|\boldsymbol{\theta})p(\boldsymbol{\theta})$$

$$\alpha\ \ p(\boldsymbol{Y}|\boldsymbol{\theta})p(\boldsymbol{\theta}) \qquad (1)$$

where p(**Y**) is the marginal, or unconditional, distribution of **Y** and "∝" denotes "proportional to". In practice, computing p(**Y**) is computationally expensive, but since it is not a function of our model parameters **θ**, it is simply a constant, denoted here by *C*. Since p(**θ** | **Y**) is a probability distribution, it must integrate to 1, thus the unknown value *C* is simply the constant that makes p(**θ** | **Y**) a valid distribution. As a result, it suffices to compute the posterior distribution as being proportional to the likelihood times the prior. More details about Bayesian methods can be found in Carlin and Louis (2009).

The Bayesian approach has several attractive features. One is the ability to incorporate prior information into the model. Prior information in the occupational hygiene field could come from a variety of sources including previously collected data, professional judgment, and mathematical models. In the GuLF STUDY, another source of prior information is the correlation between THC and each of the BTEX chemicals. For example, the concentration of a less censored chemical (THC) could be used to predict the concentration of BTEX chemicals and that information can be used as prior information for the Bayesian model. The other attractive feature is the ability to provide the full posterior distribution for the calculation of all model parameters (e.g., AM, GM, GSD, 95^{th} percentile). Using this posterior distribution, we can obtain point estimates, such as the posterior median, as well as the 95% credible intervals.

Research objectives

The goal of this research is to develop accurate estimates of occupational exposures via the inhalation route to support the GuLF STUDY. The scope of this dissertation limits to assessing inhalation exposures on the four rig vessels in the hot zone. This goal can be achieved by fulfilling the following aims.

1. Evaluate classical statistical methods for analyzing data with detection limits

The goal is to identify a classical method that will provide accurate estimate of the arithmetic mean (AM), geometric mean (GM), geometric standard deviation (GSD), and 95th percentile ($X_{0.95}$). Chapter II describes the background of the methods evaluated (β-substitution, ML, and K-M methods), the computer simulation design and metrics used to compare these methods. The β-substitution appeared to be the best methods of the three methods in most of our simulated scenarios. The manuscript of this work has been accepted for publication in the Annals of Occupational Hygiene.

2. Evaluate a Bayesian method for analyzing data with detection limits

Chapter III explores the Bayesian method for handling data with detection limits and it compares the performances the Bayesian method and β-substitution method using similar simulation design and metrics as in Chapter II. The Bayesian method was considered because we prior information could potentially be used to improve the model's estimates in scenarios that have small sample sizes or high level of censoring. The manuscript for this research will be submitted to the Annals of Occupational Hygiene.

3. Develop estimates of occupational inhalation exposures on the four rig vessels in the hot zone.

Chapter IV reports the exposure estimates on the four rig vessels in the hot zone that will be used in the GuLF STUDY. It also describes the methodology for classifying exposure groups, the data collection process, the Bayesian censored data analysis model, and priors derived from the correlation of the chemicals that were used in the Bayesian model. The manuscript of this work will be submitted to the Annals of Occupational Hygiene.

References

Amat-Bronnert A, Castegnaro M, Pfohl-Leszkowicz A. (2007) Genotoxic activity and induction of biotransformation enzymes in two human cell lines after treatment by Erika fuel extract. Environ. Toxicol. Phar- macol. 23: 89–95

Aguilera F, Méndez J, Pásaro E, Laffon B. (2010). Review on the effects of exposure to spilled oils on human health. *Journal of applied toxicology : JAT*, *30*(4), 291–301.

Campbell D, Cox D, Crum J, Foster K, Christie P. (1993) Initial effects ofthe grounding of the tanker Braer on health in Shetland. *British Medical Journal*, *307*, 1251–1255.

Campbell, D., Cox, D., Crum, J., Foster, K., Riley, A., Manfredini, R., Gallerani, M., et al. (1994). Later effects of grounding of tanker Braer on health in Shedland. *British Medical Journal*, *309*, 773–774.

Carrasco, J. M., Lope, V., Pérez-Gómez, B., Aragonés, N., Suárez, B., López-Abente, G., Rodríguez-Artalejo, F., et al. (2006). Association between health information, use of protective devices and occurrence of acute health problems in the Prestige oil spill clean-up in Asturias and Cantabria (Spain): a cross-sectional study. *BMC Public Health*, *6*

Carrasco, J. M., Pérez-Gómez, B., García-Mendizábal, M. J., Lope, V., Aragonés, N., Forjaz, M. J., Guallar-Castillón, P., et al. (2007). Health-related quality of life and mental health in the medium-term aftermath of the Prestige oil spill in Galiza (Spain): a cross-sectional study. *BMC public health*, *7*, 245.

Carlin, B.P. and Louis, T.A. (2009). Bayesian Methods for Data Analysis, 3rd ed. Boca Raton, FL: Chapman and Hall/CRC Press.

Chaty S, Rodius F, Lanhers M-C, Burnel D, Vasseur P. 2008. Induction of CYP1A1 in rat liver after ingestion of mussels contaminated by Erika fuel oils. Arch. Toxicol. 82: 75–80.

Cleveland, C. (2013). Deepwater Horizon oil spill . Retrieved from http://www.eoearth.org/view/article/161185

Cohen, AC. (1959) Simplified estimators for normal distribution when samples are singly censored or truncated. Technometrics 1, 217-37.

Cohen, AC. (1961) Tables for maximum likelihood estimates: Singly truncated and singly censored samples. Technometrics 3, 535-541.

Crum, J. E. (1993). Peak expiratory flow rate in schoolchildren living close to Braer oil spill. *British Medical Journal*, *307*, 23–24.
Dor, F., Bonnard, R., Gourier-Fréry, C., Cicolella, A., Dujardin, R., & Zmirou, D. (2003). Health risk assessment after decontamination of the beaches polluted by the wrecked ERIKA tanker. *Risk analysis : an official publication of the Society for Risk Analysis*, *23*(6), 1199–208.
Draxler, R.R., Hess G.D. (1998). An overview of the HYSPLIT_V modeling system for trajectories, dispersion, deposition,. Aust. Meteor. Mag, 47, 295-308.
European Food Safety Authority. (2010). Management of left-censored data in dietary exposure assessment of chemical substances. *EFSA Journal*, *8*(3), 1–96. Available at: www.efsa.europa.eu.
Finkelstein MM, and Verma DK. (2001) Exposure estimation in the presence of nondetectable values: another look. Am Ind Hyg Assoc J; 62:195-8.
Fisher RA. (1925) Theory of statistical estimation. Proceedings of the Cambridge Philosophical Society 22, 700-725.
Fingas MF, Halley G, Ackerman F, Nelson R, Bissonnette M, Laroche N, Wang Z, et al. (1993). *The Newfoundland offshore burn experiment (NOBE)*. Ottawa, Ontario CA.
Gallacher, J., Bronstering, K., Palmer, S., Fone, D., & Lyons, R. (2007). Symptomatology attributable to psychological exposure to a chemical incident: a natural experiment. *Journal of epidemiology and community health*, *61*(6), 506–12.
Ganser, G. H., & Hewett, P. (2010). An accurate substitution method for analyzing censored data. *Journal of occupational and environmental hygiene*, *7*(4), 233–44.
Gelman, A., Carlin, J.B., Stern H.S. and Rubin D.B. (2004). Bayesian Data Analysis. Chapman and Hall/CRC, Boca Raton.
Gill D, Picou J. (1998) Technological disaster and chronic community stress. Soc. Natur. Resour. 11: 795–815
Hill AB. 1965.The evaluation of disease: association or causation? Proc R Soc Med 1965 58:295-300.
Hornung RW, and Reed LD. (1990) Estimation of average concentration in the presence of non-detectable values. *Appl. Occup. Envir. Hyg.* 5(1):46–51.

Helsel DR. (2005) Nondetects and data analysis. New York: John Wiley & Sons, Inc.

Helsel, D. (2010). Much ado about next to nothing: incorporating nondetects in science. *The Annals of occupational hygiene*, *54*(3), 257–62.

Hewett, P., & Ganser, G. H. (2007). A comparison of several methods for analyzing censored data. *The Annals of occupational hygiene*, *51*(7), 611–32.

Janjua, N. Z., Kasi, P. M., Nawaz, H., Farooqui, S. Z., Khuwaja, U. B., Najam-ul-Hassan, Jafri, S. N., et al. (2006). Acute health effects of the Tasman Spirit oil spill on residents of Karachi, Pakistan. *BMC public health*, *6*, 84.

Kaplan, E. L., & Meier, P. (1958). Nonparametric Estimation from Incomplete Observations. *Journal of the American Statistical Association*, *53*(282), 457.

Laffon, B., Fraga-Iriso, R., Pérez-Cadahía, B., & Méndez, J. (2006). Genotoxicity associated to exposure to Prestige oil during autopsies and cleaning of oil-contaminated birds. *Food and chemical toxicology : an international journal published for the British Industrial Biological Research Association*, *44*(10), 1714–23.

Lemiere S, Cossu-Leguille, Bispo A, Jourdain MJ, Lanhers MC, Burnel D, Vasseur P. (2005) DNA damage measured by the single-cell gel elec- trophoresis (comet) assay in mammals fed with mussels contami- nated by the 'Erika' oil-spill. Mutat. Res. 581: 11–21.

Lyons, R. Temple, J. M., Evans, D., Fone, D. L., & Palmer, S. R. (1999). Acute health effects of the Sea Empress oil spill. *Journal of epidemiology and community health*, *53*(5), 306–10.

Meo, S. A., Al-Drees, A. M., Rasheed, S., Meo, I. M., Al-Saadi, M. M., Ghani, H. a, & Alkandari, J. R. (2009). Health complaints among subjects involved in oil cleanup operations during oil spillage from a Greek tanker "Tasman Spirit". *International journal of occupational medicine and environmental health*, *22*(2), 143–8.

Morita, a, Kusaka, Y., Deguchi, Y., Moriuchi, a, Nakanaga, Y., Iki, M., Miyazaki, S., et al. (1999). Acute health problems among the people engaged in the cleanup of the Nakhodka oil spill. *Environmental research*, *81*(3), 185–94.

Mulhausen, J., and J. Damiano: Chapter 4, Establishing Similar Exposure Groups. In: A Strategy for Managing Occupational Exposures. J.S. Ignacio, W. H. Bullock, (Eds), Third Edition, AIHA Press, Fairfax, VA, p. 33-46 (2006).

National Institute for Occupational Safety and Health (NIOSH) (2010a). *Health Hazard Evaluation of Deepwater Horizon Response Workers*. Retrieved from http://www.cdc.gov/niosh/h.he/reports/pdfs/2010-0115-0129-3138.pdf

National Institute for Occupational Safety and Health. (2010b). Reducing Occupational Exposures while Working with Dispersants During the Deepwater Horizon Response. Retrieved from http://www.cdc.gov/niosh/topics/oilspillresponse/dispersants.html

National Institute for Occupational Safety and Health. (2011) NIOSH Deepwater Horizon Roster Summary Report. Retrieved from http://www.cdc.gov/niosh/docs/2011-175/pdfs/2011-175.pdf

Palinkas AL, Russell J, Downs AM, Petterson, J. (1992). Ethnic differences in stress, coping, and depressive symptoms after the Exxon Valdez oil spill. *Journal of Nervous and Mental Disease*, *180*(5), 287–295.

Palinkas, L., Petterson, J., Russell, J., & Downs, M. (1993). Community patterns of psychiatric disorders after the Exxon Valdez oil spill. *American Journal of Psychiatry*, *150*(10), 1517–1523.

Palinkas LA, Petterson JS, Russell J, DownsMA (2004) Ethnic differences in symptomsof post-traumatic stress after the Exxon Valdez oil spill.PDM 19: 102–112.

Pérez-Cadahía, B., Laffon, B., Pásaro, E., & Méndez, J. (2006). Genetic damage induced by accidental environmental pollutants. *TheScientificWorldJournal*, *6*, 1221–37.

Pérez-Cadahía, B., Lafuente, A., Cabaleiro, T., Pásaro, E., Méndez, J., & Laffon, B. (2007). Initial study on the effects of Prestige oil on human health. *Environment international*, *33*(2), 176–85.

Pérez-Cadahía B, Méndez J, Pásaro E, Lafuente A, Cabaleiro T, Laffon B. 2008a. Biomonitoring of human exposure to Prestige oil: Effects on DNA and endocrine parameters. Environ. Health Insights 2: 83–92.

Pérez-Cadahía B, Laffon B, Porta M, Lafuente A, Cabaleiro T, López T, Caride A, Pumarega J, Romero A, Pásaro E,Méndez J. 2008b. Relationship between blood concentrations of heavy metals and cytogenetic and endocrine parameters among subjects involved in cleaning coastal areas affected by the 'Prestige' tanker oil spill. Chemosphere 71: 447–455.

Pérez-Cadahía B, Laffon B, Valdiglesias V, Pásaro E,Méndez J. 2008c. Cyto- genetic effects inducedby Prestige oil onhumanpopulations:The role of polymorphisms in genes involved in metabolism and DNA repair. Mutat. Res. 653: 117–123.

Rappaport, S.M. (1991). Assessment of long-term exposures to toxic substances in air – Review. *Ann of Occup Hyg*, 35:61-121.

R Development Core Team. (2011) R: a language and environment for statistical computing. R Foundation for Statistical Computing: Vienna. http://www.R-project.org.

Sabucedo, J. M., Arce, C., Ferraces, M. J., & Merino, H. (2009). Psychological impact of the Prestige catastrophe. *International Journal of Clinical and Health Psychology*, *9*, 105–116.

Sandler DP, Kwok RK, Engel LS, Parks C, London SJ, Miller AK, Blair A, Stenzel M, Suárez, B., Lope, V., Pérez-Gómez, B., Aragonés, N., Rodríguez-Artalejo, F., Marqués, F., Guzmán, a, et al. (2005). Acute health problems among subjects involved in the cleanup operation following the Prestige oil spill in Asturias and Cantabria (Spain). *Environmental research*, *99*(3), 413–24.

SkyTruth. 2010. BP / Gulf Oil Spill - Cumulative Oil Slick Footprints. Retrieved from http://blog.skytruth.org/2010/09/bp-gulf-oil-spill-cumulative-oil-slick.html

Stewart PA, Ramachandran G, Sudipto B, et al. (2014) The NIEHS GuLF STUDY: Overview of the assessment process for estimating exposures to volatile oil-related substances for subjects on the water (in press).

US Coast Guard (USCG). (2011). *On Scene Coordinator Report Deepwater Horizon Oil Spill*. Retrieved from http://www.uscg.mil/foia/docs/dwh/fosc_dwh_report.pdf

U.S. Environmental Protection Agency. *Integrated Risk Information System (IRIS) on Xylene*. National Center for Environmental Assessment, Office of Research and Development, Washington, DC. 2003.

U.S. Environmental Protection Agency. *Integrated Risk Information System (IRIS) on Benzene*. National Center for Environmental Assessment, Office of Research and Development, Washington, DC. 2003.

U.S. Environmental Protection Agency. *Integrated Risk Information System (IRIS) on Toluene*. National Center for Environmental Assessment, Office of Research and Development, Washington, DC. 2005.

U.S. Environmental Protection Agency (USEPA). (2010) . Questions and Answers on Dispersants. Retrieved from http://www.epa.gov/bpspill/dispersants-qanda.html#appl.

U.S. Environmental Protection Agency (USEPA). (2007) ProUCL Version 4.0 Technical Guide. EPA/600/R-07/0

U.S. Coast Guards. On Scene Coordinator Report on Deepwater Horizon Oil Spill(Report). September 2011. Retrieved from http://www.uscg.mil/foia/docs/dwh/fosc_dwh_report.pdf

Zock, J.-P., Rodríguez-Trigo, G., Pozo-Rodríguez, F., Barberà, J. a, Bouso, L., Torralba, Y., Antó, J. M., et al. (2007). Prolonged respiratory symptoms in clean-up workers of the prestige oil spill. *American journal of respiratory and critical care medicine*, *176*(6), 610–6.

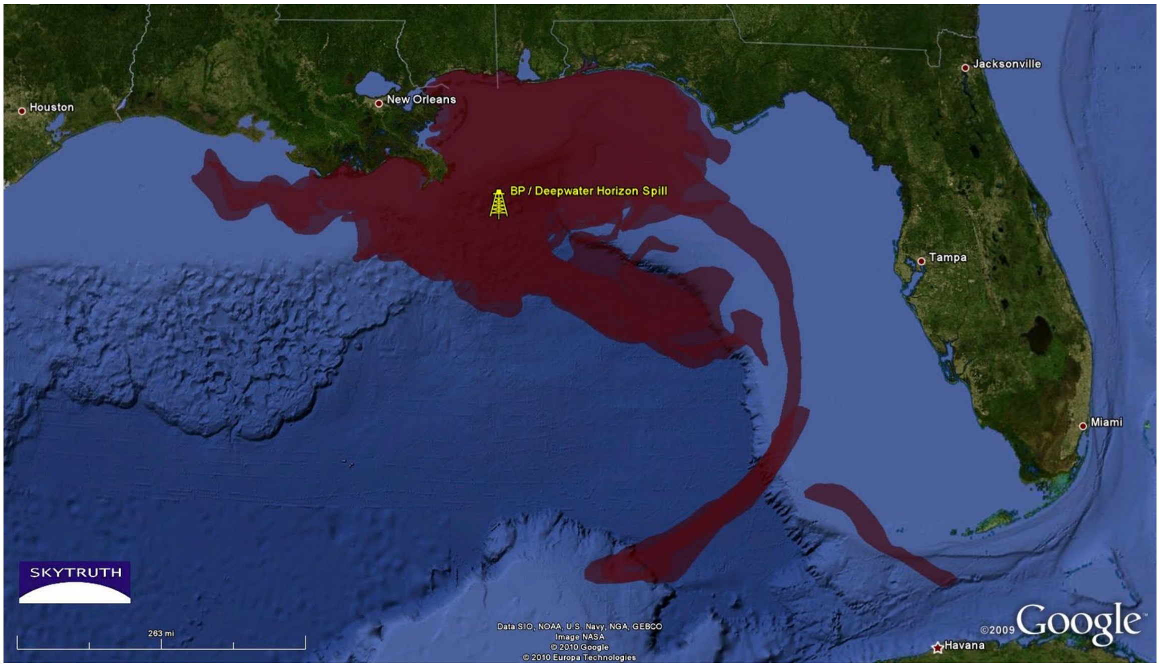

Figure 1: Cumulative *Deepwater Horizon* oil slick footprint (red) through mid July, 2010. Source: SKYTRUTH.

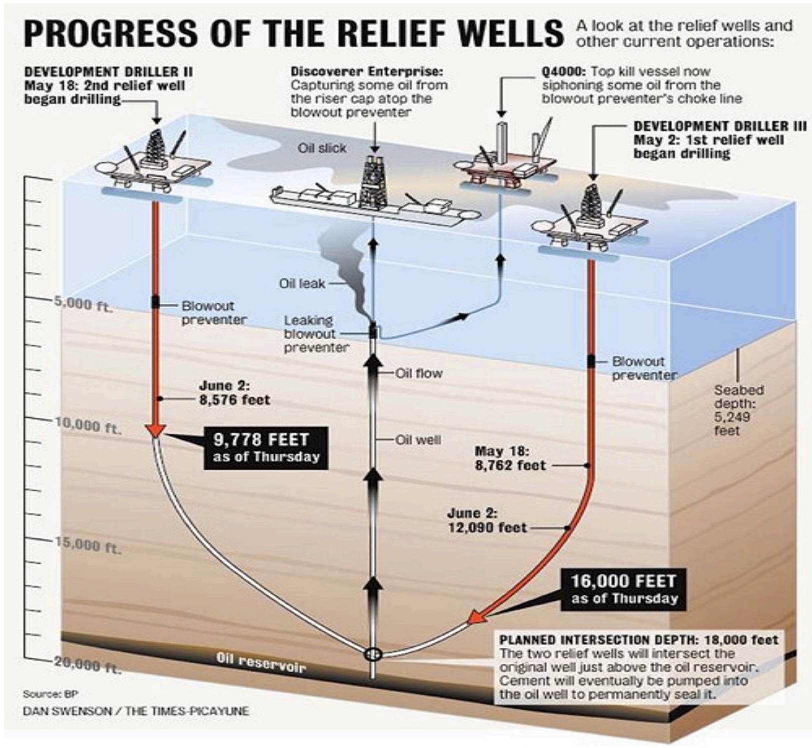

Figure 2: The location and function of the rig ships in the hot zone. Source: BP/DAN SWENSON/TIMES PICAYUNE

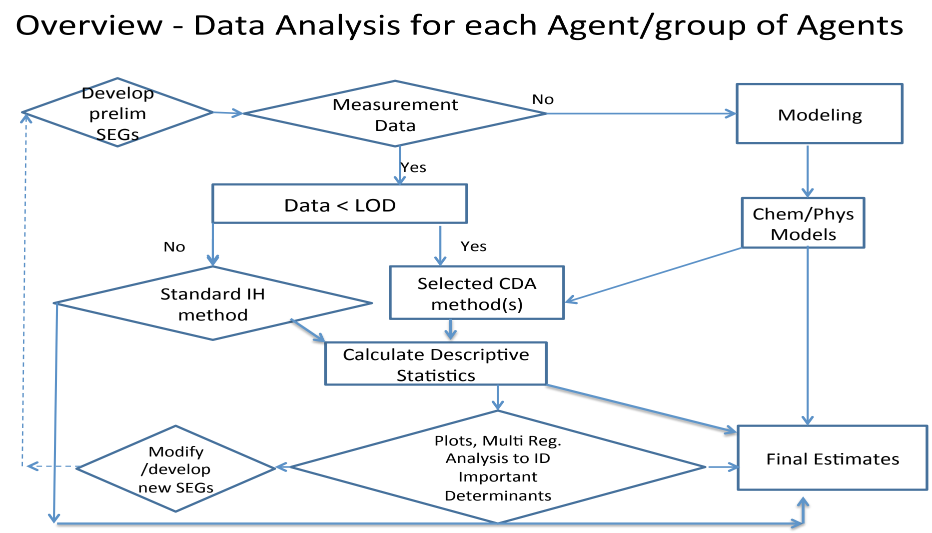

Figure 3: The overall strategy for estimating exposures for the GuLF STUDY

Table 1: Number of measurements and percent censoring for each vessel and agent

Vessel	N	% < LOD					N	%<LOD
		THC	Benzene	Toluene	Ethylbenzene	Xylene	Hexane	Hexane
DD2	339	34	96	58	61	50	200	75
Enterprise	436	11	87	36	58	41	274	22
DD3	449	30	95	24	37	34	287	45
Q4000	207	20	92	68	63	39	197	67

Chapter II
Comparison of methods for analyzing left-censored occupational exposure data

Tran Huynh, Gurumurthy Ramachandran, Sudipto Banerjee, Joao Monteiro, Mark Stenzel, Dale P. Sandler, Lawrence S. Engel, Richard K. Kwok, Aaron Blair, and Patricia A. Stewart

Accepted for publication in the Annals of Occupational Hygiene, 2014

Summary

The National Institute for Environmental Health Sciences (NIEHS) is conducting an epidemiologic study (GuLF STUDY) to investigate the potential adverse health effects associated with clean-up of the *Deepwater Horizon* oil release. The exposure assessment component of the study involves analyzing thousands of personal monitoring measurements that were collected during this effort. A substantial portion of these data has values reported by the analytic laboratories to be below the limits of detection (LOD). A simulation study was conducted to evaluate three established methods for analyzing data with censored observations to estimate the arithmetic mean (AM), geometric mean (GM), geometric standard deviation (GSD), and the 95th percentile ($X_{0.95}$) of the exposure distribution: the Maximum Likelihood (ML) Estimation, the β-substitution, and the Kaplan-Meier (K-M) methods. Each method was challenged with computer-generated exposure datasets drawn from lognormal and mixed lognormal distributions with sample sizes (N) varying from 5 to 100, GSDs ranging from 2 to 5, and censoring levels ranging from 10% to 90%, with single and multiple LODs. Using relative bias and relative root mean squared error (rMSE) as the evaluation metrics, the β-substitution method generally performed as well or better than the ML and K-M methods in most simulated lognormal and mixed lognormal distribution conditions. The ML method was suitable for large sample sizes (N ≥30) up to 80% censoring for lognormal distributions with small variability (GSD=2-3). The K-M method generally provided accurate estimates of the AM when the censoring was <50% for lognormal and mixed distributions. The accuracy and precision of all methods decreased under high variability (GSD=4 and 5) and small sample sizes (N<20) but the β-substitution was still the best of the three methods. When using the ML method, practitioners are cautioned to be aware of different ways of estimating the AM as they could lead to biased interpretation. A limitation of the β-substitution method is the absence of a confidence interval for the estimate. More research is needed to develop methods that could improve the estimation accuracy for small sample sizes and high percent censored data and also provide uncertainty estimates.

Introduction

It is estimated that more than 55,000 workers were rostered in the response and clean-up of the oil release from the *Deepwater Horizon* rig explosion that occurred on April 20, 2010 (NIOSH, 2011). As part of the comprehensive federal response to this effort, the National Institute of Environmental Health Sciences (NIEHS) initiated an epidemiological study (GuLF STUDY) to assess possible adverse health effects associated with exposures from multiple agents to the study subjects who participated in the response and cleanup work.

Exposure assessment is a critical component in the investigation of exposure-response relationship and a key criterion used to establish causality (Hill, 1965). During the response and cleanup, personal inhalation exposures were measured by BP, its contractors and governmental agencies. Over 150,000 personal exposure measurements for an array of contaminants were collected; however, a substantial number of these measurements was below the limits of detection (LOD) reported by the analytic laboratories, or left-censored (Type I censoring). These measurements are being used to characterize exposure levels for specific exposure groups defined by factors such as location, vessel, job title, and time period. Despite the large number of measurements, in many cases the number of available measurements for specific exposure groups is small (e.g., <10) and many exposure groups have a high percentage of censored data (50% - 100% for many groups). In addition, these measurements are often marked by high variability probably due to the non-routine nature of some activities and changes in these activities over time. The duration of the samples used in the STUDY varied from 4 hours to 18 hours, resulting in multiple LODs. The combination of small sample size, high percentage of censoring, high variability and multiple LODs presents many challenges to the estimation of the study subjects' exposures, including the need to identify a methodology for handling such heavily censored data sets.

Previous work by Helsel (2005; 2010), EFSA (2010), Hewett and Ganser (2007), and Ganser and Hewett (2010) together provide excellent discussions on censored data analysis (CDA) methods typically used in occupational and environmental exposure assessments. Other methods related to epidemiological studies include the multiple imputation approach (Lubin, 2004) and a variant of the K-M method, a Cox-regression-based method that was used to assess biomarkers and adverse health effects (Dinse et al, 2014). The general consensus is that all of these methods are better options that the standard substitution method (e.g., LOD/$\sqrt{2}$). While some statistical methods for handling left-censored data are available, there is no consensus on the best method for particular situations. For example, Helsel (2005) recommended the Kaplan-Meier (K-M) method, which does not assume any distributional shape of the data, over the Maximum Likelihood (ML) method for sample sizes of less than 50 and censoring less than 50%. Hewett and Ganser (2007), on the other hand, recommended the ML method over the K-M method for lognormal and mixed lognormal distributions based on their computer simulations. More recently, Ganser and Hewett (2010) developed the β-substitution method that was either comparable or superior to ML method in most simulated conditions, even for sample size of 5 to 20. However, the K-M method has not been compared with the β-substitution method. No method has been recommended for data with sample sizes of <5 or percent censoring of >80%. While other CDA methods exist, our literature review suggested that the β-substitution, the ML, and the K-M methods were the most promising candidates for further evaluation.

We evaluated these methods for estimating the arithmetic mean (AM), the 95th percentile ($X_{0.95}$), the geometric mean (GM), and the geometric standard deviation (GSD). In occupational epidemiology studies, the AM is generally considered the most appropriate metric for calculating cumulative exposure (Seixas et al, 1988; Rappaport, 1991) while the $X_{0.95}$ is a useful estimate of the upper bound of a distribution of full-shift exposures. The three candidate methods were compared in a simulation study to identify the least biased method in the estimation of the AM, $X_{0.95}$, GM, and GSD over a wide range of sample sizes (N), variability, and censoring for lognormal and mixed lognormal distributions with a single LOD and with multiple LODs.

Background

β-substitution method: This method has its roots in the popular substitution methods where each non-detectable measurement is replaced with LOD/2 or LOD/$\sqrt{2}$ (Hornung and Reed, 1990). Unlike the standard substitution methods where 2 or $\sqrt{2}$ is arbitrarily chosen, Ganser and Hewett (2010) developed an algorithm that computes a β-factor for adjusting the LOD (i.e. β-factor*LOD). The β-substitution method consists of a series of calculations that estimates the β-factor to substitute the LOD. The β-factor varies depending on whether the AM or GM is being estimated such that bias is minimized. The AM is computed from the imputed dataset where the LOD is substituted with LOD*β-mean and the GM from dataset where the LOD is substituted with LOD*β-GM. The formulas for estimating the GSD and the 95th percentile are also modified where GSD is estimated using the AM and GM, and the 95th percentile using the GM and intermediate variables. The β-substitution method assumes a lognormal distribution of the data. The method is relatively new compared to the ML or the K-M methods. The algorithm can be easily implemented in a simple spreadsheet or statistical software. Interested readers are encouraged to explore the original paper by Ganser and Hewett (2010) to see the entire algorithm and its derivation.

Maximum Likelihood (ML) Estimation Method: The ML method can be traced back to work by Fisher (1925) and Cohen (1959; 1961). Exposure data are log-transformed where μ = ln(GM) and σ = ln(GSD). The ML estimates are values of μ and σ that maximize the likelihood function.

$$Likelihood\ function\ (x_1, .. x_{n,} x_{n+1}, .. x_{n+m}\ |\mu, \sigma)$$
$$= \prod_{i=1}^{n} PDF(x_i, \mu, \sigma) \prod_{i=n+1}^{m} CDF(LOD_i, \mu, \sigma)\ (1)$$

where n = number of detectable measurements, m = number of nondetectable measurements, x_i values = detectable measurements, LOD_i values = detection limits, PDF = normal probability density function, and CDF = normal cumulative distribution function. Most statistical programs have built-in optimization algorithms to solve this equation. There are several variations of ML methods; however, the difference in the performance of these methods was found to be minor (Hewett and Ganser, 2005). In this paper, we used a standard ML method (Cohen, 1950, 1959).

In our review of the occupational hygiene literature concerning censored data analysis, we found two approaches for estimating the AM from lognormal data. One is the maximum likelihood estimator (MLE) (Cohen, 1950 and 1959) and the other is the minimum variance unbiased estimator (MVUE) (Finney, 1946; Aitchison and Brown, 1976), which we will now name as $ML_{MLE\text{-}AM}$ and $ML_{MVUE\text{-}AM}$, respectively. The maximum likelihood estimate of the AM is:

$$\widehat{AM}_{\mathrm{MLE=}} = \exp\left(\hat{\mu} + \frac{1}{2}\hat{\sigma}\right) \quad (2)$$

where $\hat{\mu}$ and $\hat{\sigma}$ are the maximum likelihood estimates of the parameters μ and σ obtained by

by maximizing the likelihood function for the log-transformed censored data. The maximum likelihood estimates are asymptotically unbiased and efficient, thus the ML method is generally recommended for moderate to large samples (Helsel, 2005; Hewett and Ganser, 2007; Krishnamoothy et al, 2009)

The minimum variance unbiased estimator of the AM is:

$$\widehat{AM}_{\mathrm{MVUE}} = \exp(\hat{\mu}_y)\psi\left(\frac{\hat{\sigma}^2}{2}\right) \quad (3)$$

where

$$\psi(g) = \left[1 + \frac{(n-1)g}{n} + \frac{(n-1)^3 g^2}{n^2(n+1)2!} + \frac{(n-1)^5 g^3}{n^3(n+1)(n+3)3!} + \frac{(n-1)^7 g^4}{n^4(n+1)(n+3)(n+5)4!} + \cdots\right]$$

In this expression, $\hat{\mu}$ and $\hat{\sigma}$ are the sample mean and standard deviation, $g=\frac{\hat{\sigma}^2}{2}$, and n=sample size. Calculation of the first five terms of this infinite sum usually provide sufficiently precise estimates (Hewett and Ganser, 1997).

The minimum variance unbiased estimator of the AM is typically used for noncensored data with small sample sizes. However, Hewett and Ganser (2007, 2010) have found that $ML_{MVUE\text{-}AM}$ can also provided accurate estimate of AM using ML estimates of μ and σ for censored data in place of the sample mean and standard deviation as it helps to correct the bias that occurs from transforming estimates from the logarithmic scale back to the regular concentration scale. These two computations for the AM of the ML method do not apply for computation of the $X_{0.95}$, GM and GSD.

Reverse Kaplan-Meier method: The K-M method is a non-parametric method that does not assume any underlying probability distribution of the data (Kaplan and Meier, 1958). Originally developed for analyzing right-censored survival data, the reverse algorithm was adapted to handle left-censored data. The K-M algorithm constructs a curve akin to an empirical CDF where it assigns a probability to each of the ranked detectable measurements while adjusting for censoring. If there are no censored values, the K-M curve is equivalent to the empirical CDF. Most statistical software packages (e.g., Minitab, SAS, or R) have procedures to calculate K-M estimators for right-censored survival analysis. Users can use the same procedures for left-censored data by 'flipping' the data (turning it from left-censored to right-censored) and then returning it back to the original scale after the probabilities have been computed. For the $X_{0.95}$ calculation, we used the algorithm denoted as Q6 in Hewett and Ganser (2010): 1) sort the data from low to high; 2) calculate i = integer portion of 0.95(n+1); 3) $\hat{X}_{0.95} = x_i + (0.95(n+1) - i)(x_{i+1} - x_i)$. The required minimum sample size of computing the $X_{0.95}$ is 20. We used the reverse K-M algorithm published in the US EPA ProcUCL 4.0 Software Technical Guide (U.S. EPA, 2007) and examples by Beal (2010) for our simulation study.

Methods

Simulation design

The accuracy of an estimation method to treat censored data depends on the assumed distributional shape of the data, the sample size, the degree of censoring, and the variability of the data. To assess the effect of each of these conditions, we generated simulated censored data sets from lognormal distributions and from mixed lognormal distributions for varying sample sizes (N), degrees of censoring, and GSDs with a single and with multiple LODs.

Figure 1 summarizes our simulation design for the three types of simulations that were conducted for this study. In Simulation 1, uncensored values of various sizes were randomly drawn from lognormal distributions with a true GM=1 and true GSDs =2, 3, 4, and 5. Sample sizes were 5, 10 and then were incrementally increased in steps of 10 up to 100. The target censoring was first fixed at 10% and was then incrementally increased in steps of 10% to 90%. We then selected a LOD value from each distribution that corresponded to the expected censoring level. For example, if the percent censoring was p, then LOD = X_p, where X_p is the value at the p percentile. Values that were less than or equal to the LOD were censored to create each final dataset. For each combination of N, GM, GSD, and expected percent censoring, 1000 datasets were generated and the four parameters (AM, GM, GSD, $X_{0.95}$) were estimated using the β-substitution, the ML, and the K-M methods.

Simulation 2 was similar to Simulation 1 except that three LODs were generated. LOD_1 was set equal to the value at the percentile p_1 of the distribution that corresponded to the expected level of censoring. LOD_2 was at the percentile that was 5% less than p_1 (i.e. p_1*0.95), and LOD_3 was at the percentile 10% less than p_1 (p_1*0.90). For each data set, values were randomly assigned to one of the three LODs with equal probability. A value was censored if it was less than or equal to its assigned LOD. The choice of these levels of LODs was based on the rationale that differing LODs in a dataset are typically close to one another due to day-to-day instrument analytical variability, lab-to-lab variability, or duration of the samples when collected over the approximate same duration. Because LOD_2 and LOD_3 are simulated to be less than LOD_1, which is at the assigned percent censoring of the distribution, the final percent censoring in the multiple LODs simulation is always slightly less than the expected percent censoring.

In Simulation 3, a mixed lognormal distribution was created by combining two randomly drawn lognormal distributions, each of which contributed 50% to the mixed distribution. The mixed distributions were simulated to represent conditions that comprise two exposure groups with different exposure distributions that cannot be distinguished from each other because of limited descriptive information.

Two types of mixed lognormal distributions were simulated. The first comprised two lognormal distributions with a GM_1 = 1 and a GM_2 = 5 and the second with a GM_1 = 1 and a GM_2 = 10. In both simulations, the GSDs for each contributing lognormal distribution were fixed at 2, 3, 4, and 5. For example, at GSD = 2 the first distribution had a GM_1=1 and a GSD_1=2 and the second distribution had a GM_2=5 and a GSD_2=2. Three LODs were generated and values were censored as in Simulation 2. Only the AM and $X_{0.95}$ parameters were evaluated for the mixed distribution conditions.

As in all simulations, the observed censoring for a given dataset may deviate from the expected censoring. Datasets that were 100% censored (observed) were discarded because all methods used here are inappropriate for 100% censored data. 100% censored datasets occurred more frequently under high censoring and small sample size conditions. For a given expected censoring probability, p, and a given sample size, N, we expect to observe datasets with 100% censoring with a probability p^N(e.g., when N=5 and p=80%, the expected percent of 100% censored datasets = 32.76%).

Simulations were programmed in statistical computing software R (R Development Core Team, 2013).

Evaluation metrics

We compared the methods using relative bias and relative root mean squared error (rMSE) found in Hewett and Ganser (2007). Relative bias (called bias hereafter) is the difference between the average of estimated values and the true value relative to the true value.

$$Relative\ bias = 100 \times \frac{\bar{x} - \theta}{\theta} \qquad (4)$$

where θ is the true value of the parameter of interest (i.e., AM, GM, GSD, and $X_{0.95}$) and $\bar{x}$ is the mean of the 1000 estimates. Bias can be negative or positive. Negative bias means the method underestimated the true value of the parameter while positive bias denotes overestimation of the true value.

The relative root mean squared error (called rMSE hereafter) is a measure that combines the bias and the precision of the method relative to the true value. rMSE can only be positive.

$$Relative\ rMSE = 100 \times \frac{1}{\theta}\sqrt{(\bar{x} - \theta)^2 \quad + \frac{\Sigma(x_i - \bar{x})^2}{N-1}} \qquad (5)$$

The smaller the bias and rMSE, the better the performance of the method.

Results

The estimates of bias and rMSE for the AM, $X_{0.95}$, GM, and GSD in Simulations 1 and 2 (lognormal distribution with a single LOD and lognormal distribution with multiple LODs) were very similar. Only figures for the AM, $X_{0.95}$, and GSD with multiple LODs are shown in this paper. Results for the GM from the lognormal distributions and multiple LODs (Simulation 2, Figures 1-2), the lognormal distributions with a single LOD (Simulation 1, Figures 3-10), and for mixed distributions (Simulation 3, Figures 11-16) are provided in the On-line Supplemental Materials. The bias and rMSE were computed from each set of conditions for each of the four parameters. The size of the circle corresponds to the magnitude of the bias or the rMSE on a continuous scale. The legend on the right side of the figures shows circles corresponding to specific values. Thus, the size of each circle in the figures generally falls between two circles in the legend. The largest circle in the legend is interpreted as either equal to or greater than 50% (for bias) or 150% (for rMSE). The figures are intended to provide an overview of the range of conditions evaluated. The actual numbers are included in the On-line Supplemental Materials (Excel file). In this presentation, we refer to GSDs of 2-3 as low variability and GSDs of 4-5 as high variability. Sample sizes of 5, 10, and $\geq$20 are considered to be small, moderate and large sample sizes, respectively. Censoring of <50 and $\leq$80 percents are described because they appear to be the breakpoints where one or more of the methods' performances changed. Generally, the bias and rMSE increased (the estimates became less accurate and precise) as the sample size decreased, the percent censoring increased, and/or variability (GSD) was high.

Lognormal distribution for multiple LODs

Arithmetic mean

Figures 2 and 3 present the bias and rMSE results for the AM for a lognormal distribution with multiple LODs. Overall, the β-substitution method generally produced comparable or smaller bias and rMSE compared to the ML and K-M methods, even under small and moderate sample size conditions. When the variability was high, the β-substitution and the $ML_{MVUE\text{-}AM}$ methods produced similar bias and rMSE. Under small N, high variability and high percent censoring conditions, the bias and rMSE for the

$ML_{MLE\text{-}AM}$ method were higher compared to the β-substitution and the $ML_{MVUE\text{-}AM}$ methods, indicating that the $ML_{MLE\text{-}AM}$ method was generally less accurate and less precise than the β-substitution and the $ML_{MVUE\text{-}AM}$ methods. The poor performance of the $ML_{MLE\text{-}AM}$ approach was mainly due to the transformation bias. The use of the MVUE equation appeared to mitigate this problem in the ML method as evident by the small bias and rMSE of the $ML_{MVUE\text{-}AM}$ method.

The K-M method's bias, for the most part, was comparable to the β-substitution and $ML_{MVUE\text{-}AM}$ methods when censoring was <50% and sample sizes were moderate to large, regardless of the GSD evaluated. The rMSE of the K-M method generally increased under small to moderate sample conditions and/or high censoring (≥80%), indicating less precision under those conditions.

All three methods generally were observed to have similar rMSEs under the condition of large Ns and censoring level approximately ≤80%. The β-substitution and the $ML_{MVUE\text{-}AM}$ methods tended to underestimate the true AM (negative mean bias), whereas the $ML_{MLE\text{-}AM}$ and K-M methods tended to overestimate the true AM (positive mean bias).

We also found that the distributions of the estimates from the 1000 datasets for small sample sizes conditions were typically skewed. Figure 4 shows an example of the distributions of the AM estimates under the condition of N=5, GM=1, GSD=4, and p=40%. The $ML_{MLE\text{-}AM}$ equation tended to produce more extreme estimates of AM (longer tail) compared to the other methods, resulting in the average of the AM estimates being much larger than the median and the true AM (although this was less of an issue when Ns were very large). This is an example of where the use of bias could be misleading (i.e., extreme values resulted in a higher average value than the value of a large majority of the estimates). In other instances, a method may yield a low bias even though it is a result of averaging very low and very high estimates of the parameter compared to the true value. For such cases, the rMSE is a better metric of the method's performance.

95th percentile

Figure 5 and 6 show that the bias and rMSE in the estimation of the $X_{0.95}$ from the β-substitution method generally were similar or smaller than for the ML and the K-M methods, even for small and moderate samples. The bias and rMSE of the ML and K-M methods was adversely affected by both small sample size and by high level of censoring. As the variability increased, the bias and rMSE for all three methods increased, particularly for small to moderately sample sizes. The β-substitution was least affected of the three methods. The β-substitution method tended to underestimate the true $X_{0.95}$ whereas the ML and the K-M methods generally overestimated the true $X_{0.95}$.

GSD

As shown in Figures 7 and 8, the bias and rMSE found for the estimation of the GSD from the β-substitution and the ML methodss were comparable in most simulated conditions; however, at high variability under the conditions of small to moderate sample sizes, the ML method's bias and rMSE were greater than those of the β-substitution method. Both the ML and the β-substitution methods tended to overestimate the GSD and produce extreme values as indicated by the high rMSE. The non-parametric K-M method does not compute the GM and GSD.

Mixed lognormal distribution and multiple LODs

Arithmetic mean

The bias and rMSE of the mixed distributions where the true GMs were 1 and 5 were similar to the results from those of a lognormal distribution (Figures 11-12 in the On-line Supplemental Materials). This is probably due to the modes (GMs) of the two distributions being relatively close to each other so that the resultant distributions more resembled lognormal distributions than the intended bimodal distributions.

Figures 9 and 10 show the bias and rMSE from the mixed distribution where the true GMs were 1 and 10. The β-substitution method generally produced comparable or smaller bias and rMSE than the ML and the K-M methods. The β-substitution method, although being a parametric method, appeared to be less sensitive to the biomodal distribution than the ML methods. The K-M method had low bias for censoring up to 50% and high rMSE under small sample sizes conditions.

Discussion

Our simulations covered a wider range of N, GSD, and percent censored than previous studies. First, we compared four important parameters: the AM, the $X_{0.95}$, the GM and the GSD. We covered sample sizes as low as 5 and GSDs that represented very routine, well controlled situations (GSD=2) and unusual situations where non-routine work is being carried out (GSD=5). We evaluated high levels of censoring (up to 90%) that may be applicable to many of today's workplaces that are very well controlled. We simulated the condition of multiple LODs, which can occur due to varying durations of the measurements and laboratory conditions. Finally, we evaluated mixed distributions, which can occur when grouping disparate measurements that have limited sampling documentation. Thus, most of the conditions likely to be encountered by the practitioner have been investigated.

In comparing the ML and the K-M methods, we found inconsistent recommendations in the literature. Hewett and Ganser (2007) recommended the ML method over the K-M method at percent censoring ≤50% even for small N and mixed distributions. Helsel (2005 and 2010), on the other hand, preferred the K-M method over the ML method for N≤ 50 and censoring ≤50%, although the ML method was recommended for N≥50 and censoring 50-80%. His suggestion was based on reviews of published studies (e.g., Shumway, 2002; Antweiler and Taylor, 2008) that mostly compared the population mean (and other non-lognormal summary statistics, such as the median and standard deviation). These inconsistent recommendations for the mean could be due to their using different equations for calculating the AM in the ML method. Our simulation results showed that the K-M method generally had a lower bias and rMSE than the $ML_{MLE\text{-}AM}$ method for censoring ≤50% in the estimation of the AM, particularly for small and moderate sample sizes. However, at censoring >50% and large sample size conditions, the $ML_{MLE\text{-}AM}$ was generally comparable or better than the K-M method. Our results comparing the $ML_{MLE\text{-}AM}$ with the K-M methods were generally in line with Helsel's recommendation of the K-M method and other studies (Cohen, 1988; Shumway et al., 2002). A comparison of the bias and rMSE of the K-M method with the ML method using the MVUE equation, however led us to similar conclusions as Hewett and Ganser's.

We therefore suggest that practitioners apply these recommendations with care, taking into considerations their data, their needs and the uses of the statistical analysis. For example, Minitab and the NADA package that were referenced in Helsel's book (Helsel, 2005) used the $ML_{MLE\text{-}AM}$ equation. If a practitioner follows Hewett and Ganser's recommendation to use the ML method even for small sample sizes, without knowing which formula the package uses to calculate the mean, a misinterpretation of the results could occur.

Those practitioners who are assessing compliance with occupational exposure limits may be interested in the $X_{0.95}$ of the measurements and therefore may not be concerned with the different formulas used to calculate AMs. Occupational exposure assessment strategies often rely on the lognormality assumption to obtain recommended statistics (Ignacio and Bullock, 2006; Ramachandran, 2005), which typically do not include the AM, but include the $X_{0.95}$ and the upper confidence limit of the $X_{0.95}$ (that are dependent on the GM and the GSD). These statistics are used for evaluation of compliance. Compliance assessments also typically involve prioritization of exposure groups with only the highest exposure groups being monitored. This prioritization process, therefore, likely results in exposure groups with a lower degree of censoring.

Our purpose of assessing occupational exposures for the GuLF STUDY is different from the above. The AM (dependent on the GM and the GSD) can be used to compare and contrast exposure groups in the epidemiologic study looking at chronic effects while the $X_{0.95}$ could possibly be used for studying effects of peak exposures. Since all exposure groups must be assessed in an epidemiologic study, it is more likely that there will be exposure groups with highly censored data (> 50%). Thus, we needed to identify a method that developed estimates with acceptable bias and error in the presence of high levels of censoring.

As with any computer simulation study, it is worthy to note that for a given dataset (especially a small dataset), the true underlying distribution is often unknown, and the percent censoring from the data does not necessarily correspond to the actual percentile in the true distribution used in our simulations. Hence, the true bias will probably differ from the composite bias obtained from these simulations and thus, the bias reported here cannot be assumed to be the bias of any particular dataset even though it meets the conditions we evaluated. However, these simulations serve as good evaluation tools to compare the methods when subjected to the same conditions. The mixed distribution simulation resembled more a slightly contaminated data rather than extreme cases of mixed distributions. If the modes of the data were clearly not lognormal (GM_1=1 and GM_2=500), then any of the parametric method might not be appropriate. Another limitation of our study is that we did not evaluate the uncertainty of the estimates. Estimation of uncertainty was, however, beyond the scope of this work. Also, we simulated conditions that are generally found in typical, routine workplace operations; the operations in the GuLF STUDY were often non-routine, and therefore may not have been covered by our simulation conditions.

Conclusions

This simulation study was conducted to identify a methodology to handle heavily censored data in the GuLF STUDY. The β-substitution method performed better than the ML and the K-M methods under most conditions of our study (including low N, high censoring, high variability, multiple LODs, and mixed distributions) using relative bias and relative rMSE as the evaluation metrics. The β-substitution method's accuracy and precision decreased at small and moderate sample sizes ($N \leq 10$), but was still the best of the three methods. Estimates for sample size <5 are likely to be unreliable. The ML generally did well with large samples sizes and lognormal distributions. The use of the minimum variance unbiased estimator equation in the estimation for the AM using the ML estimates of the GM and the GSD reduced the ML's transformation bias to the AM for small to moderate sample sizes. The K-M method was generally less biased at censoring levels <50%. Though very robust, a major limitation of the β-substitution method is the lack of a confidence interval around the mean, whereas confidence intervals can be computed for the ML and the K-M methods. This study suggests that none of the statistical methods evaluated in this paper are recommended for datasets that have a combination of small to moderate sample sizes, high level of censoring, or high variability. There is a need for the development of other methods that could improve the accuracy under those conditions and could also provide the uncertainty estimates. Bayesian approaches may offer useful insights in this regard.

Acknowledgement

This research was supported by the Intramural Research Program of the NIH, National Institute of Environmental Health Sciences.

References

Aitchison, J and Brown, J.A.C (1969). The Lognormal Distribution, Cambridge: Cambridge University Press.

Antweiler RC, Taylor HC (2008) Evaluation of statistical treatments of left-censored environmental data using coincident uncensored data sets: I. Summary statistics. Env Sci Technol; 42: 3732–8

Beal, DJ (2010) A macro for calculating summary statistics on left-censored environmental data using the Kaplan-Meier method. Proceedings of the 18th Annual Conference of the Southeast SAS Users Group.

Cohen, A.C. (1950). Estimating the Mean and Variance of Normal Populations from Singly Truncated and Double Truncated Samples. Ann. Math. Statist., Vol. 21, pp. 557-569.

Cohen, AC. (1959) Simplified estimators for the normal distribution when samples are singly censored or truncated. Technometrics 1, 217-37.

Cohen, AC. (1961) Tables for maximum likelihood estimates: Singly truncated and singly censored samples. Technometrics 3, 535-541.

Cohn, TA (1988) Adjusted maximum likelihood estimation of the moments of log-normal populations from type I censored samples: U.S. Geological Survey Open-File Report 88-350, 34

Dinse GE, Jusko TA , Ho LA, Annam K, Graubard BI, Hertz-Picciotto I, Miller WF, Gillespie BW, and Weinberg CR* American Journal of Epidemiology Vol. 179, No. 8 DOI: 10.1093/aje/kwu017

European Food Safety Authority. (2010) Management of left-censored data in dietary exposure assessment of chemical substances. EFSA Journal; 8(3). Available at: www.efsa.europa.eu.
Finkelstein MM, and Verma DK. (2001) Exposure estimation in the presence of nondetectable values: another look. Am Ind Hyg Assoc J; 62:195-8.
Fisher, RA. (1925) Theory of statistical estimation. Proceedings of the Cambridge Philosophical Society 22, 700-725.
Ganser GH, and Hewett P. (2010) An accurate substitution method for analyzing censored data, J of Occup and Environ Hyg, 7: 4, 233-44.
Gillespie BW, Chen Q, Reichert H et al. (2010) Estimating population distributions when some data are below a limit of detection by using a reverse Kaplan-Meier estimator. Epidemiology; 21: S64–70.
Helsel DR. (2005) Nondetects and data analysis. New York: John Wiley & Sons, Inc.
Helsel DR. (2010) Much ado about next to nothing: Incorporating nondetects in science. Ann Occup. Hyg ;54:257-62.
Hewett, P., & Ganser, G. H. (1997). Simple Procedures for Calculating Confidence Intervals around the Sample Mean and Exceedance Fraction Derived from Lognormally Distributed Data. *Applied Occupational and Environmental Hygiene*, *12*(2), 132–142.
Hewett P, and Ganser GH. (2007) A comparison of several methods for analyzing censored data. Ann Occup Hyg;51:611-32.
Hill AB (1965) The evaluation of disease: association or causation? Proc R Soc Med 58:295-300.
Hornung RW, and Reed LD. (1990) Estimation of average concentration in the presence of non-detectable values. Appl. Occup. Envir. Hyg. 5(1):46–51.
Ignacio JS, Bullock WH. editors. (2006) A strategy for assessing and managing occupational exposures. 3rd ed. Fairfax, VA: AIHA Press.
Kaplan EL, and Meier P. (1958) Nonparametric estimation from incomplete observations. J Am Stat Assoc **1958**; 53:457-81

Krishnamoorthy K, Mallick A, Mathew T. (2009) Model based imputation approach for data analysis in the presence of non-detectable values: Normal and Related Distributions. Ann Occup Hyg; 59: 249–68.

Kroll CN, Stedinger JR. (1996) Estimation of moments and quantiles using censored data. Water Resour Res; 32: 1005–12

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

National Institute for Occupational Safety and Health. (2011) NIOSH Deepwater Horizon Roster Summary Report. Available at: http://www.cdc.gov/niosh/docs/2011-175/pdfs/2011-175.pdf

Occupational Safety & Health Administration. (2011) Deepwater Horizon oil spill: OSHA's role in the response. Available at: http://www.osha.gov/oilspills/dwh_osha_response_0511a.pdf.

R Development Core Team. (2013) R: a language and environment for statistical computing. R Foundation for Statistical Computing: Vienna. http://www.R-project.org.

Ramachandran G. (2005) Occupational exposure assessment for air contaminants. Boca Raton, FL: CRC Press. ISBN: 1-56670-609-2.

Rappaport, S.M. (1991). Assessment of long-term exposures to toxic substances in air – Review. *Ann of Occup Hyg*, 35:61-121.

Seixas, N., Robins, T., Moulton, L. The use of geometric mean and arithmetic mean exposures in occupational epidemiology. *Am J Ind Med* 1988: 14: 465–77.

Shumway RH, Azari RS, and Kayhanian M. (2002). Statistical approaches to estimating mean water quality concentrations with detection limits: Environmental Science and Technology 36, 3345-3353

Succop PA, Clark S, Chen M et al. (2004) Imputation of data values that are less than a detection limit. J Occup Environ Health; 1: 436–41.

U.S. Environmental Protection Agency. 2007. *ProUCL Version 4.0 Technical Guide*. EPA/600/R-07/041

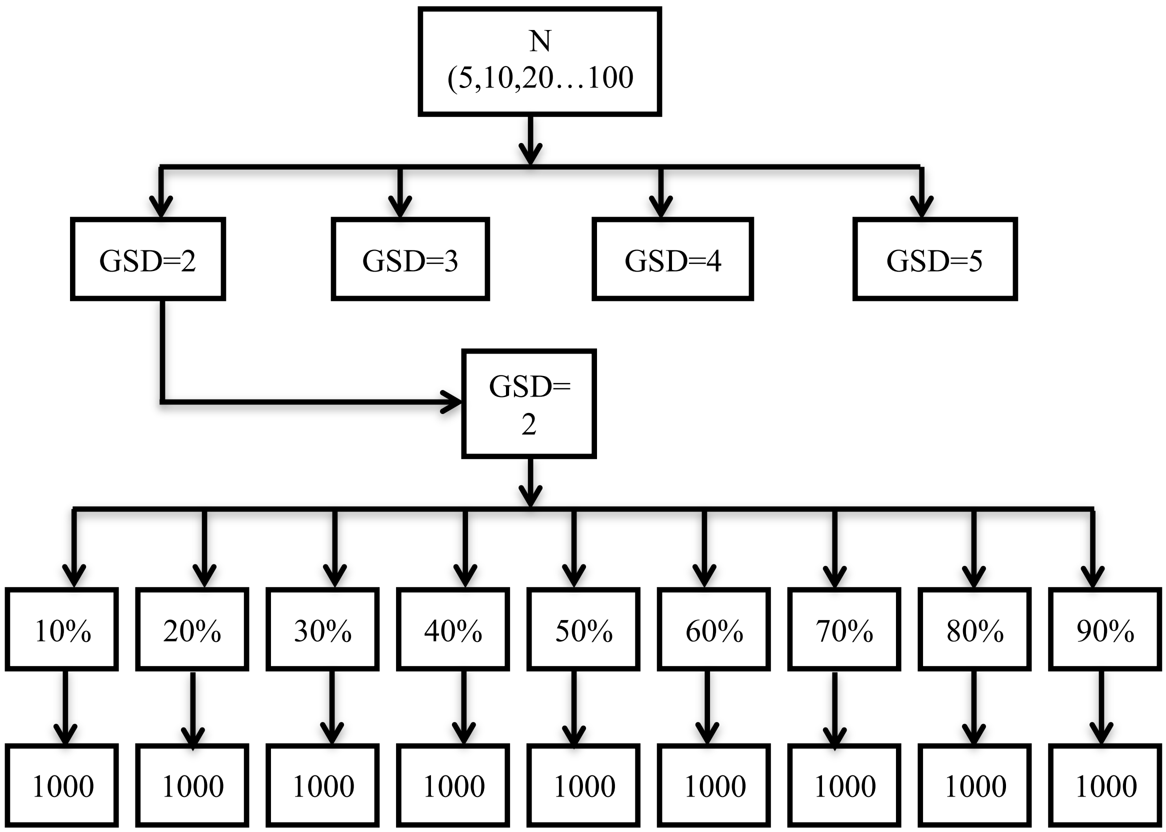

Figure 1: A graphical depiction of the simulation design. Sample sizes (N) were fixed at 5, 10, 20, 30, 40, 50, 60, 70, 90, and 100. For each sample size, data were drawn from a lognormal distribution with a true GM=1 and true GSDs of 2, 3, 4, and 5, respectively. Datasets were censored in increments of 10% with either a single LOD value or multiple LODs. For each combination of N, GM, GSD, and percent censored, 1000 datasets were generated and analyzed using the β-substitution, the ML, and the K-M methods. A mixed lognormal distribution is created by combining two lognormal distributions with GM_1=1 and GM_2=5 or 10.

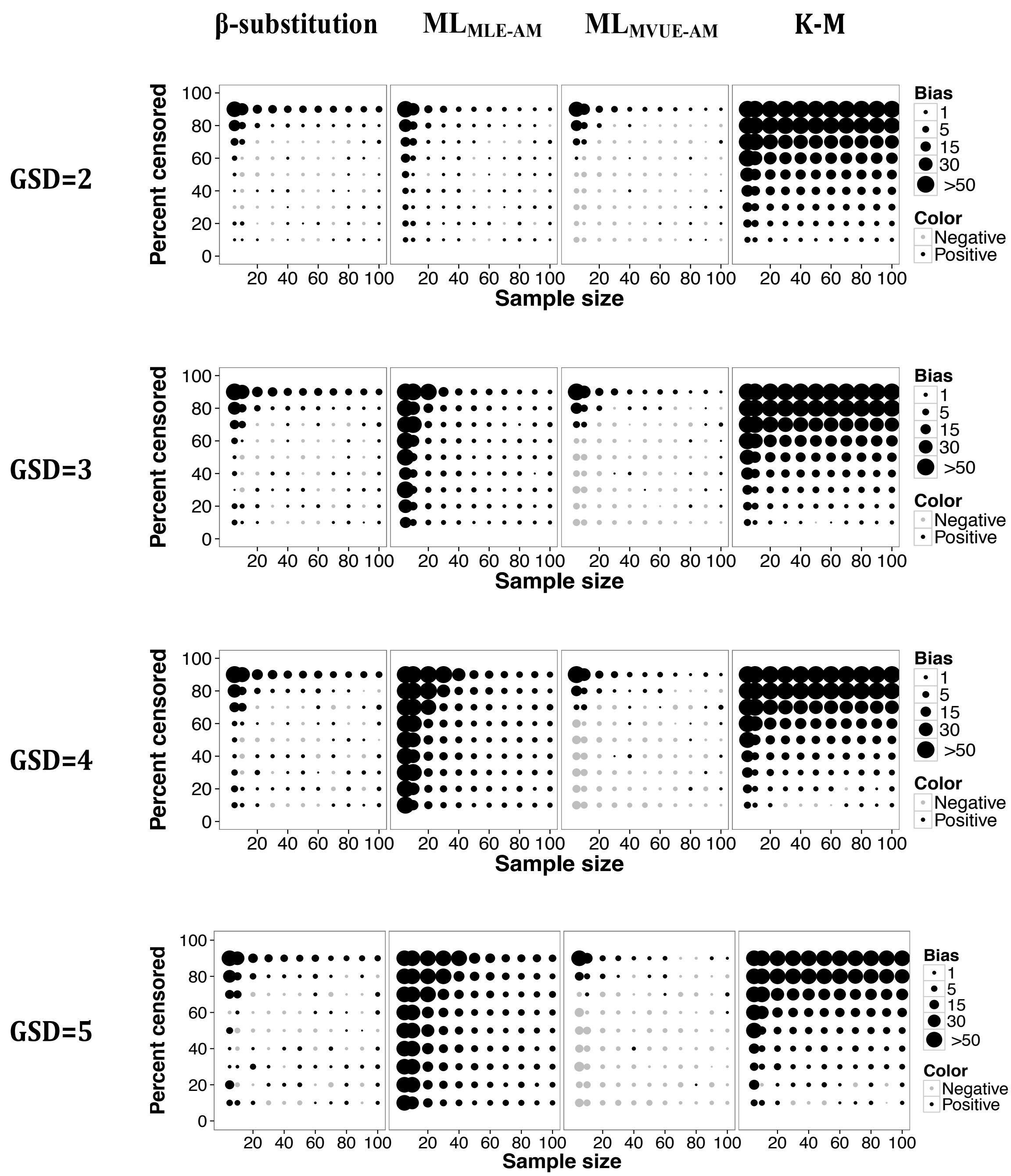

Figure 2: Relative bias in the estimate of the AM of a lognormal distribution and multiple LODs for different sample sizes, percent censoring, and GSDs for the β-substitution, the ML, and the K-M methods. ML$_{MLE-AM}$ and ML$_{MVUE-AM}$ denote two different ways of estimating the AM from the ML estimates of μ and σ from the log-transformed data.

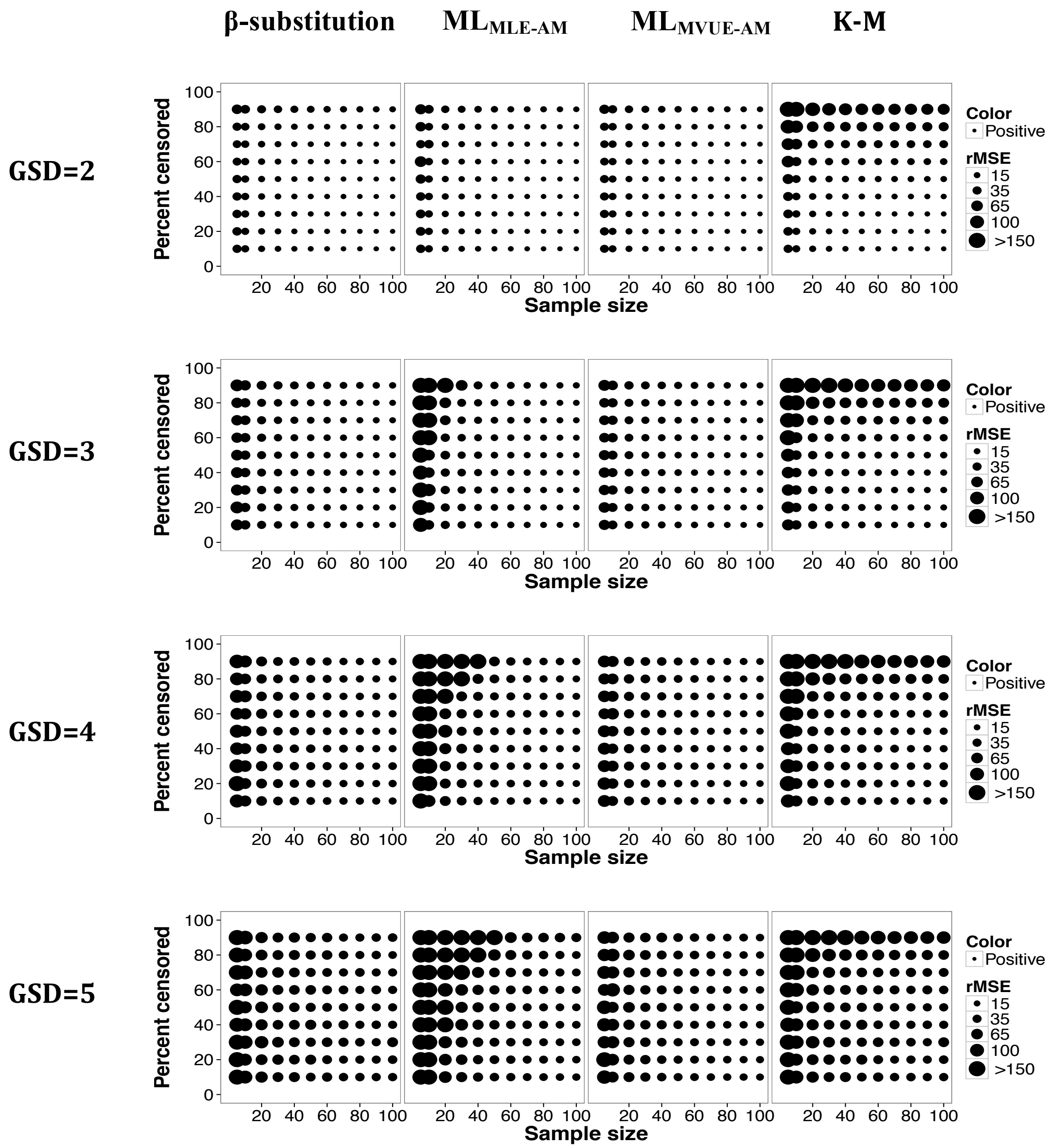

Figure 3: Relative rMSE in the estimate of the AM of a lognormal distribution and multiple LODs for different sample sizes, percent censoring, and GSDs for the β-substitution, the ML, and the K-M methods. ML_{MLE-AM} and $ML_{MVUE-AM}$ denote two different ways of estimating the AM from the ML estimates of μ and σ from the log-transformed data.

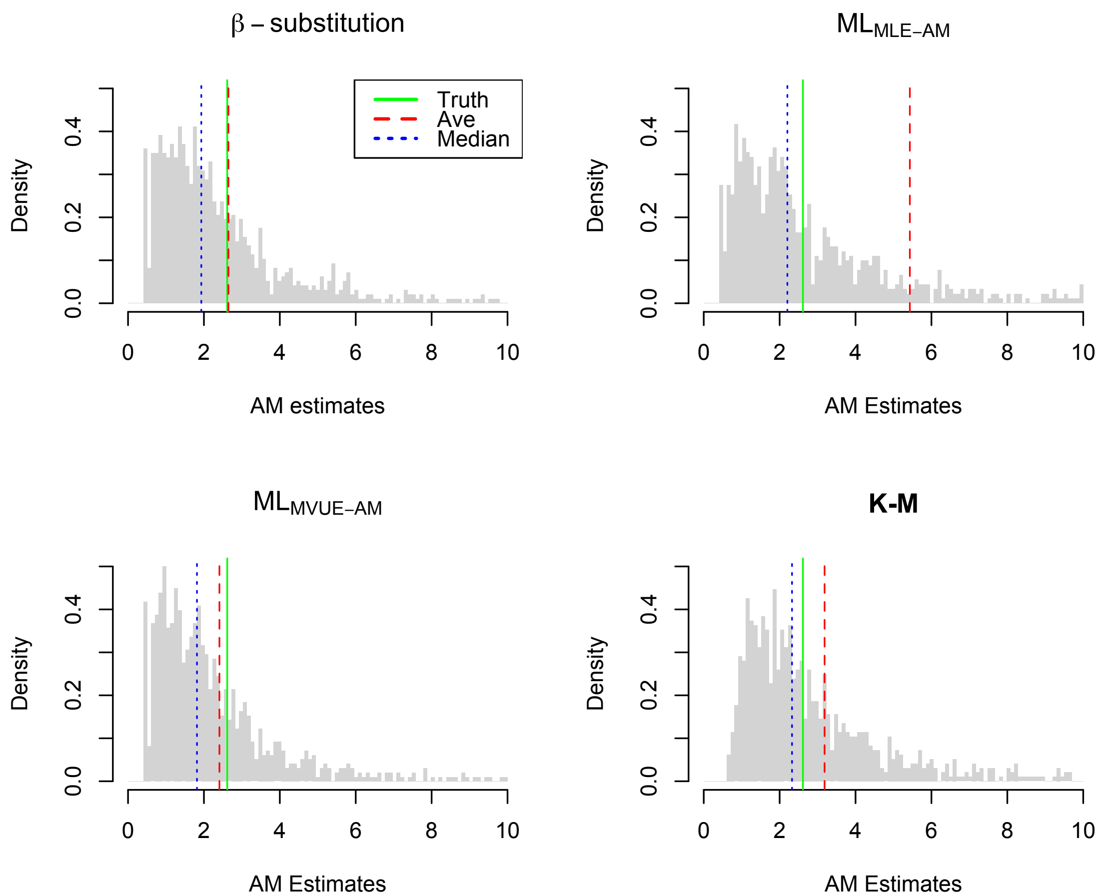

Figure 4: Histograms of the AM estimates from 1000 simulated datasets under the condition of N=5, GM=1, GSD=4, and percent censoring = 40 for three estimation methods. $ML_{MLE\text{-}AM}$ and $ML_{MVUE\text{-}AM}$ denote two different ways of estimating the AM from the ML estimates of μ and σ from the log-transformed data. The average, the median, and the true AM vertical lines showed the sensitivity of the average in a skewed distribution. The $ML_{MLE\text{-}AM}$ had a large variability, resulting in a higher average AM value compared the other methods.

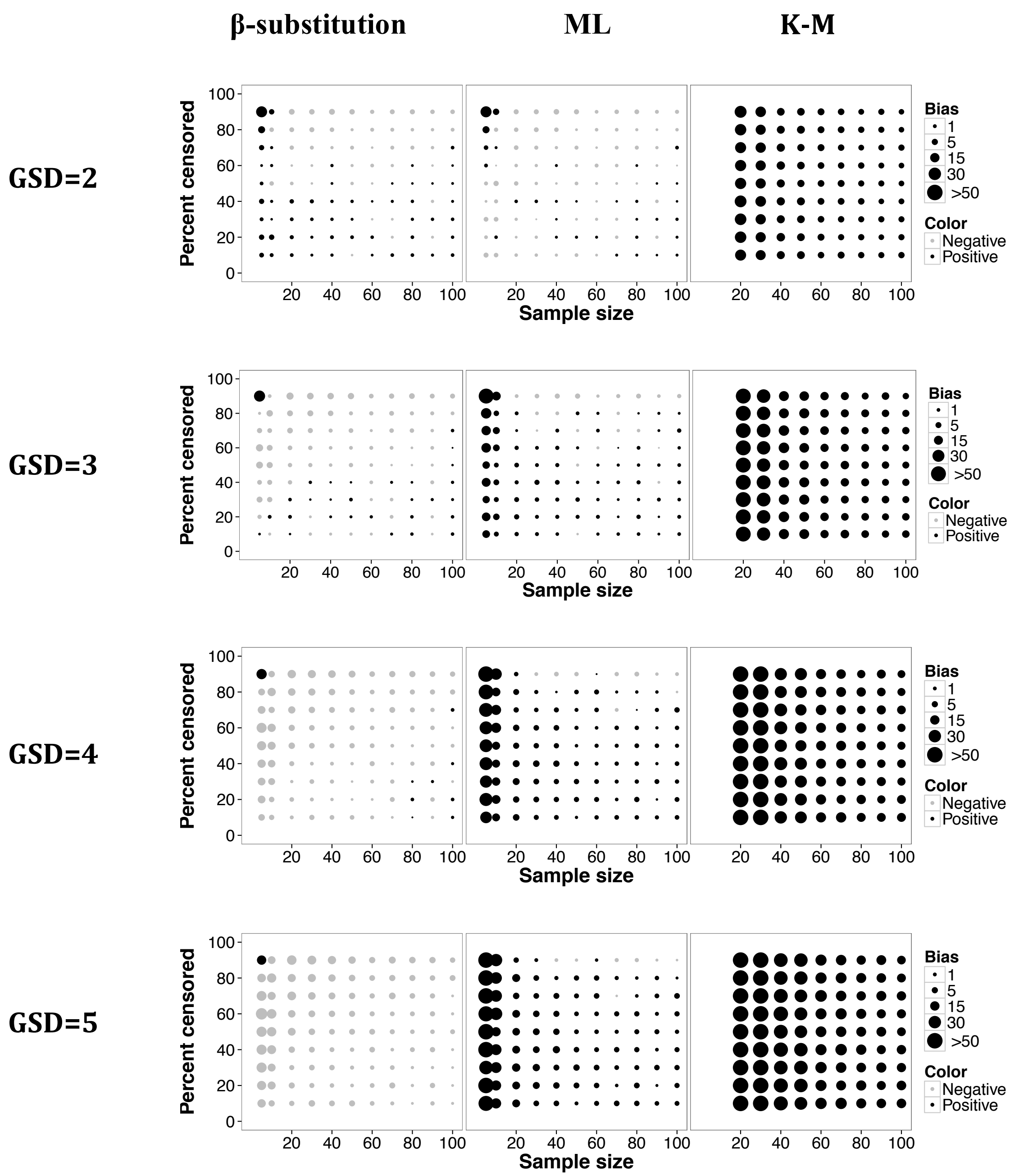

Figure 5: Relative bias in the estimate of the 95th percentile of a lognormal distribution and multiple LODs for different sample sizes, percent censoring, and GSDs for the β-substitution, the ML, and the K-M methods. The K-M method required a minimum sample size of 20 to estimate the 95th percentile.

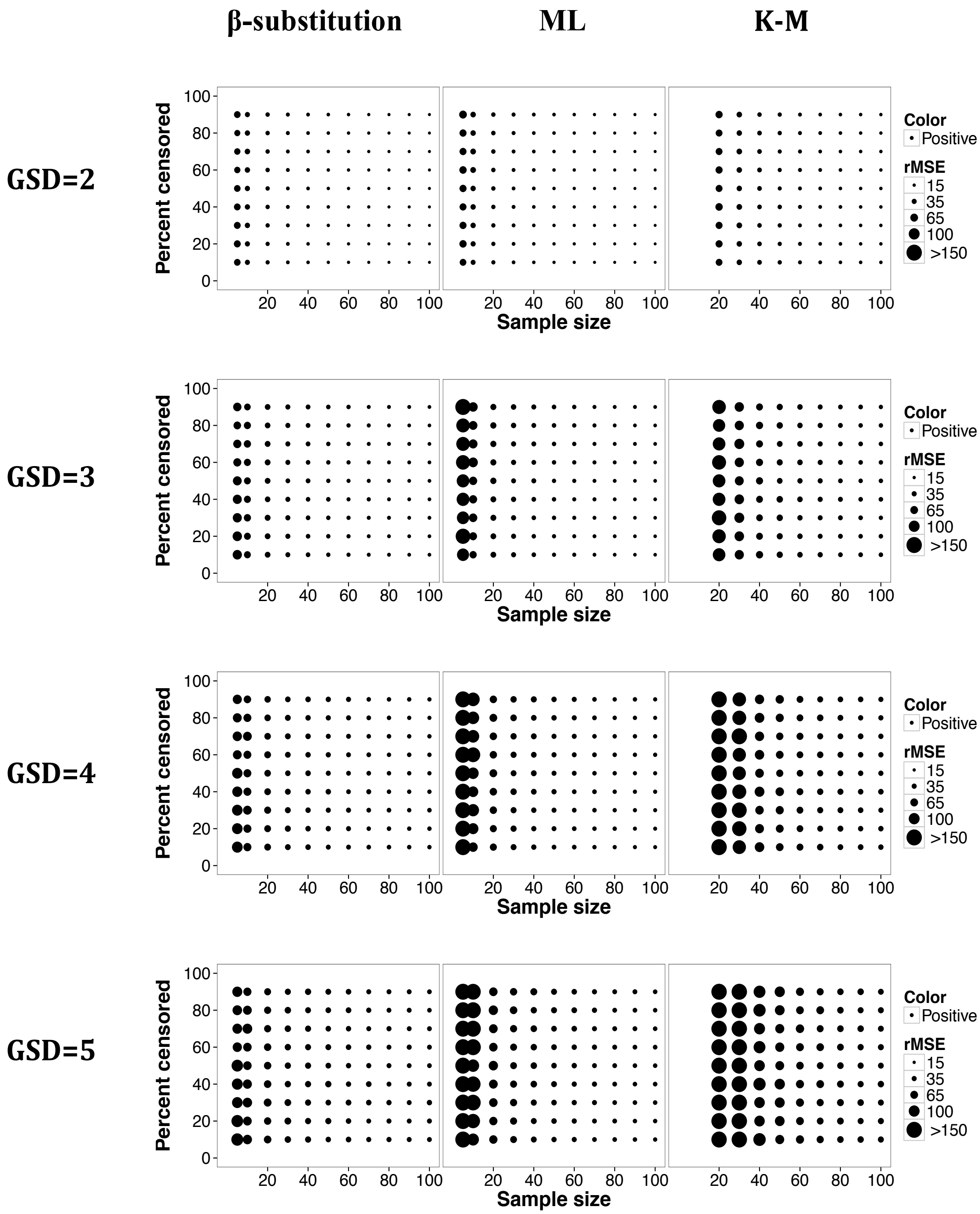

Figure 6: Relative rMSE in the estimate of the 95th percentile of a lognormal distribution and multiple LODs for different sample sizes, percent censoring, and GSDs for the β-substitution, ML, and K-M methods. The K-M method required a minimum sample size of 20 to estimate the 95th percentile.

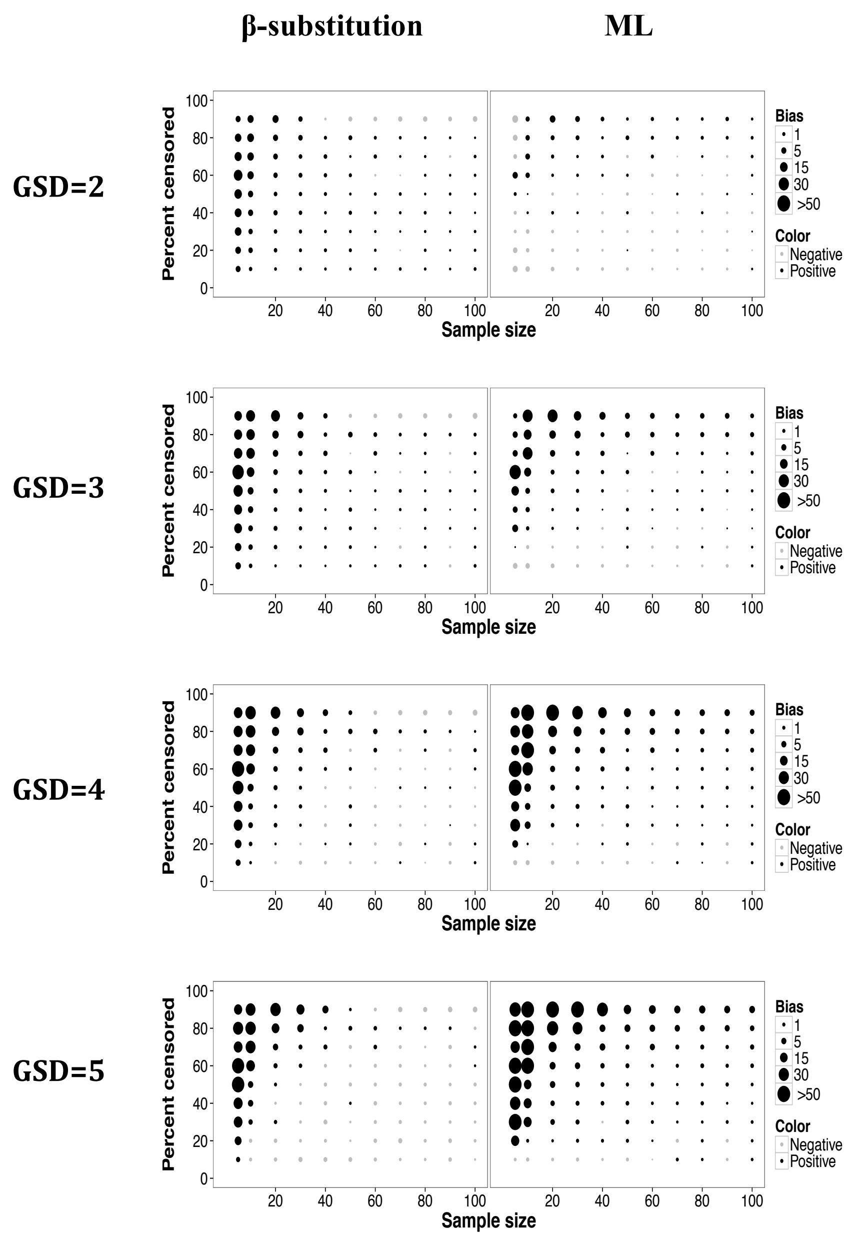

Figure 7: Relative bias in the estimate of the GSD of a lognormal distribution and multiple LODs for different sample sizes, percent censoring, and GSDs for the β-substitution and the ML methods. The K-M method does not compute GSD.

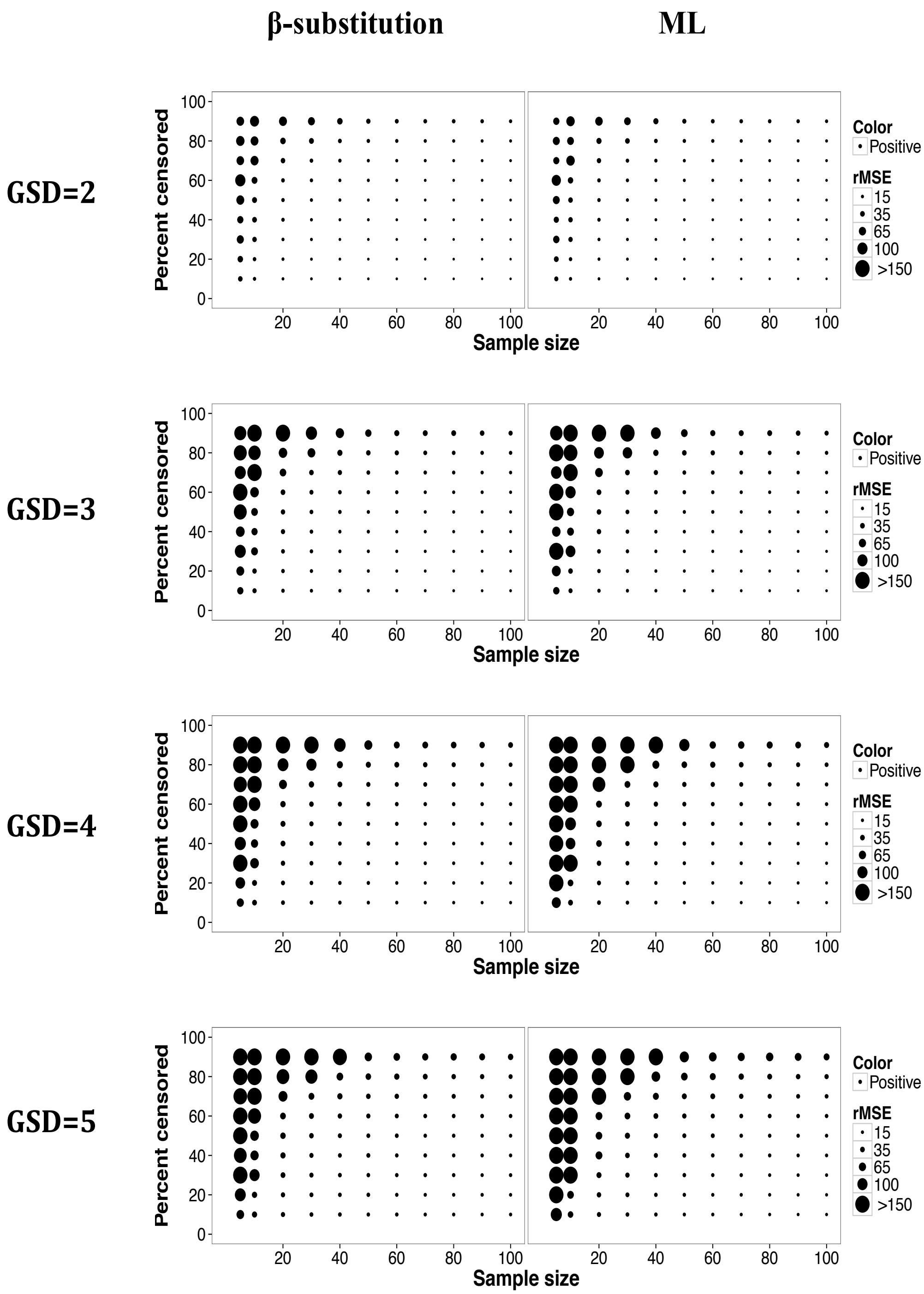

Figure 8: Relative rMSE in the estimate of the GSD of a lognormal distribution and multiple LODs for different sample sizes, percent censoring, and GSDs for the β-substitution and the ML methods. The K-M method does not compute GSD.

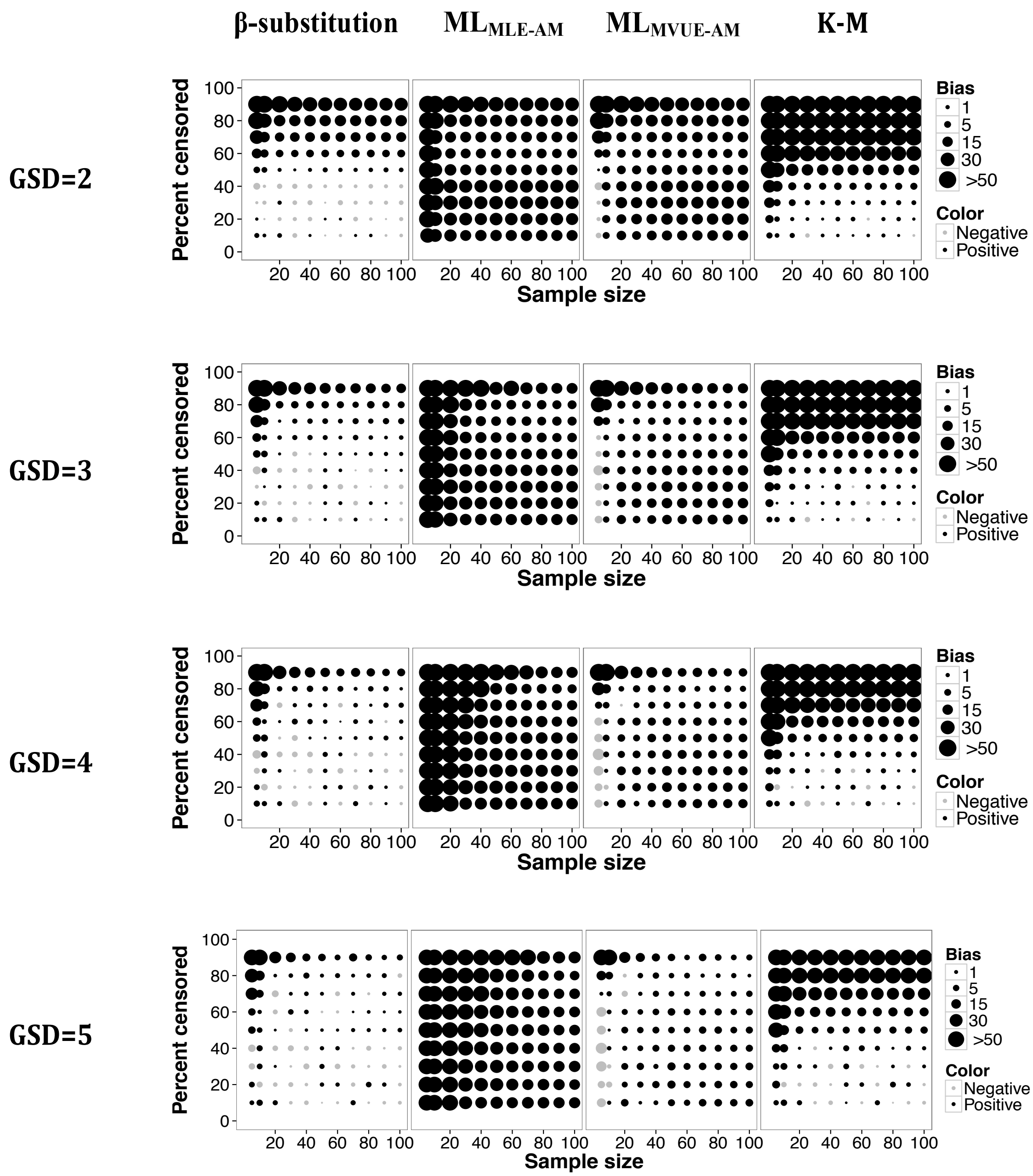

Figure 9: Relative bias in the estimate of the AM of a mixed distribution (GM_1=1 and GM_2 =10) and multiple LODs for different sample sizes, percent censoring, and GSDs for the β-substitution, the ML, and the K-M methods. $ML_{MLE\text{-}AM}$ and $ML_{MVUE\text{-}AM}$ denote two different ways of estimating the AM from the ML estimates of μ and σ from the log-transformed data.

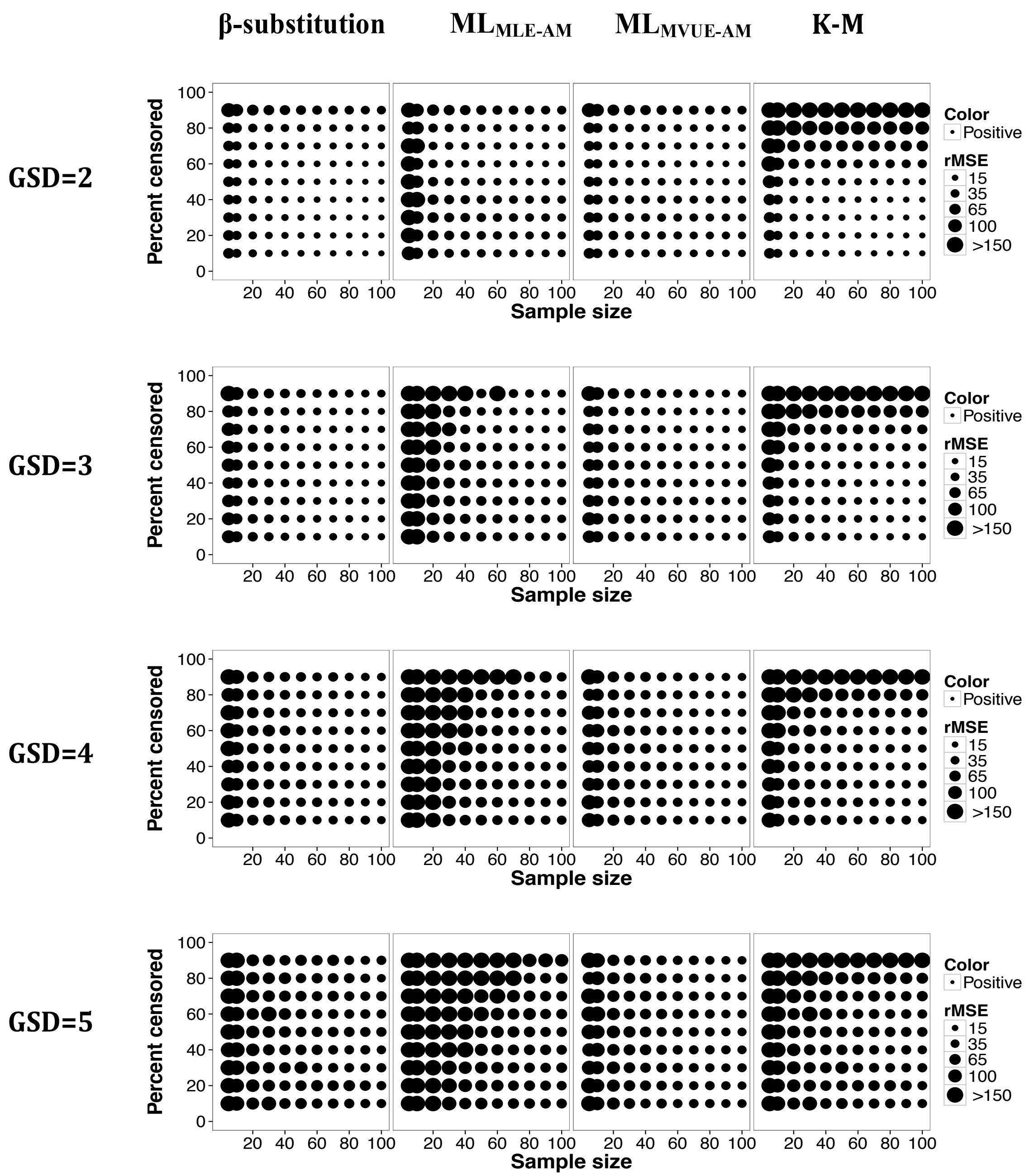

Figure 10: Relative rMSE in the estimate of the AM of a mixed distribution (GM_1=1 and GM_2 =10) and multiple LODs for different sample sizes, percent censoring, and GSDs for the β-substitution, the ML, and the K-M methods. $ML_{MLE\text{-}AM}$ and $ML_{MVUE\text{-}AM}$ denote two different ways of estimating the AM from the ML estimates of μ and σ from the log-transformed data.

Chapter III

Comparison of the β-substitution Method and a Bayesian Method for Analyzing Left-Censored Data

Tran Huynh, Harrison Quick, Gurumurthy Ramachandran, Sudipto Banerjee, Joao Monteiro, Caroline Groth, Mark Stenzel, Aaron Blair, Dale P. Sandler, Lawrence S. Engel, Richard K. Kwok, and Patricia A. Stewart

To be submitted to the Annals of Occupational Hygiene

Summary

Classical statistical methods for analyzing exposure data with values below the detection limits are well-described in the occupational hygiene literature, but an evaluation of a Bayesian approach for handling such data is currently lacking. Here, we first describe a Bayesian framework for analyzing censored data. We then present the results of a simulation study conducted to compare the β-substitution method with a Bayesian method for exposure datasets drawn from lognormal distributions and mixed lognormal distributions with varying sample sizes, geometric standard deviations (GSDs), and censoring for single and multiple limits of detection. For each set of factors, estimates for the arithmetic mean (AM), geometric mean (GM), GSD, and the 95thpercentile ($X_{0.95}$) of the exposure distribution were obtained. We evaluated the performance of each method using relative bias, the root mean squared error (rMSE), and coverage (the percentage of the computed 95% uncertainty intervals containing the true value). The Bayesian method using non-informative priors and the β-substitution method were generally comparable in bias and rMSE when estimating the AM and GM. For the GSD and the 95th percentile, the Bayesian method with non-informative priors was more biased, and had a higher rMSE, than the β-substitution method but the use of more informative priors generally improved the Bayesian method's performance, making both the bias and the rMSE more comparable to the β-substitution method. The advantage of the Bayesian method is that it allowed the use of prior information and also provided estimates of uncertainty for these parameters of interest, whereas the β-substitution method only provided estimates of uncertainty for the AM, and the coverage was not as consistent. Selection of one or the other method depends on the needs of the practitioner, the availability of prior information and the distribution characteristics of the exposure data. We suggest the use of the Bayesian method if the practitioner has the computational resources and prior information, as the method generally provides accurate estimates and also provides the distributions of all of the parameters, which could be useful for making decisions in exposure management applications.

Introduction

As exposure limits continue to decrease due to growing evidence of detrimental health effects at air concentrations lower than previously identified, the need to incorporate data below the limits of detection (LOD) of the reporting analytic laboratory in the data analysis is an important part of the exposure assessment strategy. Analysis of this type of censored data (Type I left-censoring) requires a statistical approach that not only accounts for the detection limits but also accurately estimates exposure distributional parameters such as the arithmetic mean (AM), which is often used for occupational epidemiological studies (Seixas et al, 1988; Rappaport, 1991), and the geometric mean (GM), the geometric standard deviation (GSD), and the 95th percentile ($X_{0.95}$), parameters often used for exposure management purposes.

Many statistical methods for dealing with censored data have been discussed in the occupational and environmental exposure assessment literature including the standard substitution method, variations of the maximum likelihood (ML) method, the probability plot based method, the Shapiro–Wilk W-statistic-based approach, the non-parametric Kaplan-Meier method, and the β-substitution method. The limitations of these methods motivated the implementation and evaluation of a Bayesian approach for handling censored occupational monitoring data. Bayesian models for left-censored data have been used in other fields and have been shown to perform as well as the ML method (e.g., Busschaert et al., 2011; Paulo et al, 2005). The goal of this paper is to compare the β-substitution method to a Bayesian method because the β-substitution method was found to perform as well as or better than the ML and K-M methods (Huynh et al, 2014).

Background

The standard substitution method that substitutes the censored data (i.e., those < LOD) with LOD/2 (or LOD/√2) (Hornung and Reed, 1990) is the easiest to use but has been shown to perform poorly in several comparison studies (Helsel, 2005 and 2010; Singh et al,. 2006, Hewett and Ganser, 2007). In the β-substitution method, the LOD is substituted with a data-dependent β-factor multiplied by the LOD. It was found to exhibit less bias than the standard substitution method and was comparable to the ML method (Ganser and Hewett, 2010). The ML is a parametric method (i.e. requires the assumption of an underlying distribution) in which parameters are estimated by maximizing the likelihood function, a product of the probability density function (PDF) for the measurements greater than the LOD and the cumulative distribution function (CDF) for the measurements less than the LOD (Fisher,1925; Cohen, 1959;1961; Finkelstein and Verma, 2001; Krishnamoothy et al., 2011). The probability plot based methods (also known as Regression on Order Statistics (ROS) or log-probit regression (LPR)) assumes a lognormal distribution of the data and computes the mean and standard deviation by fitting a linear regression of the log-transformed data versus their normal scores on a normal or lognormal probability plot (Kroll and Stedinger, 1996; Helsel and Cohen, 1988; Gilliom and Helsel,1986; Helsel, 2005). The LPR method was generally comparable to the ML method in most simulation studies, although its variant (the robust LPR) might show slight improvement over the ML under a simulated mixed distribution (Gilliom and Helse, 1986). Hewett and Ganser (2007) showed that the ML method generally did better than the LPR in estimating the mean while the LPR was better at estimating the $X_{0.95}$. They also concluded that little is gained from variations of LPR or the MLE methods. In the approach based on the Shapiro–Wilk W-test statistic, the appropriate underlying distribution (e.g., normal or lognormal) is selected and nondetected values are calculated by maximizing the W-statistic with a constrained optimization algorithm (Flynn, 2010). The estimates provided by this method were comparable to the restricted MLE method and its main advantage is its ease of implementation using Microsoft Excel Solver tool. The Kaplan-Meier (K-M) method, which does not assume any distributional shape, estimates summary statistics by constructing a curve akin to an empirical CDF while adjusting for censoring (Kaplan and Meier, 1958; Gillespie et al, 2010).

Several evaluation studies have been published recommending different methods for different measurement conditions. While these recommendations have varied somewhat, the MLE method has generally been recommended for large datasets that meet the distributional shape assumption (Helsel, 2005 and 2010, Hewett and Ganser 2007). The K-M method may be preferred for moderately censored datasets with smaller sample sizes and under conditions where the distributional shape assumption is less likely to be met (Antweiler and Taylor, 2008; Helsel, 2005 and 2010). The β-substitution method has been shown to perform as well as or better than the ML method and the K-M method, particularly for small sample sizes (Ganser and Hewett, 2010; Huynh et al., 2014). Despite the accuracy of the β-substitution method, it is limited by its inability to calculate uncertainty intervals (Huynh et al., 2014). Huynh et al (2014) also found that the K-M method generally resulted in a reasonably small bias when estimating the AM under the conditions where the degree of censoring was less than 50% regardless of the sample size, and of the three methods, it was least affected by the variability in the data and the distributional shape. When comparing AMs from the parametric ML method with the K-M and the β-substitution methods, Huynh et al (2014) also found that conclusions drawn from the comparison depended on the equation used to estimate the AM. Using the uniform minimum variance unbiased estimator (MVUE) equation (Finney, 1941; Aitchison and Brown, 1957) to estimate the AM, the ML method was less biased than the K-M method for small sample sizes and was comparable to the β-substitution method. If the standard maximum likelihood (ML) equation (Cohen, 1961; Leidel et al, 1977; Selvin and Rappaport, 1989) was used, the ML method generally performed worse than the K-M and the β-substitution methods under small sample sizes and moderately censored conditions. The different equations only applied to the AM calculation, and not the GM, GSD, and $X_{0.95}$. None of the methods was found to perform satisfactorily under very small sample size (e.g., $\leq$10) or under high censoring (>80%) conditions or a combination (Huynh et al, 2014).

Methods

The Bayesian approach

Bayesian inference is based on conditional probabilities through the use of Bayes' Theorem. A likelihood for the data vector, $\boldsymbol{Y}$, given a vector of model parameters, $\boldsymbol{\theta}$, is denoted by $p(\boldsymbol{Y}|\boldsymbol{\theta})$. Bayesian inference combines $p(\boldsymbol{Y}|\boldsymbol{\theta})$ with prior information in the form of the prior distribution for $\boldsymbol{\theta}$, denoted by $p(\boldsymbol{\theta})$. Inference is then made based on the posterior distribution, $p(\boldsymbol{\theta}|\boldsymbol{Y})$, obtained via Bayes' Theorem:

$$p(\boldsymbol{\theta}|\boldsymbol{Y}) = \frac{p(\boldsymbol{Y}|\boldsymbol{\theta})p(\boldsymbol{\theta})}{p(\boldsymbol{Y})} = \frac{p(\boldsymbol{Y}|\boldsymbol{\theta})p(\boldsymbol{\theta})}{\int p(\boldsymbol{Y}|\boldsymbol{\theta})p(\boldsymbol{\theta})d\boldsymbol{\theta}} = C\ p(\boldsymbol{Y}|\boldsymbol{\theta})p(\boldsymbol{\theta})$$

$$\alpha\ \ p(\boldsymbol{Y}|\boldsymbol{\theta})p(\boldsymbol{\theta}) \qquad (1)$$

where $p(\boldsymbol{Y})$ is the marginal, or unconditional, distribution of $\boldsymbol{Y}$ and "α" denotes "proportional to". In practice, computing $p(\boldsymbol{Y})$ is computationally expensive, but since it is not a function of our model parameters $\boldsymbol{\theta}$, it is simply a constant, denoted here by *C*. Since $p(\boldsymbol{\theta}|\boldsymbol{Y})$, is a probability distribution, it must integrate to 1, thus the unknown value *C* is simply the constant that makes $p(\boldsymbol{\theta}|\boldsymbol{Y})$, a valid distribution. As a result, it suffices to compute the posterior distribution as being proportional to the likelihood times the prior. More details about Bayesian methods can be found in Carlin and Louis (2009).

The Bayesian approach has several attractive features. One is the ability to provide the full posterior distribution for the calculation of all the model parameters (e.g., mean and variance parameters) and all *functions* of model parameters. Using this posterior distribution, we can obtain point estimates, such as the posterior median, as well as 95% credible intervals. These 95% credible intervals are the Bayesian analogue to the 95% confidence intervals used in classical statistics. Theoretically, these quantities have different interpretations, but in practice they are often viewed as being equivalent. As such, notationally we will use the expression "95% CI" to refer to both the Bayesian credible interval and the classical confidence interval.

Another attractive feature of the Bayesian method is the use of prior information (e.g., priors based on expert opinions or pilot study data that are expressed in the form of probability distributions). In the absence of prior information, it is common to use the so-called "non-informative" priors for the model parameters (e.g., distributions with large variances or uniform distributions with large ranges) that let the data drive the posterior distribution and thus the statistical inference. However, when data are limited, whether due to small sample size (e.g., N < 10) or to a very high degree of censoring (e.g., percent censoring > 80%), informative priors are often useful. If the prior distribution is generally consistent with the observed data, the posterior estimates can be more accurate and precise than those derived from using non-informative priors. On the other hand, a poorly chosen informative prior (say, with a mean that is far from the truth and an unlikely small variance) can lead to a biased posterior inference. Thus, the process for specifying informative priors must be both transparent and scientifically acceptable and the posterior's sensitivity to the choice of the prior should be assessed.

Bayesian model for left-censored data

As is common for exposure data, we have chosen to specify our likelihood as a lognormal distribution. That is, our log-transformed exposure measurements are modeled as being normally distributed with mean and variance parameters, μ and σ^2, respectively. In the industrial hygiene literature, μ corresponds to the log of the GM and σ corresponds to the log of the GSD. Once the posterior distribution for μ and σ^2 are found, the posterior distributions for other model parameters of interest such as the GM, GSD, AM, and $X_{0.95}$ can be computed.

To model censored data, typically one would construct their likelihood as a product of the PDF for the detected observations and the CDF for the censored observations. Thus, the censored observations' contribution to the likelihood is simply the probability that a value would be censored, and parameter estimates can be obtained by maximizing the likelihood. Under the Bayesian framework, however, we can consider the censored observations, denoted as $Y_{i,cen}$, as missing values. We look to obtain a posterior distribution for these values (denoted as the vector as $\boldsymbol{Y}_{cen}$) in addition to the model parameters, μ and σ^2, given our observed (or detected) values, denoted $\boldsymbol{Y}_{det}$). Using this and Bayes Theorem from Eq. (1), we can construct our hierarchical model as

$$p(\mu, \sigma^2, \boldsymbol{Y}_{cen} | \boldsymbol{Y}_{det}) \propto p(\mu, \sigma^2) \times \prod_{\boldsymbol{Detect}} N\,(\log\,(Y_{i,det}) | \mu, \sigma^2) \times I\,\{Y_{i,det} > LOD_i\}$$

$$\times \prod_{\boldsymbol{Censored}} N\,(\log\,(Y_{i,cen}) | \mu, \sigma^2) \times I\,\{Y_{i,cen} \leq LOD_i\}$$

(2)

In this expression, $p(\mu, \sigma^2)$ denotes the prior distribution of our model parameters (to be defined later), and I{} denotes an "indicator function" which takes the value 1 when the expression inside the brackets is true and takes the value 0 otherwise; this will restrict our imputed censored observations from being larger than their respective LODs.

To fit this model, we used a Markov chain Monte Carlo (MCMC) algorithm to estimate the censored values and the model parameters (Robert and Casella, 2005; and Carlin and Louis, 2009). Imputation of censored values requires sampling from truncated distributions taking values below the LOD (Gelfand et al., 1992). Calculating the parameters from each of these distributions results in samples of our parameters from the posterior distribution in Eq. (2), from which we can obtain full posterior distributions for GM=exp(μ), GSD=exp(σ), and the 95th percentile or $X_{0.95}$=exp(μ +1.645*σ). We applied the uniform minimum variance unbiased estimator (MVUE) method for calculating the AM (Finney, 1941).:

$$\widehat{AM}_{\text{MVUE}} = \exp(\mu)\psi\left(\frac{\sigma^2}{2}\right) \quad (3)$$

where

$$\psi(g) = \left[1 + \frac{(n-1)g}{n} + \frac{(n-1)^3 g^2}{n^2(n+1)2!} + \frac{(n-1)^5 g^3}{n^3(n+1)(n+3)3!} + \frac{(n-1)^7 g^4}{n^4(n+1)(n+3)(n+5)4!} + \cdots\right]$$

In this expression, $g = \frac{\sigma^2}{2}$ and n=sample size. Using the first 5 terms of this infinite sum is usually sufficient to provide precise estimates. While complicated, this expression is generally less biased than the maximum likelihood estimate for the AM (i.e. $\mu + \frac{\sigma^2}{2}$) particularly for small sample sizes (Huynh et al., 2014).

Simulation Study

The overall accuracy of an estimation method using censored data depends on a number of factors. We conducted three sets of simulations to compare the β-substitution method and the Bayesian method similar to the procedure described in Huynh et al. (2014). Here, we briefly describe the simulation.

Our first simulation set represents the most basic case (Figure 17 in the On-line Supplement Materials). We generated simulated data from lognormal distributions with a true GM = 1, and true GSDs = 2, 3, 4, and 5, and a single LOD corresponding to an expected percent censoring, p, ranging from 10% to 90% in increments of 10%. For each of our conditions (i.e. each combination of N, GSD, and percent censoring), we generated and analyzed 1000 data sets

The second simulation set follows directly from the first set except we investigated the impact of multiple LODs. The issue of multiple LODs, however, is addressed differently by the two censored data methods. The β-substitution method takes the average of all the LODs and uses this average in the algorithm as if the dataset had a single LOD. This is in contrast to the Bayesian approach described in Equation (2), which allows each censored observation to have a unique LOD. We simulated two scenarios (with small and large differences between the LODs) to assess the effect of the difference between the LODs. The small gap multiple LODs simulation used p_1=expected censoring level, $p_2=0.95*p_1$, and $p_3= 0.90*p_1$. The large gap multiple LODs simulation used p_1=expected censoring level, $p_2=2/3* p_1$, and $p_3=1/3*p_1$. While the latter is less frequently encountered in typical industrial hygiene practices, it is more applicable to the GuLF STUDY where many exposure groups have highly varying values of LODs due to differences in sampling duration (4-18 hrs). Lastly, our third simulation attempted to model data drawn from a *mixed* distribution, which, in our case, is a bimodal lognormal distribution, which can occur when limited sampling documentation results in grouping disparate measurements. We simulated two types of mixed distributions where one mixed distribution is a combination of two lognormal distributions with GMs of 1 and 5, and the other with GMs of 1 and 10.

As in all simulations, the observed censoring for a given dataset may deviate from the expected censoring. Any dataset that was observed to be 100% censored was discarded because all methods used here are inappropriate such datasets. Fully censored datasets occurred more frequently under high censoring and small sample size conditions. For a given expected censoring probability, p, and a given sample size, N, we expect to observe datasets that are fully censored with a probability p^N(e.g., when N=5 and p=80%, the expected percent of 100% censored datasets = 32.76%).

All computation was programmed in statistical computing software R (R Development Core Team, 2014).

Priors for the simulation

In industrial hygiene practice, eliciting priors for μ and σ is often simplified by specifying bounds on the GM and the GSD, respectively. As such, for our comparisons we created a relatively non-informative (or "weakly informative") prior for μ, which was uniformly distributed between log(0.05) and log(500). These values were chosen such that the range was too broad to provide much information to model the simulation where the true GM=1. Similarly, we created a non-informative prior for σ using a uniform distribution bounded between log(1.01) and log(12). We consider the priors used here to be "non-informative" because their bounds are sufficiently large relative to the "true" parameter values. In theory, however, one may need to use larger bounds—or perhaps unbounded distributions with a very large variance—to achieve truly non-informative priors. In order to assess the effect of narrower bounds on our posterior distributions, we also analyzed the data from our first simulation scenario using the more informative priors listed in Table 1. For the informative priors we retained the lower bound for μ, but narrowed the upper bound to log(50). We used log(1.01) and log(4) as the lower and upper bounds of σ if the true GSD =2 and 3 and log(3) and log(6) if the true GSD = 4 and 5.

Evaluation metrics

Both methods were evaluated using relative bias, relative root mean squared error, and coverage probability. Relative bias (called bias hereafter) is the difference between the average estimated value and the true value relative to the true value

$$Relative\ Bias = 100 \times \frac{\bar{x}-\theta}{\theta} \qquad (4)$$

where $\bar{x}$ is the average estimate of the parameter of interest (e.g., AM, GM, GSD, or $X_{0.95}$) from the 1,000 trials, and θ is the true value. Bias can be negative or positive; negative bias indicates that the method underestimated the true value of the parameter while positive bias indicates overestimation. The relative root mean squared error (called rMSE hereafter) is a measure that combines the bias and the precision of the method relative to the true value and can only be positive. The rMSE is generally considered a better metric than bias because rMSE allows the additional evaluation of the precision of the method.

$$Relative\ rMSE = 100 \times \frac{1}{\theta}\sqrt{(\bar{x}-\theta)^2 + \frac{\sum(x_i-\bar{x})^2}{N-1}} \quad (5)$$

Coverage probability is estimated as the proportion of the estimated parameters that are located within the 95% confidence interval/credible interval . The desired coverage probability for a 95% uncertainty interval is 0.95.

Algorithms for computing the confidence interval for the mean of the β-substitution method have not been provided, so we adapted an algorithm that was intended for non-censored data (Hewett and Ganser, 1997) to compute the confidence interval for the AM from β-substitution method. This approach allows us to compare the coverage for the AM for the β-substitution and the Bayesian method. We found no method to estimate β-substitution confidence intervals for the GM, GSD, or $X_{0.95}$, and thus, make no coverage comparisons for these parameters.

Results

The results for the lognormal distributions with a single LOD and with small gap multiple LODs simulations were generally very similar. Therefore, only figures for the single LOD simulation set are shown in herein. Figures for the GM under these conditions, simulations with larger gaps between the multiple LODs and simulations with mixed distributions are shown in the On-line Supplemental Materials. In these figures, the mean relative bias and rMSE for each of the four parameters (GM, GSD, AM, $X_{0.95}$) are shown for the simulated conditions. The size of the circle in the figures corresponds to the magnitude of the mean bias or the mean rMSE on a continuous scale. The size of each circle generally falls between two of the circles identified with specific values in the legend. For the sake of brevity, we refer to GSD ≤ 3 as low variability and GSD ≥ 4 as high variability. Sample sizes of 5 are considered small, while those between 10 and 20 are considered moderate and above 20 are considered large. Generally, the smaller the bias and rMSE, the better the performance.

Lognormal distributions and single LOD

Arithmetic Mean (AM)

Figures 1 showed that the Bayesian method with non-informative priors had a slightly higher bias than the β-substitution method in the estimation of the AM. The rMSE for both methods was small and comparable under most conditions (Figure 2). The exception was under the condition of small to moderate sample sizes and high variability s, where the Bayesian method had a larger bias but a smaller rMSE. The use of informative priors generally improved the accuracy and precision of the Bayesian estimates as expected. Figure 3 shows histograms of the 1000 AM estimates for N=5, p=70, and GSD=4 to provide insight on the behavior of the two methods. This condition was chosen to investigate the small N, high censoring, and high variability scenarios where the performance of the methods deviated from each other. For small Ns, the distributions tended to be skewed but both methods provided similar distributions. The β-substitution had a slightly longer tail, which resulted in the larger rMSE compared to that of the Bayesian method.

Figure 4 illustrates the estimated coverage probabilities for the β-substitution and the Bayesian methods for the AM. The Bayesian models with non-informative and informative priors generally provided coverage comparable to the β-substitution method in many simulated conditions. The exceptions were the condition of small sample sizes and censoring >40% and also large Ns and low censoring where the β-substitution method performed worse. The results showed that the β-substitution method did not provide as good as coverage as the Bayesian method. As mentioned earlier, the β-substitution method does not provide uncertainty estimates for other parameters but the Bayesian method does.

Geometric Mean (GM)

The bias for the GM for the Bayesian method with non-informative priors and the β-substitution method were comparable for large sample sizes and when censoring ≤ 80% conditions (Figures 1 and 2 in the On-line Supplemental Materials). The β-substitution method tended to overestimate the GM under all conditions, whereas the Bayesian method with non-informative priors tended to underestimate the GM and informative priors generally had mixed overestimation and underestimation.

Despite the differences in bias, the rMSE for the GM for both methods appeared to be similar. This discrepancy is likely due to the contribution of the "precision" component of rMSE—i.e. the Bayesian approach provides equally (or perhaps more) precise results than the β-substitution method, thus offsetting the differences in bias. As for coverage, the β-substitution method does not currently provide an estimate for the uncertainty in GM, but the 95% CI from the Bayesian approach consistently provided coverage near the ideal 95% (Excel file in the Supplemental Materials)

Geometric Standard Deviation (GSD)

The β-substitution method generally had smaller bias than the Bayesian method with non-informative priors but with informative priors the Bayesian method was more comparable to the β-substitution method (Figure 5). As the GSD increased, the bias for the β-substitution method increased for small to moderate sample sizes and high censoring conditions, while the bias of the Bayesian method with non-informative priors decreased. The Bayesian method's trend was likely influenced by the fact that as the true GSD moved closer to the median of the GSD of the prior distribution (i.e. GSD prior=(1.1+12)/2 =6.5), the Bayesian model was able to provide a better estimate of the GSD (i.e. with lower bias)

The Bayesian method generally had smaller rMSE than the β-substitution method, indicating better precision (Figure 6), particularly when informative priors were used. The β-substitution method tended to provide large GSDs more frequently than the Bayesian method with non-informative priors whereas the bounded informative priors for the GSD limited the Bayesian method's ability to provide larger GSDs (data not shown).

While the β-substitution method failed to provide measures of uncertainty for estimates of the GSD, the 95% credible intervals generated by the Bayesian method provided the desired coverage probability of 0.95 (Excel file in the On-line alal Materials).

95th percentile ($X_{0.95}$)

The β-substitution method generally provided smaller bias than the Bayesian method in the estimation of the $X_{0.95}$, although the use of informative priors reduced bias for the Bayesian method (Figure 7). With respect to the rMSE metric, the β-substitution method and the Bayesian method generally provided comparable rMSE for sample sizes ≥ 20 (Figure 8). For small to moderate sample size conditions, the β-substitution method generally had smaller rMSE than the Bayesian method with non-informative priors and comparable bias to the method with informative priors. As with the other parameters the coverage probability for the Bayesian method was generally close to the target 95% coverage (Excel file in the On-line Supplemental Materials)

Other simulations

In the multiple LODs simulation where the gaps between LODs were large, the bias and rMSE in the estimation of the AM and the GM for both methods were generally comparable under most conditions (Figures 3 and 4 for the AM; Figures 6 and 7 for the GM in the On-line Supplemental Materials) The β-substitution method generally had higher bias in the estimation of the AM than the Bayesian method at censoring ≤ 60% and a higher rMSE for small to moderate sample sizes and high variability. The coverage for the β-substitution method was generally much lower than the 95% target compared to the Bayesian method's performance, especially at ≥60% censoring and higher variability, even with large sample sizes conditions (Figure 5 in the On-line Supplemental Materials). In the estimation of the GSD, the β-substitution method and the Bayesian method provided comparable bias and rMSE for large sample sizes except at censoring >80% (bias) and small to moderate sample sizes (rMSE) (Figures 8 and 9 in the On-line Supplemental Materials). Lastly, in the estimation of the 95th percentile, the β-substitution method generally provided comparable or smaller bias (particularly when the variability was low) and rMSE than the Bayesian method, particularly under small sample sizes conditions (Figures 10 and 11 in the On-line Supplemental Materials). The general poor performance of β-substitution method under some conditions might be because the β-substitution uses an averaged LOD that substantially deviated from the original LODs, whereas the Bayesian method used each of the original LODs. This type of scenario is less frequently encountered in typical exposure assessment but may occur in retrospective occupational exposure assessments that typically use multiple sources of data for estimation of historical exposure levels. The Bayesian coverage for the GM, GSD, and $X_{0.95}$ were generally close to the 95% target (Excel file in the Supplemental Materials)

The results for the simulation of mixed distributions with the GMs of 1 and 5 ss looked very similar to those of lognormal distributions with a single LOD (figures not shown). This was likely due to the fact that the overall distribution looked more lognormal than a bimodal distribution. As the distance between the modes increased (i.e. $GM_1 = 1$ and $GM_2 = 10$), the bias for the estimation of the AM derived from the Bayesian method with non-informative priors was higher than the β-substitution method but the rMSE for both methods was small and comparable (Figures 12 and 13 in the On-line SM). The β-substitution method also appeared to perform better than the Bayesian method in the estimation of the $X_{0.95}$ as shown in Figures 14 and 15 in the On-line Supplemental Materials. In terms of coverage, the Bayesian method had much lower coverage at GSD=2 compared to the other GSDs (shown in Figure 16 in the On-line Supplemental Materials). This is likely due to the fact that at low variability, the two modes were more distinct than the high variability data. At GSD=2, the coverage for the β-substitution method and the Bayesian method were equally poor but at higher GSD, the Bayesian coverage was slightly better than β-substitution method as in the single LOD and multiple LOD simulations.

An illustrated example

To further demonstrate the behavior of the β-substitution and the Bayesian methods, we analyzed a small subset of the GuLF STUDY data using both methods. The GuLF STUDY is a long-term epidemiological study initiated by the National Institute of Environmental Health Sciences (NIEHS) to investigate the potential adverse health effects associated with the exposures to multiple agents that occurred during the response and clean-up of the *Deepwater Horizon* oil release in the Gulf of Mexico. The exposure assessment component of the study involves the analysis of approximately 140,000 personal air measurements, of which about 60 percent was below the detection limits of the analytic laboratory. Despite the large number of available measurements, as we divide the data into exposure groups (EGs), defined by chemical, location, vessel, time period, tasks, and activities, the EGs will have varying level of censoring.

We used the following non-informative (weakly informative) priors for the Bayesian method:

$\mu \sim$ Normal (0, 1000),

$\sigma \sim$ Uniform (log(1.01), log(12))

The notation for μ denotes the normal distribution on the log scale with the mean of 0 and a very large variance of 1000. The prior for σ takes a uniform distribution with the minimum and maximum GSDs of 1.01 and 12, respectively. In this example, the non-informative normal distribution (Carlin and Louis, 2005; Busschaert et al., 2011; Paulo et al, 2005) was used. The Bayesian model was executed in JAGS (Plummer, 2003), and processed in R.

Figure 9 (and Table 2 in the On-line Supplemental Materials) show the results of our analysis for 13 EGs with xylene exposures measured on one of the rig vessels during on the *Deepwater Horizon* response. The AM estimates for both methods were comparable for most EGs. The upper bounds of the 95%CI for the β-substitution method, however, were much greater for a few groups. In contrast, because the GSD estimates from the Bayesian method had been restricted by the maximum prior (i.e. GSD=12), some of the Bayesian GSDs were close to the upper bound, but did not exceed it because of the boundary requirements. A sensitivity assessment of the priors (i.e. reanalyzing the data using an unbounded uniform prior or a prior with a larger upper bound) would likely result in the Bayesian method providing higher estimates of the GSD.

Discussion

The β-substitution method and the Bayesian method with non-informative priors often produced accurate point estimates for many of the simulated conditions. Both methods' performance varied, however, when estimating differing parameters. Generally, the Bayesian method with non-informative priors was comparable to the β-substitution method when estimating the AM and GM but it was more biased in the estimation of the GSD and the 95th percentile, although the use of more informative priors resulted in more comparable performances. The Bayesian method generally provided consistent and better coverage of the AM than the β-substitution method.

The β-substitution method's ease of implementation in a simple spreadsheet can be an attractive feature to many practitioners. If, however, coverage is important, this method is not recommended. It currently does not have a method for providing uncertainty estimates for the GM, GSD or $X_{0.95}$. Even for the AM, we had to adapt a method for non-censored data described by Hewett and Ganser, 1997 to allow comparison with the Bayesian 95% CI. It is possible that this adaptation was not appropriate for censored data and that this was the reason that the 95% CI of the β-substitution method was worse than the Bayesian method. Under the condition of multiple LODs that were much further apart, the β-substitution method performed worse than the Bayesian method because β-substitution method uses an averaged LOD that, in many cases, substantially deviated from the original LODs, whereas the Bayesian method used each of the original LODs. This type of scenario (i.e. multiple LODs) is less frequently encountered in typical exposure assessments but may occur in retrospective occupational exposure assessments that typically use multiple sources of data for estimation of historical exposure levels.

The performance of the Bayesian method depends on the choice of priors , which will vary with different studies. Thus, any detailed insights (such as the direction or magnitude of bias) from these informative prior simulations, other than that they improved the Bayesian method's performance over the non-informative priors, is not appropriate. That is because different informative priors have different influences on the posterior distribution that cannot be captured in a simulation study. While the Bayesian method generally can provide accurate point estimates and full distribution to all the parameters using relatively non-informative priors, good priors may be difficult to identify under conditions where they are needed most, i.e. small sample sizes or high censoring if other information is not available. In addition, some computational skills are also needed to obtain the full benefits of the Bayesian methods. The WinBUGS codes for the Bayesian censored data are provided in the On-line Supplemental Materials.

Our simulation study attempted to assess a wide range of conditions; however, as with any simulation study, these conditions might not be representative of all the conditions that one might encounter in an occupational exposure assessment study. For example, the largest GSD tested in our simulations was 5 but preliminary results of the GuLF STUDY data suggested that some EGs had data with much higher GSDs. We tested the mixed distribution simulation to assess how far we can stretch the lognormality assumption for these parametric methods and found that the Bayesian method (non-informative priors) generally did not work well for censored data that were substantially bimodal. Since most exposure data generally have a lognormal distributional shape, this assumption might not be important, and if the modes are close enough, the Bayesian method might suffice. However, if the data are distinctively bimodal, the results from the Bayesian method might be less predictable than the β-substitution method.

Conclusions

Our simulation study compared the β-substitution method and a Bayesian method for estimating parameters of exposure distributions from censored data. We have shown that both methods generally delivered accurate point estimates, while only the Bayesian method provided reliable uncertainty intervals for all four parameters investigated. The β-substitution method was generally less biased and was easier to implement but its measure of uncertainty was less reliable for the AM, and uncertainty could not be estimated for the other parameters. We recommend that the practitioner take into account the data available, the purpose of the estimation, and the need for uncertainty estimates when selecting a method with censored data. The Bayesian method may be particularly useful if the practitioner has the computational resources and prior information, as the method generally provides accurate estimates and also provides the distributions of all of the parameters.

Acknowledgements: This research was supported in part by the Intramural Research Program of the NIH, National Institute of Environmental Health Sciences.

References

Aitchison J, Brown JAC (1957) The Lognormal Distribution with Special Reference to Its Uses in Economics. cambridge University Press, Cambridge, UK.

Antweiler RC, Taylor HC. (2008) Evaluation of statistical treatments of left-censored environmental data using coinci- dent uncensored data sets: I. Summary statistics. Env Sci Technol; 42: 3732–8.

Busschaert P, Geeraerd AH, Uyttendaele M, and Van Impe JF. (2011). Hierarchical Bayesian analysis of censored microbiological contamination data for use in risk assessment and mitigation. Food microbiology, 28(4), 712–9.

Carlin, B.P. and Louis, T.A. (2009). Bayesian Methods for Data Analysis, 3rd ed. Boca Raton, FL: Chapman and Hall/CRC Press.

Cohen, AC. (1959) Simplified estimators for the normal distribution when samples are singly censored or truncated. Technometrics 1, 217-37.

Cohen, AC. (1961) Tables for maximum likelihood estimates: Singly truncated and singly censored samples. Technometrics 3, 535-541.

Finkelstein MM, and Verma DK. (2001) Exposure estimation in the presence of nondetectable values: another look. Am Ind Hyg Assoc J; 62:195-8.

Finney DJ. (1941) On the Distribution of a Variate Whose Logarithm Is Normally Distributed. Journal of the Royal Statistical Society, Supplement 7:155-1 61.

Fisher, RA. (1925) Theory of statistical estimation. Proceedings of the Cambridge Philosophical Society 22, 700-725.

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

Gillespie BW, Chen Q, Reichert H et al. (2010) Estimating population distributions when some data are below a limit of detection by using a reverse Kaplan-Meier estimator. Epidemiology; 21: S64–70.

Gelfand, AE, Smith, AFM., and Lee, TM. (1992). Bayesian analysis of constrained parameter and truncated data problems using Gibbs sampling. Journal of the American Statistical Association, 87, 523--532.

Helsel DR. (2005) Nondetects and data analysis. New York: John Wiley & Sons, Inc.

Helsel DR. (2010) Much ado about next to nothing: Incorporating nondetects in science. Ann Occup. Hyg ;54:257-62.

Hewett P and Ganser GH. (1997). Simple Procedures for Calculating Confidence Intervals around the Sample Mean andExceedance Fraction Derived from Lognormally Distributed Data. *Applied Occupational and Environmental Hygiene*, *12*(2), 132–142.

Hewett P, and Ganser GH. (2007) A comparison of several methods for analyzing censored data. Ann Occup Hyg;51:611-32.

Hewett P, Logan P, Mulhausen J, Ramachandran G, and Banerjee S. (2006). Rating exposure control using Bayesian decision analysis. *Journal of occupational and environmental hygiene*, *3*(10), 568–81.

Hornung RW, and Reed LD. (1990) Estimation of average concentration in the presence of non-detectable values. Appl. Occup. Envir. Hyg. 5(1):46–51.

Huynh T, Ramachandran G, Banerjee S, Monteiro J, Stenzel M, Sandler DP, Kwok RK, Engel LS, Blair A, Stewart PA. "Comparison of methods for analyzing left-censored data to estimate worker's exposures for the GuLF STUDY.'" Accepted for publication at the Annals of Occupational Hygiene in April 2014

Kaplan EL, and Meier P. (1958) Nonparametric estimation from incomplete observations. J Am Stat Assoc **1958**; 53:457-81

KrollCN, Stedinger JR. (1996) Estimation ofmoments and quan- tiles using censored data.Water Resour Res; 32: 1005–12

Leidel, NA, Busch KA, Lynch JR (1977) Occupational Exposure
Sampling Strategy Manual. National Institute for Occupational Safety and Health Pub. No. 77-173 (available from the National Technical Information Service, Pub. No. PB274792)

Mulhausen J, Damiano J, editors. (1998) A strategy for assess- ing and managing occupational exposures. 2nd edn. Fairfax, VA: American Industrial Hygiene Association.

Paulo, MJ, Van der Voet H, Jansen, MJW, Ter Braak, CJF, and Van Klaveren, JD . (2005). Risk assessment of dietary exposure to pesticides using a Bayesian method. *Pest management science*, *61*(8)

Plummer M. (2003) JAGS: A program for analysis of Bayesian graphical models using Gibbs sampling.

R Development Core Team. (2014) R: a language and environment for statistical computing. R Foundation for Statistical Computing: Vienna. http://www.R-project.org.

Robert, Christian and George Casella, 2005, Monte Carlo Statistical Methods, New York, Springer

Selvin S, Rappaport SM (1989). A Note on the Estimation of the Mean Value from a Lognormal Distribution. Am. Ind. Hyg. ASSOCJ. 50~627-630

Singh A, Maichle R, Lee SE. (2006) On the computation of a 95%upper confidence limit of the unknown population mean based upon data sets with below detection limit observations. Washington, DC: U.S. Environmental Protection Agency EPA/600/R-06/022

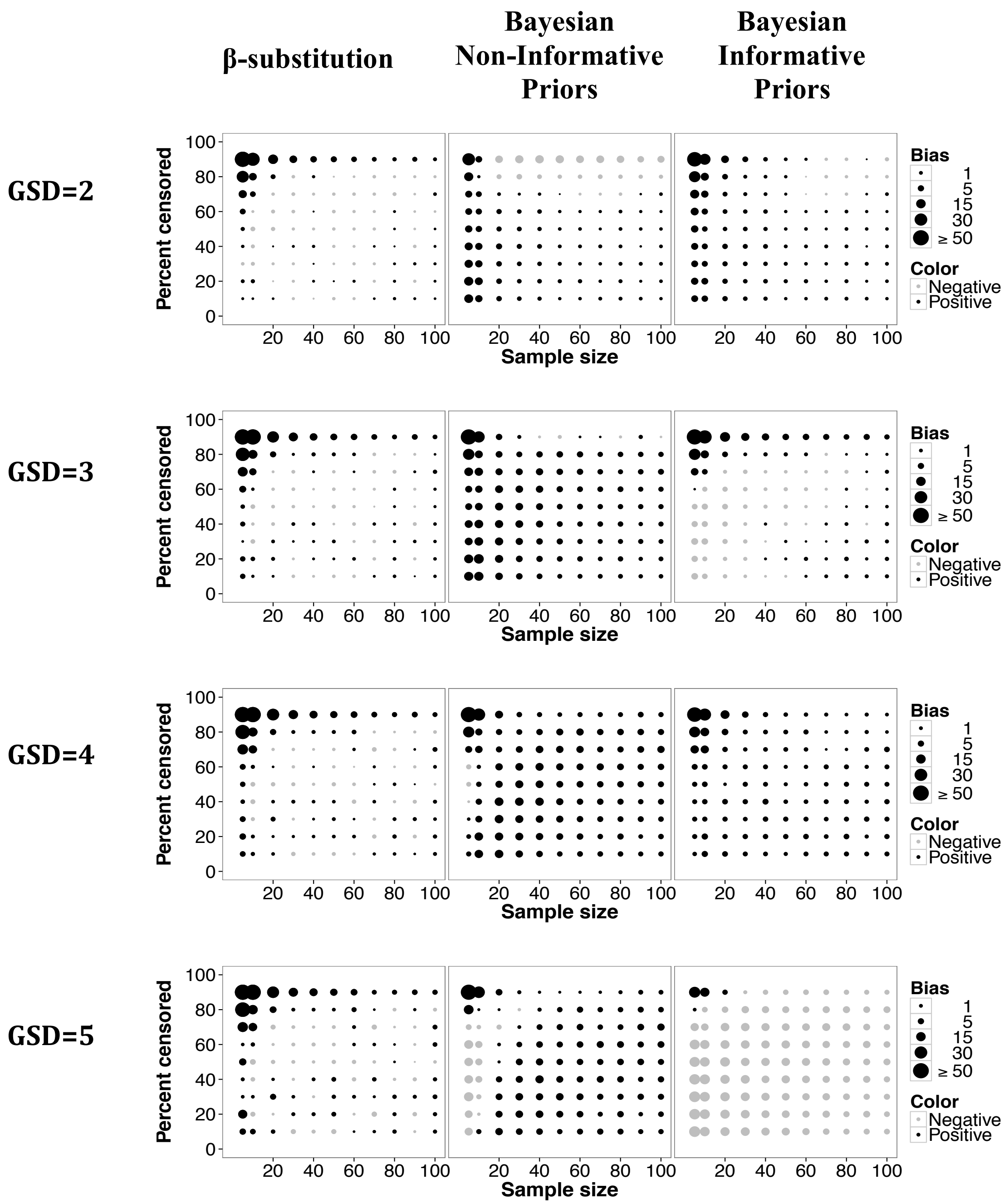

Figure 1: Relative bias in the estimate of the AM of a lognormal distribution and a single LOD for different sample sizes, percent censoring, and GSDs for the β-substitution method and the Bayesian method with non-informative and informative priors.

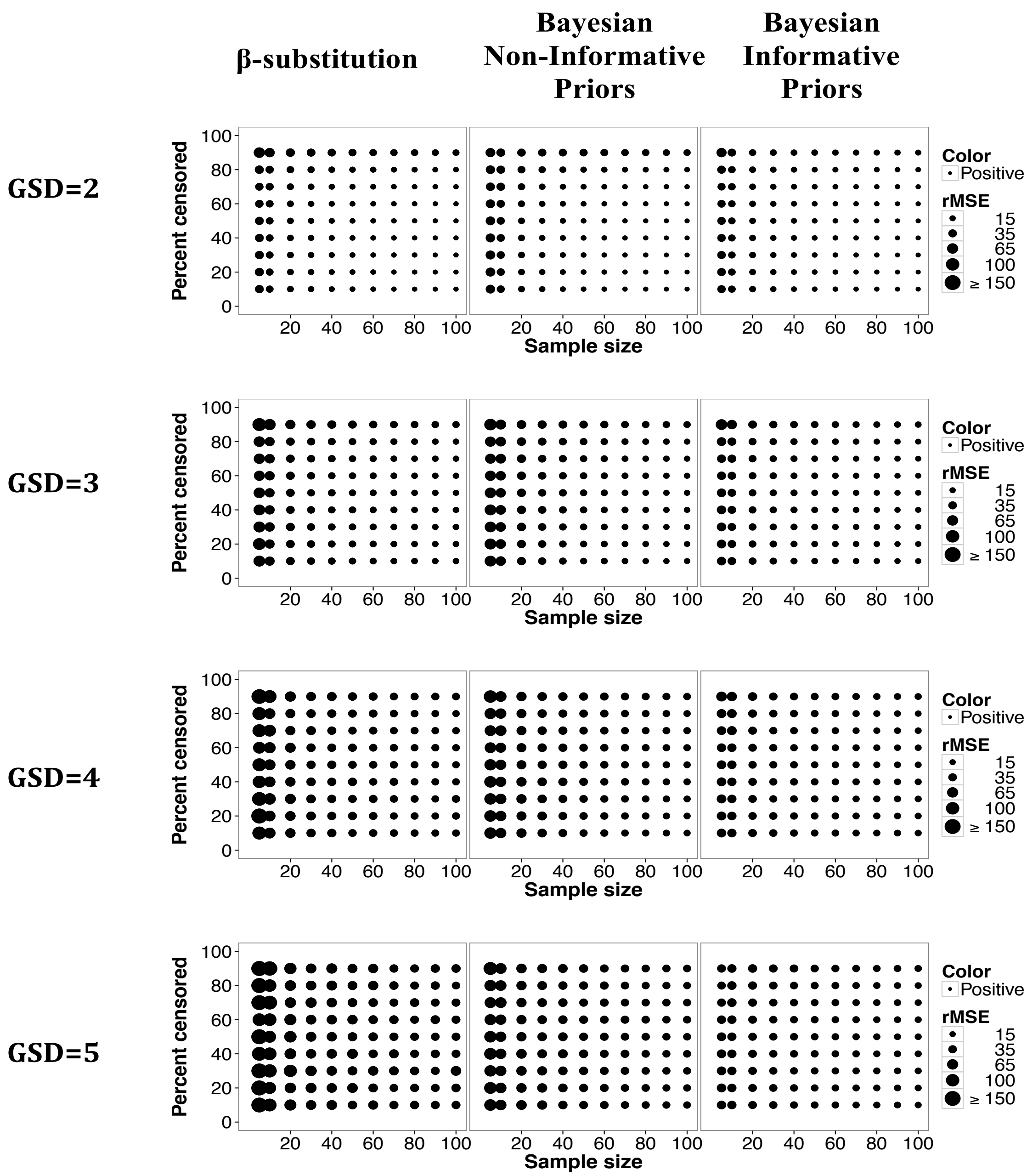

Figure 2: Relative rMSE in the estimate of the AM of a lognormal distribution and a single LOD for different sample sizes, percent censoring, and GSDs for the β-substitution method and the Bayesian method with non-informative and informative priors.

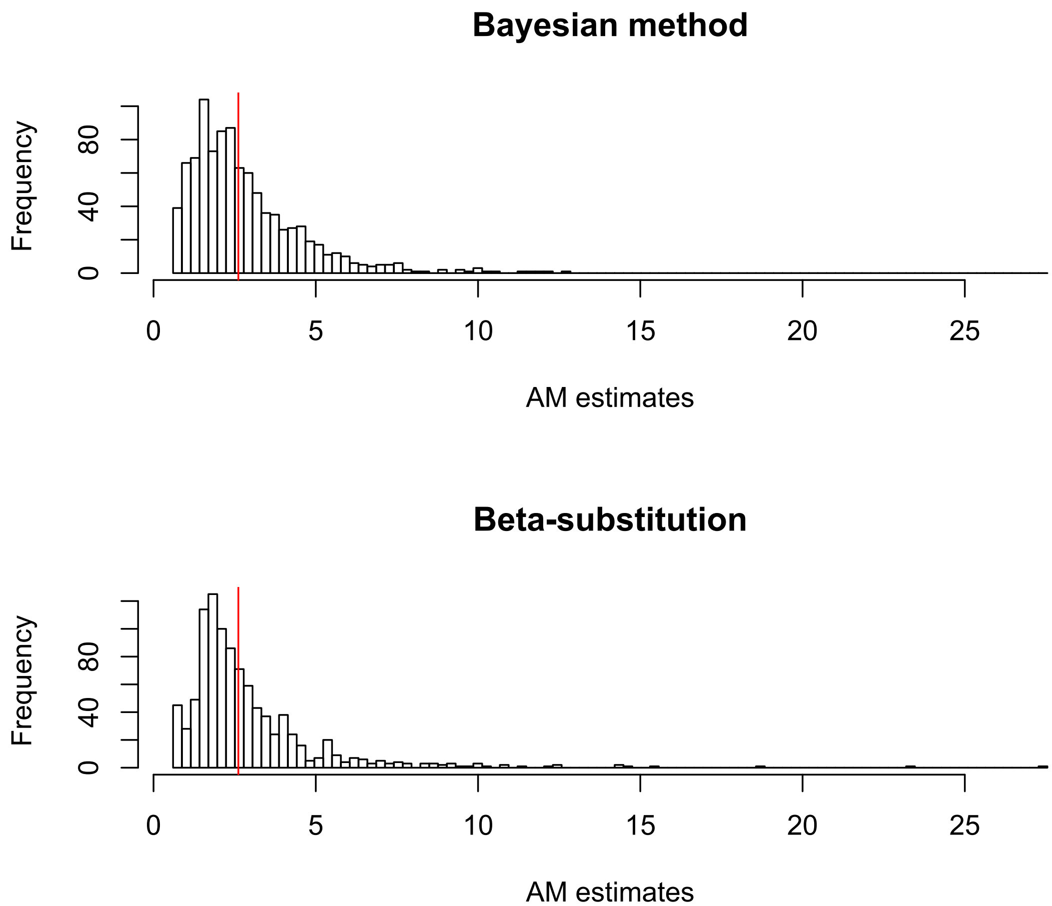

Figure 3: Histograms of the 1000 estimates of the AM for N=10, p=70, GSD=4 condition for the β-substitution method and the Bayesian method with non-informative priors. The line at approximately 2.5 indicates the true AM.

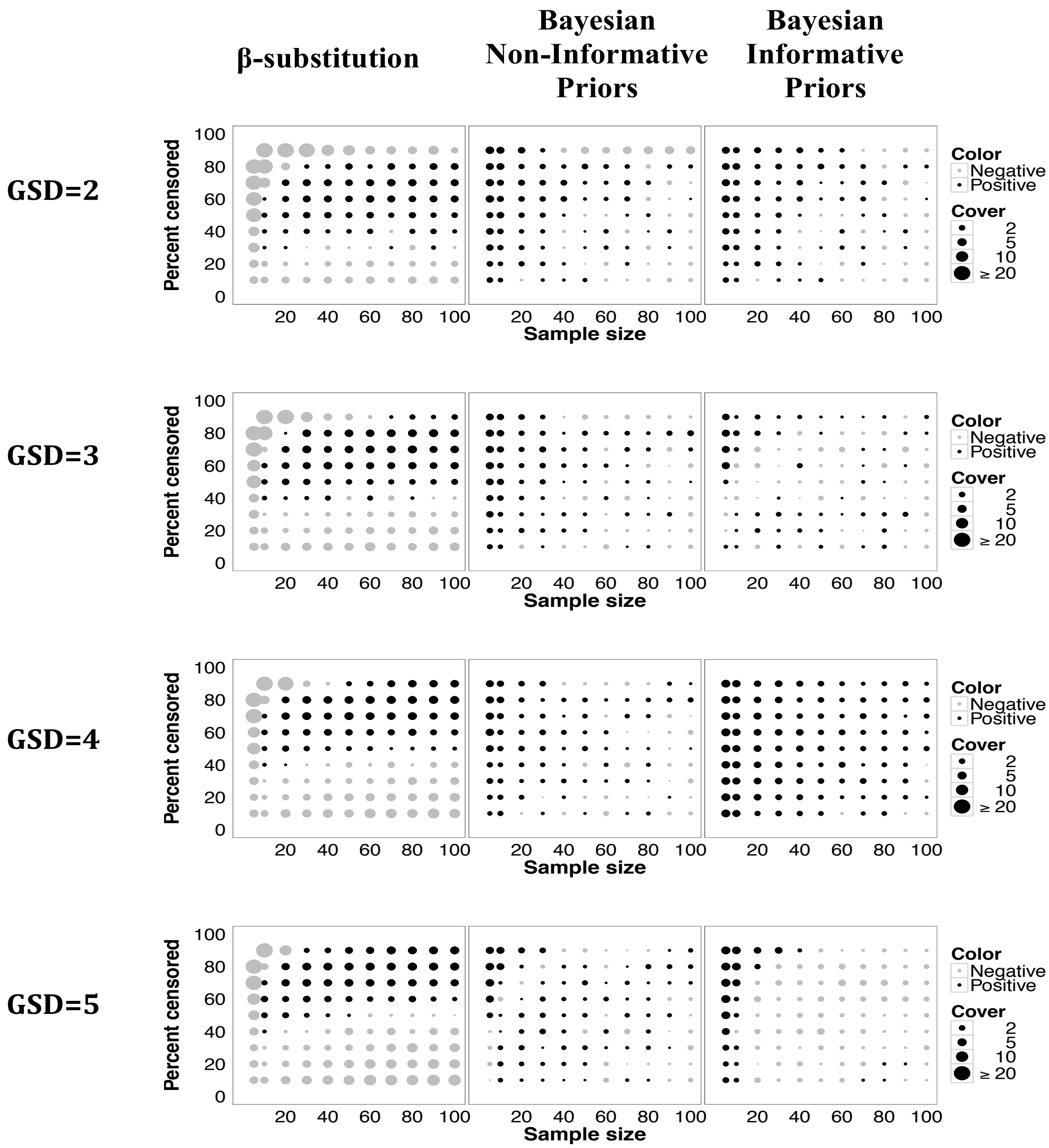

Figure 4: Coverage probabilities (in percent) for the AM of a lognormal distribution and a single LOD for the β-substitution method and the Bayesian method with non-informative and informative priors. The size of the circle represents difference between the actual coverage probability minus the target 95%. Positive dots indicate probabilities larger than 95%. While 100% coverage is not as desirable because the uncertainty estimates maybe too wide to be informative, large negative coverage (approximately <90%) might be worse because it indicates that the interval frequently missed the true value.

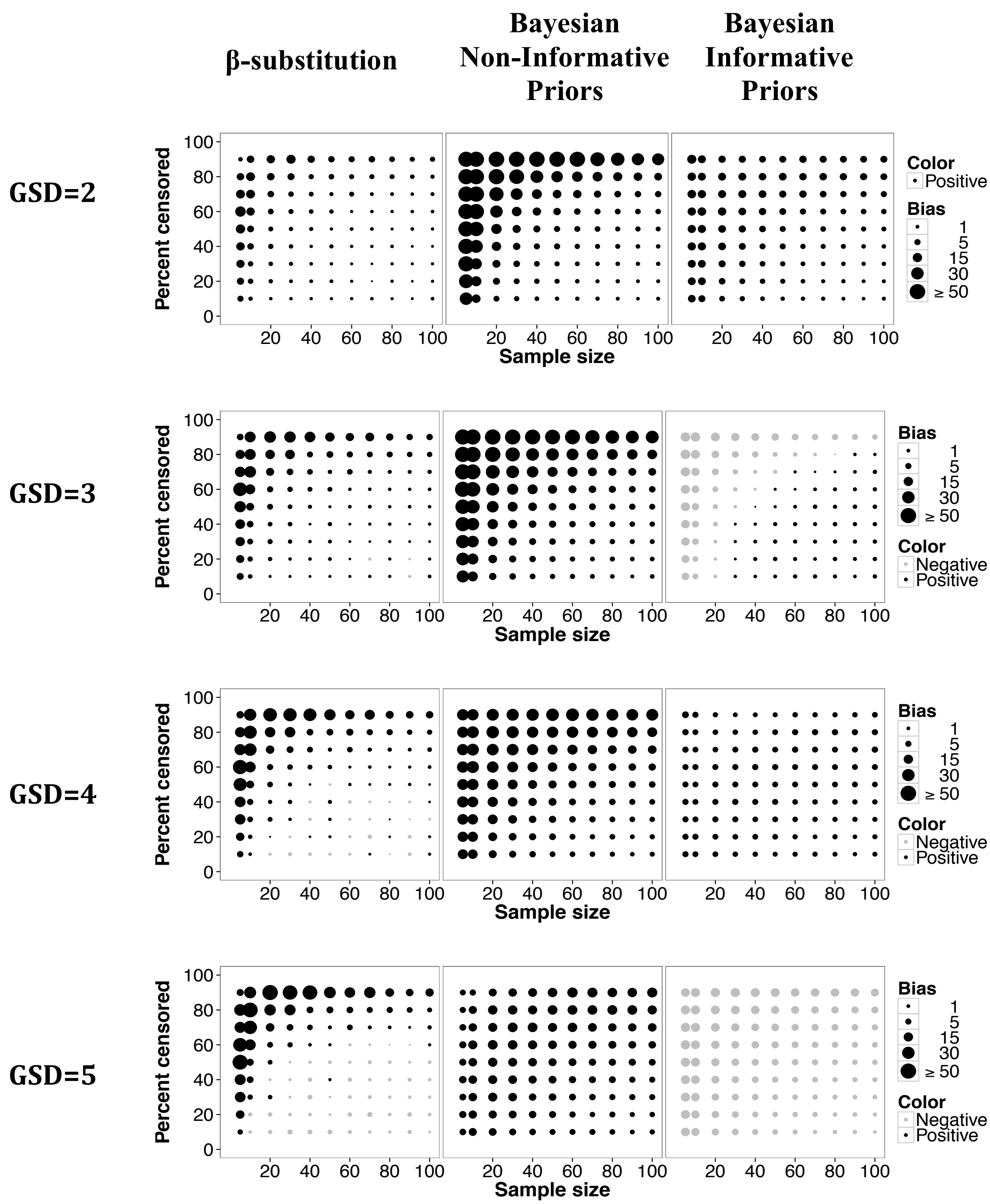

Figure 5: Relative bias in the estimate of the GSD of a lognormal distribution and a single LOD for different sample sizes, percent censoring, and GSDs for the β-substitution method and the Bayesian method with non-informative and informative priors.

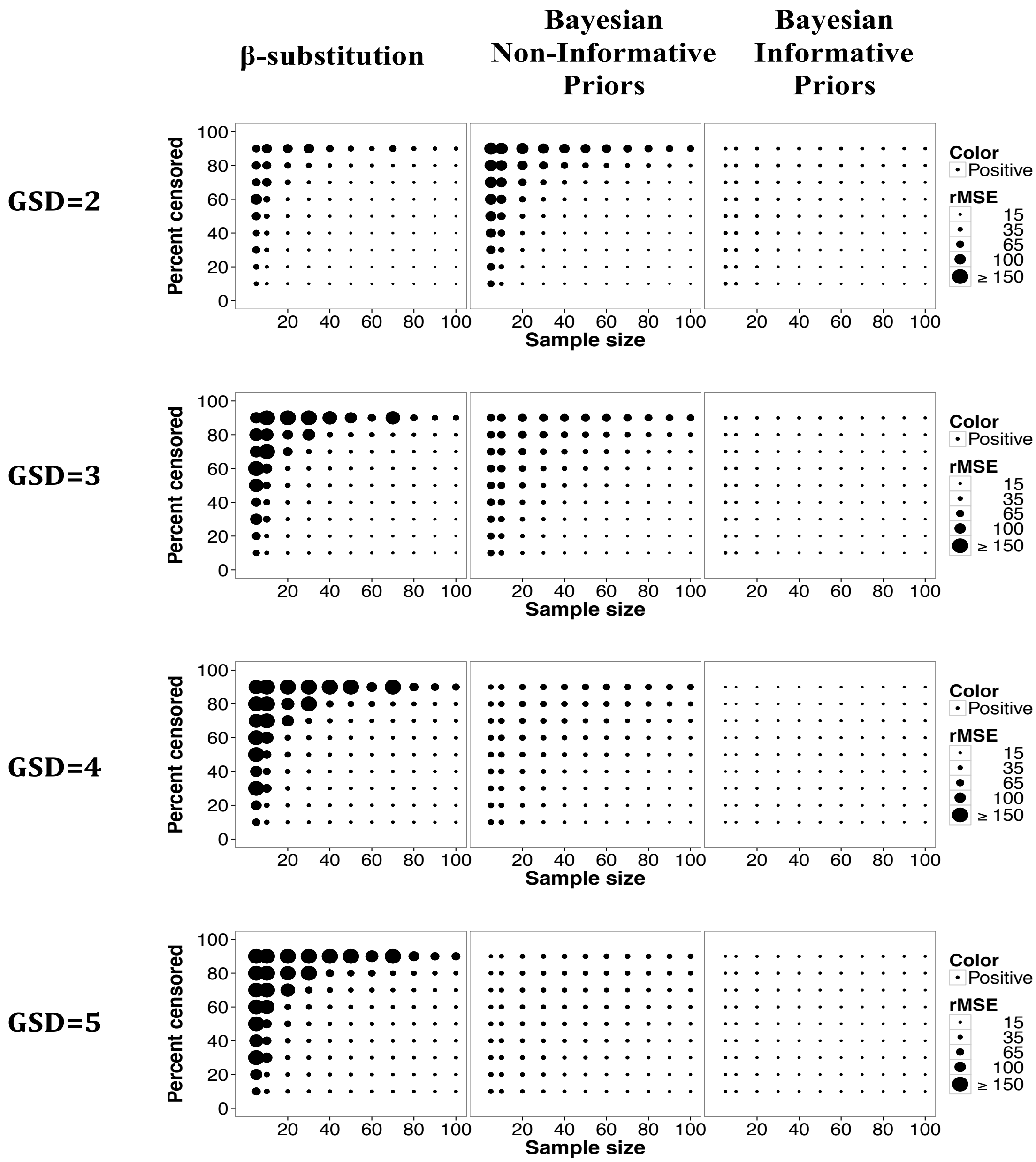

Figure 6: Relative rMSE in the estimate of the GSD of a lognormal distribution and a single LOD for different sample sizes, percent censoring, and GSDs for the β-substitution method and the Bayesian method with non-informative and informative priors.

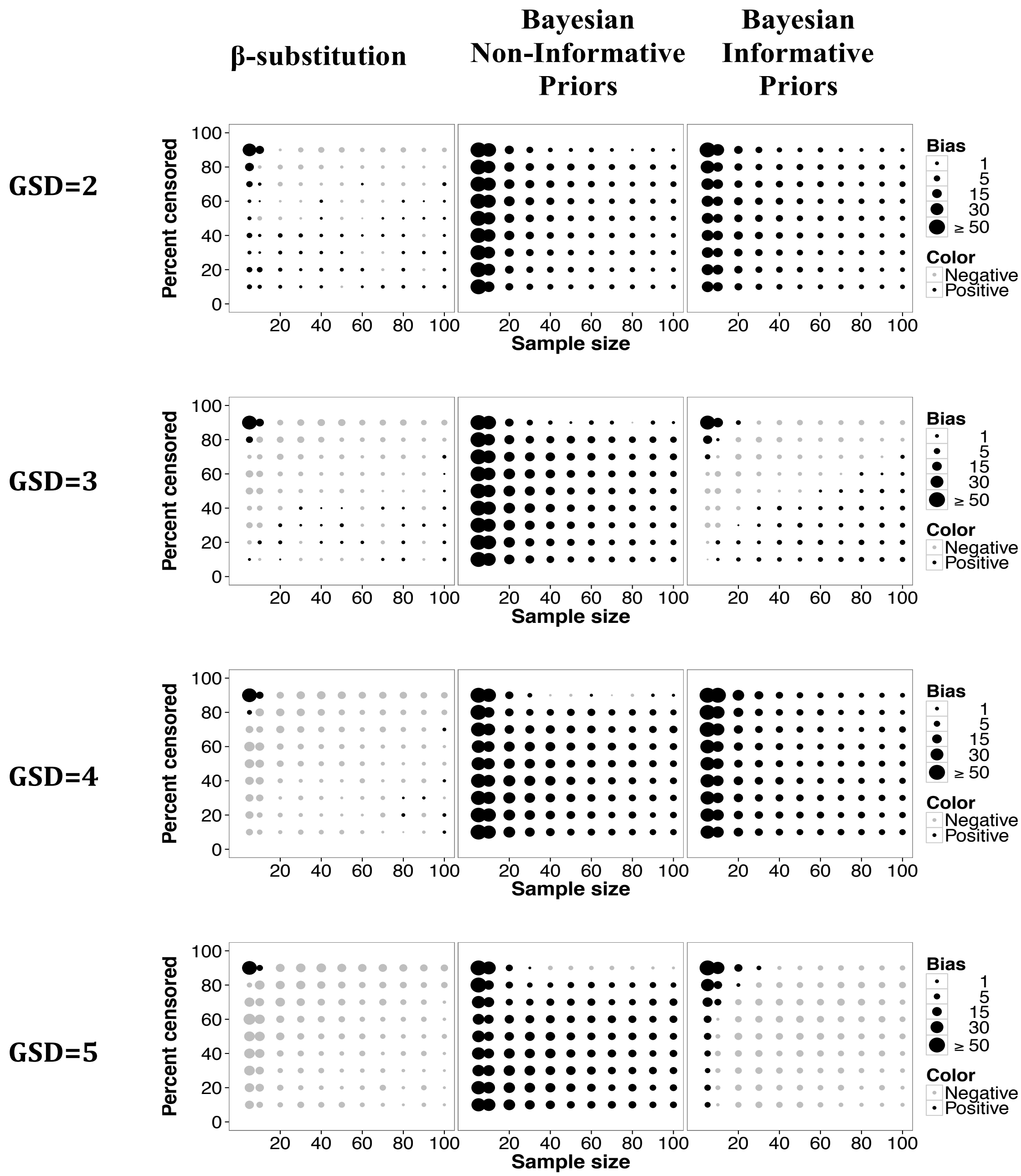

Figure 7: Relative bias in the estimate of the 95th percentile of a lognormal distribution and a single LOD for different sample sizes, percent censoring, and GSDs for the β-substitution method and the Bayesian method with non-informative and informative priors.

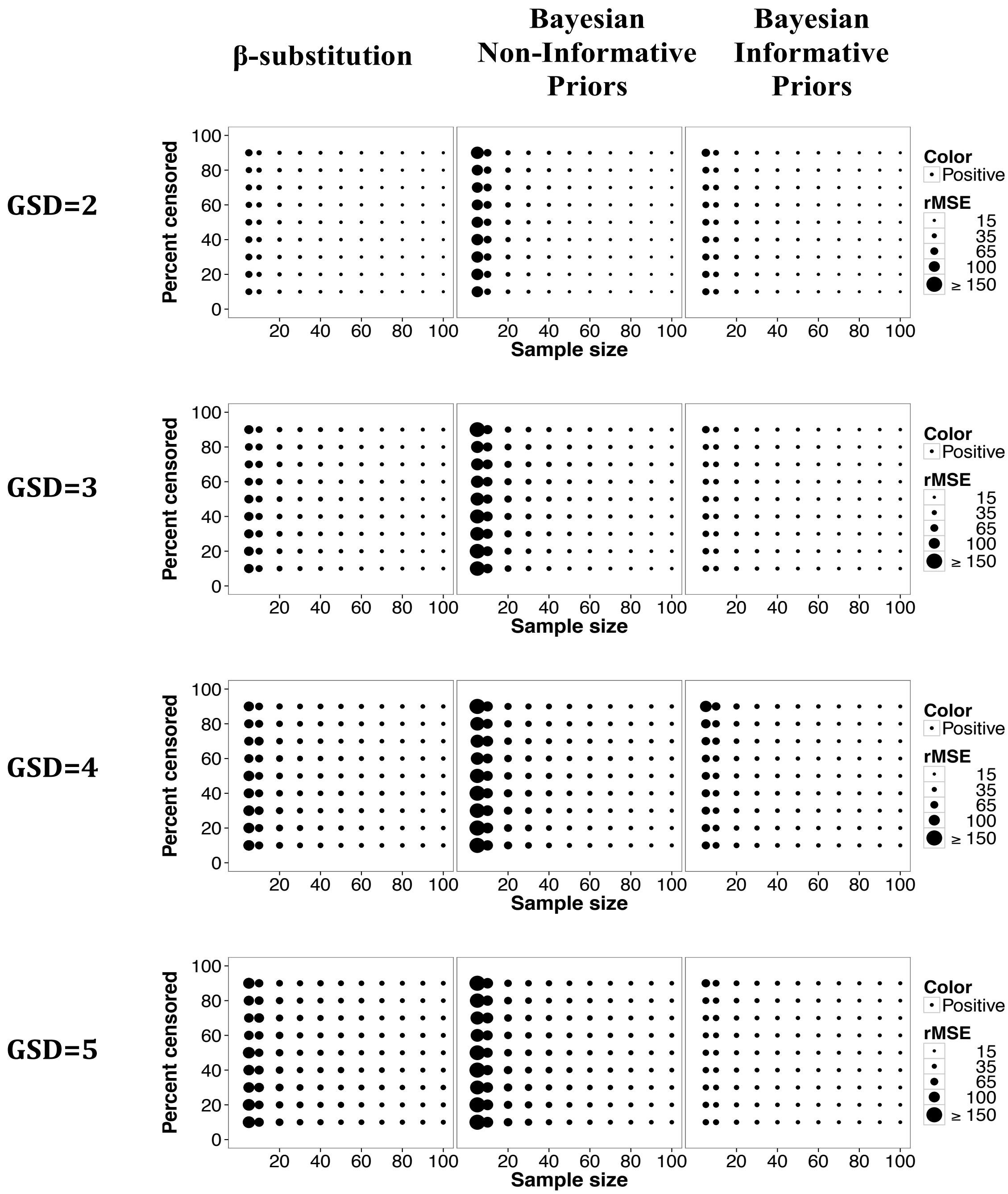

Figure 8: Relative rMSE in the estimate of the 95^{th} percentile of a lognormal distribution and a single LOD for different sample sizes, percent censoring, and GSDs for the β-substitution method and the Bayesian method with non-informative and informative priors.

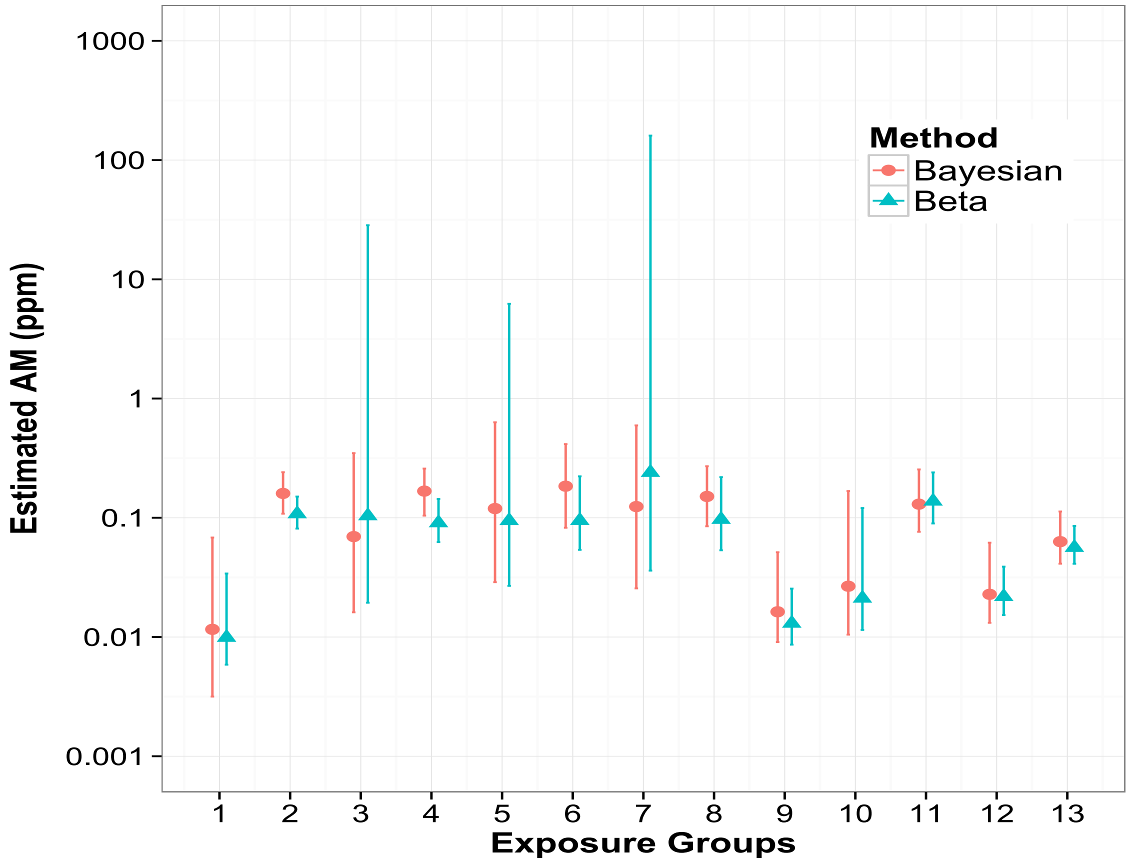

Figure 9: Estimates of the AM and their uncertainty using the β-substitution method and the Bayesian method with non-informative priors.

Table 1: Priors specification for μ and σ of the log-transformed data

Truth	Non-Informative Priors	Informative Priors
$\mu = 0$ (GM = 1)	$\mu \sim$ Uniform (ln(0.05), ln(500))	$\mu \sim$ Uniform (ln(0.05), ln(50))
$\sigma = 0.7, 1.1, 1.4, 1.6$ (GSD = 2,3,4,5)	$\sigma \sim$ Uniform (ln(1.01), ln(12))	$\sigma \sim$ Uniform (ln(1.01), ln(4)) if true GSD = 2,3 $\sigma \sim$Uniform (ln(3), ln(6)) if true GSD = 4,5

Chapter IV

Estimates of occupational inhalation exposures on the four rig vessels during the *Deepwater Horizon* oil release clean-up

Tran Huynh, Gurumurthy Ramachandran, Catherine Groth, Sudipto Banerjee, Mark Stenzel, Dale P. Sandler, Lawrence S. Engel, Richard K. Kwok, Aaron Blair, and Patricia A. Stewart

To be submitted to the Annals of Occupational Hygiene

Summary

After the 2010 *Deepwater Horizon* oil release, the National Institute of Environmental Health Sciences initiated an epidemiological study (GuLF STUDY) to investigate the potential adverse health effects associated with the oil spill response and clean-up work. Quantitative exposure assessment is a critical component of the GuLF STUDY because it allows the investigation into the exposure-disease relationship. This study presents the methodology and results of the inhalation exposures for workers on the four main rig vessels (Enterprise, DD2, DD3, and Q4000) that were responsible for stopping the leak in the hot zone closest to the well site. We used personal measurements that were collected by BP and its contractors during clean-up period. Exposure groups (EGs) were created on based on chemical, location, vessel, time period, and job titles/tasks. Bayesian method were used to analyzed exposures for total hydrocarbons (THCs), benzene, toluene, ethylbenzene, xylene (BTEX chemicals) and hexane. THC measurements were least censored compared other chemicals evaluated. THC exposures changed over time and varied by vessels and exposure groups. Highest exposures were generally observed in the time period from when the leak start until the well was successfully top capped. Exposures gradually decreased over time in most exposure groups except a few that might be involved in the decontamination effort. BTEX chemicals and hexane exposures were substantially lower than THC. The variability of the EGs for the GuLF STUDY were generally high, reflecting the non-routine, time-dependent nature of spill response efforts as well as the challenges of retrospectively constructing exposures for oil spill study.

Introduction

On April 20, 2010, the *Deepwater Horizon* rig exploded and subsequently released millions of gallons of crude oil in the Gulf of Mexico. During the response and clean-up of the oil release that lasted until December 31, 2010, more than 55,000 workers were rostered to have been involved in the clean-up (NIOSH, 2011). A large number of these workers reported ill health symptoms including headaches, nausea, lower and upper respiratory irritations, heat stress, mental and physical fatigue, eye and skin irritation (NIOSH, 2010). As part of the comprehensive federal response the incident, the National Institute of Environmental Health Sciences (NIEHS) initiated an epidemiological study (GuLF STUDY) to investigate the possible short-term and long-term adverse health effects experienced by the workers (Sandler et al., 2012).

The characterization of exposures is a critical component of any occupational and environmental epidemiological study because it allows us to investigate the exposure-disease relationship. Very few human exposure assessment and epidemiological studies for past oil spills are described in the literature. There have been many major oil spills around world in the 20th century but human health studies have been conducted only for seven major spills (Anguilera et al., 2010). Most of these were cross-sectional studies and did not have quantitative estimates of chemical exposures (Palinkas et al., 1992,1993; Campbell et al., 1993, 1994; Crum, 1993; Lyons et al., 1999; Gallacher et al., 2007; Dor et al., 2003; Suarez et al., 2005; Carrasco et al., 2006, 2007;, Zock et al., 2007; Sabucedo et al., 2009; Janjua et al., 2006). In studies that had air monitoring results, few samples were collected for a very small number of contaminants during the spill (Campbell et al., 1993; Morita et al., 1999; Meo et al., 2008, Laffon et al., 2006; Perez-Cadahia et al., 2006, 2007). Nonetheless, these studies generally reported respiratory and dermal symptoms similar to those experienced during the BP oil spill clean up. Evidence of genotoxicity and endocrine toxicity has also been found in-vitro study (Amat-Bronnet et al, 2007) and epidemiological studies (Laffon et al., 2006, Perez-Cadahia et al., 2006, 2007, 2008a, 2008b, 2008c) in addition to bioaccumulation of oil compounds in marine food (Lemiere et al., 2005, Chaty et al, 2008). Psychological and social effects associated with oil spills were also common findings in human studies (Palinkas et al., 1992, 1993, 2004; Gill and Picou, 1998; Sabucedo et al., 2009; Gallacher et al. 2007, Zock et al., 2007; Janjua et al., 2006; Carrasco et al., 2006, 2007; Morita et al., 1999).

Crude oil contains a number of volatile chemicals but only six oil-related chemicals were selected for assessment of inhalation exposures: total hydrocarbons

(THCs), benzene, toluene, ethylbenzene, xylene (BTEX chemicals), and hexane. These chemicals were selected because they have been associated with adverse health effects in previous spills and BTEX chemicals are particularly toxic (IRIS, 2003 and 2005). In addition, due to their volatility they are more likely to be inhaled by the workers. They also have the largest number of measurements collected over the clean-up period (Stewart et al., 2014)

Over the response and clean-up period, more than 150,0000 personal exposure measurements for several contaminants were collected by BP or its contractors at various locations for a variety of tasks. These measurements will be used to characterize exposure for specific exposure groups (EGs), defined by location, time period, vessel, and job titles or tasks. For example, the location of the spill remediation was divided into five areas: the hot zone (within a mile radius from the well head), the source (~5 mile radius from the well, excluding the hot zone), off shore (other than the source), and land. Exposures to oil components changed over time due to weathering and could also be impacted by decisions to contain the oil such as the use of dispersants, capping of well at the wellhead on July 15, bottom capping, and refurbishing of the vessels. Over 9000 large and small vessels were involved in the clean-up operations including the large oil drilling platforms responsible for controlling the leak, research and governmental marine vessels, ships with remotely operated vehicles (ROVs), large and small ships called vessels of opportunities (VoOs) that assisted in the oil skimming and burning, and other supporting ships and barges that carried fuel and personnel (USCG, 2011). Some of these vessels had very unique responsibilities that would lead to different level of exposures of the oil components.

This paper focuses the exposures on the four oil drilling rig vessels in the hot zone: the DD2, DD3, Enterprise, and Q4000. The Enterprise was located directly above the well head and was collecting oil and gas while supporting the containment efforts. At the same time, BP tried several attempts to control leak using remotely operated vehicles (ROVs). First BP tried activate the safety device called the blowout preventer (BOP) using the ROVs but failed. Then they tried to put a temporary dome on the BOP but had to remove it because the dome's opening was plugged by methane hydrate crystals. Next, they inserted a riser insertion tube tool to divert the flow through the riser so that some of the oil can be collected by the Enterprise. They also tried to pump various materials through the BOP called 'top kill' procedure but also failed. They finally put a custom-made cap over damaged well and stopped the majority of the leak on July 15, 2010. The Q4000 was involved in the failed 'top kill' operation and then was used to flare gas and pump mud. The DD2 and DD3 drilled the two relief wells so that they can pump mud through the main well (Cleveland, 2013). The majority of the personal measurements were collected on these four rig vessels, and the rest on the ROVs, VoOs, and on land.

Despite the large number of measurements available, a substantial number of these measurements were below the limits of detection (LOD) reported by the analytical laboratory (left-censored). In order to provide accurate exposure estimates and accounting for these censored values, we investigated various statistical methods for analyzing censored data in the literature and conducted simulation studies to identify a method that is most suitable for our needs (Huynh et al, 2014a and 2014b). A Bayesian method was selected for this study because it was found to provide point estimates that were comparable to classical censored data analysis methods such as the maximum likelihood (ML) method and the β-substitution method in many simulated scenarios, and also provided accurate credible intervals for all parameters including the arithmetic mean (AM), the geometric mean (GM), the geometric standard deviation (GSD), and the 95th percentile ($X_{0.95}$) (Huynh et al, 2014b).

The exposure assessment component of the GuLF STUDY comprises of four sub-studies: occupational inhalation exposure, spatial/environmental exposure, dermal exposure, and linkage of estimates to the workers. This article presents occupational inhalation exposure estimates to THC, BTEX, and hexane on the four rig vessels in the hot zone.

Methods

Classification of EGs

The overall strategy for developing exposure groups for the GuLF STUDY is discussed extensively in Stenzel et al. (2015). Here, we summarized the strategy to provide the context in which the EGs on the rig vessels were developed. A comprehensive list of exposure groups was developed using information from a number of sources including site visits, time history reports from the personal samples, persons-on-board lists, public domain documents, and completed questionnaires from the workers.

Exposure groups were classified based on the chemical, type of vessel, geographical location, job group (e.g., job titles, tasks, and activities) and time period. For location, the spill response and clean-up effort was divided into five areas: the hot zone (within a nautical mile radius around the well site), the source (outside the hot zone, within a five nautical miles of the well site), off shore (other than source), near shore (approximately three nautical miles from shoreline), and land. The hot zone was a restricted area where the four rig vessels were stationed to repair the main well, collect oil and gas, and drill the relief wells. Using records on the job titles, tasks, and activities, ahierarchy of EGs was established for each vessel. A total of approximately 65 exposure groups were developed for four vessels. Some groups were present on more than one vessel but some groups were also unique to a particular vessel depending on the functions of the vessel. The hierarchy has four tiers, with Tier 1 being the most general and Tier 4 most specific. Figure 1 shows the hierarchy with selected exposure groups on the Enterprise as an example.

The EGs were also divided into time periods to reflect key events that might have a significant impact on the exposures. Time period 1a (TP1a) was from April 20 – May 14 and it represents the period when the oil was leaking and dispersants were not used.

Time period 1b (May 15 – July 15) began with the first injection of dispersants at the well head (~ 5000 feet below the sea surface) on May 15 and simultaneous application at the sea surface and ended with the successful top capping on July 15. The dispersants chemical mixtures (Corexit 9527 and Corexit 9500) were used to break down the oil so that the oil was more rapidly digested by bacteria. The majority of the leak stopped during time period 2 (July 15^{th} – August 10^{th}) until the well was mostly sealed at the bottom on August 10. Decontamination, equipment refurbishing and other clean-up activities lasted until Dec 31, 2010 in time period 3 (August 10 – December 30, 2010).

Data collection

BP contractors monitored personnel with the highest potential exposure using organic vapor badges (3M 3500 or 3520, or Assay Technology 521) and the badges were analyzed for multiple chemicals including THCs, BTEX, hexane, heptane, cyclohexane, trimethylbenzenes, tetrahydrofuran (Stewart et al, 2014). The duration of these samples varied from less than one hour to more than 24 hours. We excluded samples with duration < 4 hours and >18 hours so that our estimates would better reflect full-shift exposures. The excluded samples made up less than 6% of the total (Stewart et al, 2014).

Table 1 shows the final number of measurements for the chemicals of interest on each rig vessel. THC and the BTEX chemicals had the same number of measurements because these chemicals were analyzed using the same badge across all time periods. Hexane, on the other hand, was not analyzed until after time period 1b and thus had fewer measurements. The highest number of measurements were collected on the DD3 and Enterprise, followed by the DD2 and the Q4000. Benzene had the highest percent of censored data, followed by ethylbenzene, hexane, toluene and xylene, with THC having the lowest percent censoring.

Analysis of censored data using the Bayesian approach

We analyzed the censored data using a hierarchical Bayesian framework presented in Huynh et al. (2014b). Exposures data are log-transformed. Using Bayes Theorem, the hierarchical model is constructed as follow:

$$p(\mu, \sigma^2, \boldsymbol{Y}_{cen}|\boldsymbol{Y}_{det}) \propto p(\mu, \sigma^2) \times \prod_{\boldsymbol{Detect}} N\,(\log{(Y_{i,det})}|\mu, \sigma^2) \times I\,\{Y_{i,det} > LOD_i\}$$
$$\times \prod_{\boldsymbol{Censored}} N\,(\log{(Y_{i,cen})}|\mu, \sigma^2) \times I\,\{Y_{i,cen} \le LOD_i\} \quad (1)$$

In this expression, p(μ, σ^2) denotes the prior distribution of our model parameters, and I{} denotes an "indicator function" which takes the value 1 when the expression inside the brackets is true and takes the value 0 otherwise; this will restrict our imputed censored observations from being larger than their respective LODs. A Markov chain Monte Carlo (MCMC) algorithm was used obtain posterior distributions for censored values (denoted by the vector $\mathbf{Y}_{cen)}$ and also the model parameters, μ and σ^2, given our observed (or detected) values, denoted $\mathbf{Y}_{det}$. (Robert and Casella, 2005; and Carlin and Louis, 2009). Censored values were imputed by sampling from the normal truncated distributions taking values below the LOD (Gelfand et al., 1992).

We then estimated the posterior distributions of the model parameters GM = exp(μ), GSD = exp(σ), the AM using the minimum variance unbiased estimator (Finney, 1941), and the 95th percentile or $X_{0.95}$=exp(μ +1.96*σ) for each EG on each of the rig vessels.

Bayesian computation was performed in JAGS (Just Another Gibbs Samplers) (Plummer, 2013) and data processing in R (R Core Development Team, 2013)

Priors for THC

For THC analysis, we specified uniform priors for μ and σ. The minimum and maximum for μ or ln(GM) were set at ln(0.025) and ln(50) ppm, respectively. The minimum GM was based on prior knowledge/professional judgment that it should not be lower than 10*the lowest LOD (0.25). BP was using an exposure limit of 100 ppm for THC. If exposures were at the exposure limit, the average would be in the range of 25 to 30 ppm. The GM would likely be less than 20 ppm. BP intervened with actions such as surface application of dispersant, moving the ships to assure that concentration did not exceed the exposure limit. Therefore in the GuLF STUDY, the upper GM was conservatively specified at 50 ppm.

The minimum and maximum priors for σ or ln(GSD) were specified at ln(1.01) and ln(12), respectively. While the minimum GSD for an EG cannot be lower than 1, the maximum GSD observed using uncensored measurements for any EG in the GuLF STUDY was 12. Such high GSDs are likely due to the highly variable and non-routine tasks being performed during an oil spill clean-up operations.

Priors for BTEX and hexane

The priors for BTEX and hexane analysis were developed based on the correlation between THC and each of the BTEX chemicals and hexane. A separate Bayesian regression analysis that accounted for censored observations for both analytes provided the distributions of the regression coefficients β_0 and β_1 (Groth et al., 2015). All analytes were log-transformed and reported in ppb units from the model

$$\ln(\text{analyte}_i) = \beta_0 + \beta_1 * \ln(\text{THC}_i) + \varepsilon_i \qquad \text{(Equation 1)}$$

Groth et al. (2015) also found significant differences in the values of β_1 between ships, time periods, inside living quarters of the ships (Inside) versus outside operations on the ships (Outside).

Table 2 shows the median and credible intervals for the intercepts and slopes separated by Inside and Outside, chemicals, and by time period on the Enterprise (results for other ships are in the On-line Supplement Materials). The Inside measurements for four rig ships were combined because we expect the exposures inside living quarters across four ships to be similar The Outside measurements were analyzed separately for each vessel. Our regression analysis was performed at the broad level (i.e., EGs were combined) because the correlation was more easily observed at this level than at the EGs level. The assumption is that the observed correlation at the broad level would be similar at the EGs level. Some chemicals such as benzene or ethylbenzene in time period 3 were highly censored so their results were not used.

Figure 2 illustrates the resampling procedure for using the regression coefficients to generate priors for each EG using xylene as an example. From the posterior distributions of ln(GM) and ln(GSD) for THC, a random value of GM and GSD is selected to simulate one distribution of THC exposures. A random value of the intercept, the slope, and the error from their respective posterior distributions is selected to develop a distribution of xylene using Equation 1; the the ln(GM) and ln(GSD) of that distribution were computed. The algorithm is repeated 1000 times to obtain the prior distributions of the GM and GSD of xylene. The prior distributions ln(GM) and ln(GSD) for most EGs generally resembled normal and gamma distributions respectively.

Results

In this paper, we report the results and discussion for the AM and $X_{0.95}$ exposures because the AM is considered an appropriate metric for estimating cumulative exposure for chronic disease investigation in occupational epidemiologic study (Seixas et al, 1988; Rappaport, 1991) and $X_{0.95}$ can be used as a measure of peak exposure for acute symptoms. For presentation purposes, the AM and $X_{0.95}$ exposures were presented for selected exposure groups across time periods and across ships. These exposure groups were selected if they had estimates in more than one time period within a ship or were present in more than one vessel in order to facilitate between-ship comparisons. Estimates of the GM, GSD, AM and $X_{0.95}$ for all EGs are included in the Excel On-line Supplemental Materials. Within each vessel, there were many EGs but results for those with $N < 5$ and/or censoring level > 80% were not reported. THC estimates were reported in parts per million (ppm) and BTEX and hexane in parts per billion (ppb).

THC estimates

Figure 3 shows THC estimates of the AM for selected EGs on four rig vessels. The Enterprise generally had highest exposures compared to other three ships before the well was successfully top capped. Exposures after top capping and bottom capping across four ships appeared to be similar. The DD2 and Q4000 did not have estimates for TP1a. On the DD2, TP1b generally had higher exposures than TP2 and TP3 except the Outside.Operations.ROV group in TP3 which had higher exposures than TP2 and TP1b. TP2 exposures were higher than TP3 for Crew exposure groups (e.g., Outside.Crew.CraneOperator and Outside.Crew.FloorhandOrRoughneck) but not for

Operations (Outside.Operations.ROV). The relative ranking of exposures among groups appeared to vary by time periods. For example, the Outside.Crew.CraneOperator had highest exposure in TP1b but the Outside.Operations had highest exposure in TP2. Surprisingly, the Inside group had similar exposure as some of the outside groups.

On the Enterprise, exposures were highest during TP1b, followed by TP1a, then TP2, and TP3 for Crew exposure groups that had estimates for at least two time periods. This trend appeared to be consistent across those EGs. No estimates were reported for sub-groups in the Operations for TP1a, TP2 and TP3 due to lack of samples. For TP1b, the Inside Group had the lowest exposure compared to all the Outside EGs. Substantial reduction in exposures was observed in TP2 compared to TP1b, and a slight reduction in TP3 compared to TP2.

The DD3 had highest exposures during TP1a, followed by TP1b across all EGs. TP2 exposures were lower than TP1a and TP1b and higher than TP3 for some exposure groups (Inside and IHSafety) but not others (Outside.Crew, Ouside.Crew.Roustabout, and Outside.Operations). The Inside group had similar exposures to those of the Outside group.

The Q4000 did not have measurements for Inside and also had fewer EGs than the other three ships. Highest exposures were observed in TP1b, followed by TP3 and TP2. The uncertainty intervals overlapped some EGs but not others across four rig ships.

In the estimates of the 95th percentile, the exposure trend across time periods for each ship were similar to the AM estimates (Figure 4). The estimates were approximately an order of magnitude higher than the AM estimates. Similarly to AM estimates, the uncertainty intervals for the 95th percentile overlap some EGs but not others.

The GSD estimates for THC varied from 1.5 to 9.6. Inside EGs generally had lower variability than most Outside EGs. The uncertainty intervals for GSD estimates were within 12.

BTEX estimates

The benzene measurements on all four ships were highly censored (>80% to 100%) and so their estimates were not reported.

Ethylbenzene AM estimates were similar cross time periods on the DD2. The Enterprise had high exposure on time period 1b and time period 3 were mostly censored.

Toluene AM estimates for some EGs on the Enterprise, DD3, and Q4000 had similar exposure trends over time as those for THC (Figure 5). The DD2, however, has higher AM estimates in TP2 and TP3 than in TP1. The 95th percentile of toluene estimates generally had similar trends as the AM (Figure 1 in the On-line Supplemental Materials).

Xylene AM estimates appeared to have similar trend as toluene (Figure 8) and 95th percentile trend is shown in Figure 2 in the On-line Supplement Materials.

The GSD estimates for toluene, ethylbenzene, and xylene were highly variable and generally higher than THC (Excel file in the On-line Supplement Materials)

Hexane estimates

No hexane estimates were report in TP1a because hexane was not analyzed until the beginning of TP1b. The exposures in TP1b were higher than in TP2 across all four rig ships and TP2 higher than TP3 on the DD3. The 95th percentile had similar exposure trends as the AM (Figure 3 in the On-line Supplement Materials)

Discussion

Our study presents exposure estimates for a comprehensive list of EGs that covered all tasks on the four rig vessels. The exposure trends across EGs, time periods, and vessels appeared to be influenced by a number of factors including the vessels' dates of arrival in the hot zone, function of vessel, job responsibilities of the crew, the physical and chemical properties of the chemicals, and the weathering effects that can also changed the composition of the oil.

The Enterprise and the DD3 were two of four rig vessels to arrive in the hot zone first. Therefore, there were measurements on these vessels only for time period 1a. The

Enterprise was positioned directly above the wellhead and it had equipment to process hydrocarbons. Its main responsibility was to process hydrocarbons and later (after June 3, 2010) to collect oil and gas. Thus, THC exposures on the Enterprise were generally higher than those on the DD2, DD3, and Q4000 for TP1b, TP2, and TP3. The DD3 started drilling the relief well on May 2, 2010 before the injection of dispersants started on May 15th, 2010. Generally there is not much oil or gas fumes associated with the drilling process. The crew on the DD3 were exposed to chemicals from the drilling process and background vapors associated with being in the hot zone. The DD3 exposures were similar to that on the Enterprise in TP1a. There were only very few measurements (< 5 sample badges) reported on the DD2 in TP1a because the DD2 came to the hot zone much later and started drilling the second relief well on May 16, 2010.

Up until the well was successfully top capped on July 15, 2010 , oil was gushing out of the well resulting high concentration of vapors. Despite other containment efforts such as dispersant injection below and application at the sea surface, oil skimming and burning that were simultaneously taking place, highest exposure was generally observed in this time period and then reduced only after the leaking was stopped in TP2 (top capping and bottom capping) because no more fresh oil had risen up to the surface. In TP3 when most volatile oil components had evaporated, THC exposures were generally lowest except in a few instances such as the Outside.Operations.ROVs on the DD2 and DD3 or Crew and Operations groups on the Q4000 where exposures were higher. The high exposure might be due to materials from the vessel decontamination processes. BTEX and hexane exposures generally were very low or not reported due to high percentage of undetected values.

Within each ship, exposures varied by EGs due to the variety of tasks. Except for the Q4000, three other ships monitor exposures inside the vessel. The Inside jobs mainly included cooks and utility operators, administrators, and those relating to living spaces. By policy, these inside areas were required to have carbon filter and we would expect the inside group to have lower exposures than those working outside. While the inside exposures appeared comparable on the three rig vessels, the inside group for TP1b looks similar to the Outside Group on the DD2 and DD3 (except the Enterprise where it was much lower than the outside).

The Outside group consisted of the Crew and Operation groups. The crew members performed day-to-day tasks to maintain the ships while the operation people, who were brought in from contractors, worked on oil-spill related tasks. The THC exposures on these EGs did not appear to have a consistent trend across ships. Even within each vessel, the relative ranking of exposures for EGs was challenging because they also differed by time period.

Although grouped in the same EGs, the exposure trends over time for toluene, ethylbenzene, xylene, and hexane differed from that of THC for each ship due to a combination of factors. These included the different evaporation rates of various oil-related components, and also oil weathering that affected the composition of THC. The BTEX chemicals generally had a higher level of censoring than THC, and much lower concentration than THC.

The GSD estimates for THC exposures were high compared to most typical occupational exposure groups. This is a reflection of the non-routine, time dependent nature of the spill response work. Other factors that contribute to the high variability/bias in the estimates are the small sample sizes and high degree of censoring. Toluene, ethylbenzene, xylene and hexane general had higher range of GSDs than THC because they have higher of percentage of censored observations in their data. In addition, the priors can also be a contributing factor. THC analysis used bounded priors which restrict the maximum GSD to be less than 12 whereas BTEX and hexane analysis used informative priors (normal and gamma) that were approximated from our resampling technique.

This study presents an approach to classifying exposure groups for an oil spill study and the Bayesian method for develop exposure estimates. As with most occupational epidemiological studies, potential bias from misclassification of EGs is also a limitation of the study as indicated by the large variability of the GSD and large credible intervals in many of our estimates. The misclassification bias could come from a number of sources including incomplete records of formal tasks or assignments on boards, non-routine nature of spill response work, undocumented tasks, recall bias during the completion of the questionnaire months after the event. We tried to minimize bias to the fullest extent as possible using the most current and systematic approach and based on a wealth of information from a variety of sources including site visits, public records, interviews with workers on the ships, questionnaires, years of combined professional judgments from research collaborators in industrial hygiene and exposure assessment. Despite the limitation, the study is an important contribution to the oil spill heath studies field which often lacks comprehensive quantitative exposure assessments. This study also offers a unique approach that could be used for future oil spill studies.

Conclusions

In summary, our study used a Bayesian method that can analyze data with detection limits to develop estimates of occupational exposures to THC, BTEX chemicals, and hexane on the four rig vessels in the hot zone. THC measurements were least censored compared to other chemicals evaluated. THC exposure trends over time varied by time period, ships, and EGs. Generally highest exposures were observed in TP1b when the oil was continuously leaking and rise to the surface until the pipe was successfully capped. Exposures gradually decreased over time in selected EGs except a few that might be involved in the decontamination effort. BTEX chemicals and hexane exposures were substantially lower than THC. The variability of the EGs for the GuLF STUDY were generally high, reflecting the non-routine, time-dependent nature of any spill response efforts as well as the challenges of retrospectively constructing exposures for oil spill study.

References

Amat-Bronnert A, Castegnaro M, Pfohl-Leszkowicz A. (2007) Genotoxic activity and induction of biotransformation enzymes in two human cell lines after treatment by Erika fuel extract. Environ. Toxicol. Phar- macol. 23: 89–95

Aguilera F, Méndez J, Pásaro E, Laffon B. (2010). Review on the effects of exposure to spilled oils on human health. *Journal of applied toxicology : JAT*, *30*(4), 291–301.

Campbell D, Cox D, Crum J, Foster K, Christie P. (1993) Initial effects ofthe grounding of the tanker Braer on health in Shetland. *British Medical Journal*, *307*, 1251–1255.

Campbell, D., Cox, D., Crum, J., Foster, K., Riley, A., Manfredini, R., Gallerani, M., et al. (1994). Later effects of grounding of tanker Braer on health in Shedland. *British Medical Journal*, *309*, 773–774.

Carrasco, J. M., Lope, V., Pérez-Gómez, B., Aragonés, N., Suárez, B., López-Abente, G., Rodríguez-Artalejo, F., et al. (2006). Association between health information, use of protective devices and occurrence of acute health problems in the Prestige oil spill clean-up in Asturias and Cantabria (Spain): a cross-sectional study. *BMC Public Health*, *6*

Carrasco, J. M., Pérez-Gómez, B., García-Mendizábal, M. J., Lope, V., Aragonés, N., Forjaz, M. J., Guallar-Castillón, P., et al. (2007). Health-related quality of life and mental health in the medium-term aftermath of the Prestige oil spill in Galiza (Spain): a cross-sectional study. *BMC public health*, *7*, 245.

Carlin, B.P. and Louis, T.A. (2009). Bayesian Methods for Data Analysis, 3rd ed. Boca Raton, FL: Chapman and Hall/CRC Press.

Chaty S, Rodius F, Lanhers M-C, Burnel D, Vasseur P. 2008. Induction of CYP1A1 in rat liver after ingestion of mussels contaminated by Erika fuel oils. Arch. Toxicol. 82: 75–80.

Cleveland, C. (2013). Deepwater Horizon oil spill . Retrieved from http://www.eoearth.org/view/article/161185

Crum, J. E. (1993). Peak expiratory flow rate in schoolchildren living close to Braer oil spill. *British Medical Journal*, *307*, 23–24.

Dor, F., Bonnard, R., Gourier-Fréry, C., Cicolella, A., Dujardin, R., & Zmirou, D. (2003). Health risk assessment after decontamination of the beaches polluted by the wrecked ERIKA tanker. *Risk analysis : an official publication of the Society for Risk Analysis*, *23*(6), 1199–208.

Gill D, Picou J. (1998) Technological disaster and chronic community stress. Soc. Natur. Resour. 11: 795–815

Groth C, Supdipto B, Huynh T, et al., (2014) Correlations of rig ship measurements (in press).

Huynh T, Ramachandran G, Banerjee S, Monterior J, Stenzel M, Sandler D, Kwok R, Engel L, Blair A, Stewart PA "Comparison of methods for analyzing left-censored data to estimate workers exposures for the GuLF STUDY." Accepted with revision at the Annals of Occupational Hygiene.

Huynh T, Quick H, Ramachandran G, Banerjee S, Monteiro J, Groth C, Stenzel M, Sandler D, Kwok R, Engel L, Blair A, Stewart PA. "A comparison of the β-substitution method and a Bayesian approach for handling left-censored data" To be submitted to the Annals of Occupational Hygiene July.

Hill AB. 1965.The evaluation of disease: association or causation? Proc R Soc Med 1965 58:295-300.

Janjua, N. Z., Kasi, P. M., Nawaz, H., Farooqui, S. Z., Khuwaja, U. B., Najam-ul-Hassan, Jafri, S. N., et al. (2006). Acute health effects of the Tasman Spirit oil spill on residents of Karachi, Pakistan. *BMC public health*, *6*, 84.

Laffon, B., Fraga-Iriso, R., Pérez-Cadahía, B., & Méndez, J. (2006). Genotoxicity associated to exposure to Prestige oil during autopsies and cleaning of oil-contaminated birds. *Food and chemical toxicology : an international journal published for the British Industrial Biological Research Association*, *44*(10), 1714–23.

Lemiere S, Cossu-Leguille, Bispo A, Jourdain MJ, Lanhers MC, Burnel D, Vasseur P. (2005) DNA damage measured by the single-cell gel elec- trophoresis (comet) assay in mammals fed with mussels contami- nated by the 'Erika' oil-spill. Mutat. Res. 581: 11–21.

Lyons, R. Temple, J. M., Evans, D., Fone, D. L., & Palmer, S. R. (1999). Acute health effects of the Sea Empress oil spill. *Journal of epidemiology and community health*, *53*(5), 306–10.

Meo, S. A., Al-Drees, A. M., Rasheed, S., Meo, I. M., Al-Saadi, M. M., Ghani, H. a, & Alkandari, J. R. (2009). Health complaints among subjects involved in oil cleanup operations during oil spillage from a Greek tanker "Tasman Spirit". *International journal of occupational medicine and environmental health*, *22*(2), 143–8.

Morita, a, Kusaka, Y., Deguchi, Y., Moriuchi, a, Nakanaga, Y., Iki, M., Miyazaki, S., et al. (1999). Acute health problems among the people engaged in the cleanup of the Nakhodka oil spill. *Environmental research*, *81*(3), 185–94.

National Institute for Occupational Safety and Health (NIOSH) (2010). *Health Hazard Evaluation of Deepwater Horizon Response Workers*. Retrieved from http://www.cdc.gov/niosh/h.he/reports/pdfs/2010-0115-0129-3138.pdf

National Institute for Occupational Safety and Health. (2010). Reducing Occupational Exposures while Working with Dispersants During the Deepwater Horizon Response http://www.cdc.gov/niosh/topics/oilspillresponse/dispersants.html (accessed January 3, 2013)

National Institute for Occupational Safety and Health. (2011) NIOSH Deepwater Horizon Roster Summary Report. Retrieved from http://www.cdc.gov/niosh/docs/2011-175/pdfs/2011-175.pdf

Palinkas AL, Russell J, Downs AM, Petterson, J. (1992). Ethnic differences in stress, coping, and depressive symptoms after the Exxon Valdez oil spill. *Journal of Nervous and Mental Disease*, *180*(5), 287–295.

Palinkas, L., Petterson, J., Russell, J., & Downs, M. (1993). Community patterns of psychiatric disorders after the Exxon Valdez oil spill. *American Journal of Psychiatry*, *150*(10), 1517–1523.

Pérez-Cadahía, B., Laffon, B., Pásaro, E., & Méndez, J. (2006). Genetic damage induced by accidental environmental pollutants. *TheScientificWorldJournal*, *6*, 1221–37.

Pérez-Cadahía, B., Lafuente, A., Cabaleiro, T., Pásaro, E., Méndez, J., & Laffon, B. (2007). Initial study on the effects of Prestige oil on human health. *Environment international*, *33*(2), 176–85.

Pérez-Cadahía B, Méndez J, Pásaro E, Lafuente A, Cabaleiro T, Laffon B. 2008a. Biomonitoring of human exposure to Prestige oil: Effects on DNA and endocrine parameters. Environ. Health Insights 2: 83–92.

Pérez-Cadahía B, Laffon B, Porta M, Lafuente A, Cabaleiro T, López T, Caride A, Pumarega J, Romero A, Pásaro E,Méndez J. 2008b. Relationship between blood concentrations of heavy metals and cytogenetic and endocrine parameters among subjects involved in cleaning coastal areas affected by the 'Prestige' tanker oil spill. Chemosphere 71: 447–455.

Pérez-Cadahía B, Laffon B, Valdiglesias V, Pásaro E,Méndez J. 2008c. Cyto- genetic effects inducedby Prestige oil onhumanpopulations:The role of polymorphisms in genes involved in metabolism and DNA repair. Mutat. Res. 653: 117–123.

Martyn Plummer (2003). JAGS: A Program for Analysis of Bayesian Graphical Models Using Gibbs Sampling, Proceedings of the 3rd International Workshop on Distributed Statistical Computing (DSC 2003), March 20–22, Vienna, Austria. ISSN 1609-395X.

Rappaport, S.M. (1991). Assessment of long-term exposures to toxic substances in air – Review. *Ann of Occup Hyg*, 35:61-121.

R Development Core Team. (2011) R: a language and environment for statistical computing. R Foundation for Statistical Computing: Vienna. http://www.R-project.org.

Sabucedo, J. M., Arce, C., Ferraces, M. J., & Merino, H. (2009). Psychological impact of the Prestige catastrophe. *International Journal of Clinical and Health Psychology*, *9*, 105–116.

Sandler DP, Kwok RK, Engel LS, Parks C, London SJ, Miller AK, Blair A, Stenzel M, Stewart PA. NIEHS GuLF STUDY: GuLF Long-Term Follow-Up Study. Retrieved from http://www.niehs.nih.gov/research/atniehs/labs/epi/studies/gulfstudy/publications/index.cfm

Seixas, N., Robins, T., Moulton, L. The use of geometric mean and arithmetic mean exposures in occupational epidemiology. *Am J Ind Med* 1988: 14: 465–77.

Suárez, B., Lope, V., Pérez-Gómez, B., Aragonés, N., Rodríguez-Artalejo, F., Marqués, F., Guzmán, a, et al. (2005). Acute health problems among subjects involved in the cleanup operation following the Prestige oil spill in Asturias and Cantabria (Spain). *Environmental research*, *99*(3), 413–24.

Stenzel M. (2014) Development of exposure groups (in press).

Stewart PA, Ramachandran G, Sudipto B, et al., The NIEHS GuLF STUDY: Overview of the assessment process for estimating exposures to volatile oil-related substances for subjects on the water (in press)

U.S. Coast Guard (USCG). (2011). *On Scene Coordinator Report Deepwater Horizon Oil Spill*. Retrieved from http://www.uscg.mil/foia/docs/dwh/fosc_dwh_report.pdf

U.S. Environmental Protection Agency. *Integrated Risk Information System (IRIS) on Xylene*. National Center for Environmental Assessment, Office of Research and Development, Washington, DC. 2003.

U.S. Environmental Protection Agency. *Integrated Risk Information System (IRIS) on Benzene*. National Center for Environmental Assessment, Office of Research and Development, Washington, DC. 2003.

U.S. Environmental Protection Agency. *Integrated Risk Information System (IRIS) on Toluene*. National Center for Environmental Assessment, Office of Research and Development, Washington, DC. 2005.

U.S. Environmental Protection Agency (USEPA). (2010) . Questions and Answers on Dispersants: http://www.epa.gov/bpspill/dispersants-qanda.html#appl. (Accessed January 4, 2013)

Zock, J.-P., Rodríguez-Trigo, G., Pozo-Rodríguez, F., Barberà, J. a, Bouso, L., Torralba, Y., Antó, J. M., et al. (2007). Prolonged respiratory symptoms in clean-up workers of the prestige oil spill. *American journal of respiratory and critical care medicine*, *176*(6), 610–6.

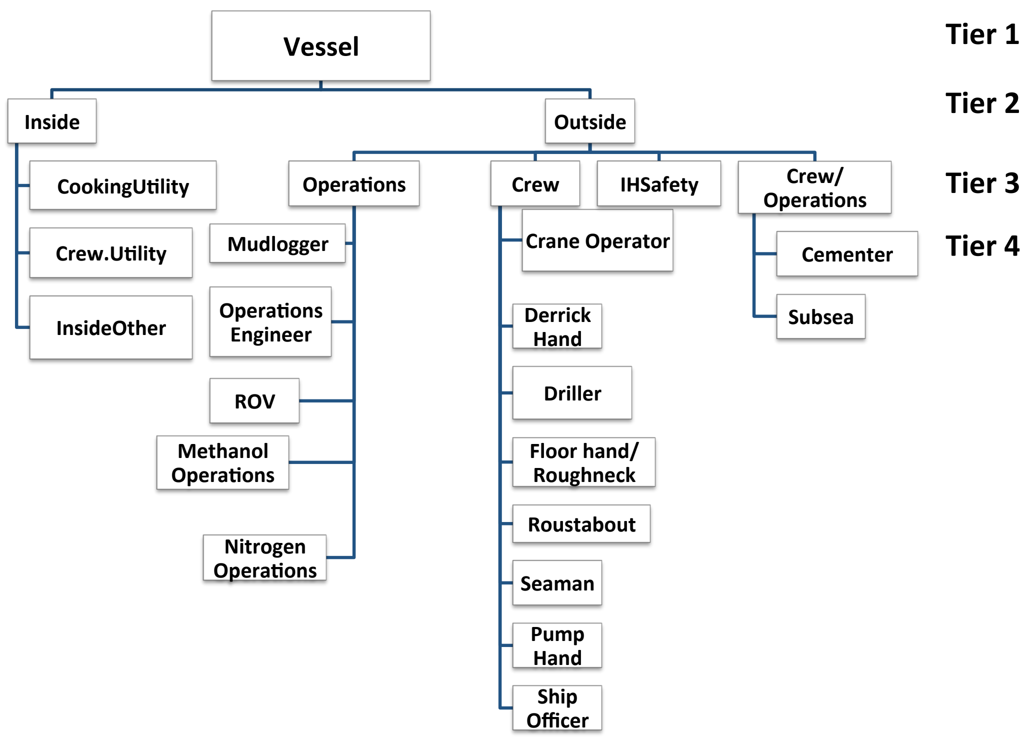

Figure 1: Hierarchy of selected exposure groups (EGs) on the Enterprise. Tiers 1 to 4 represent the levels of EG classification from broad to most specific.

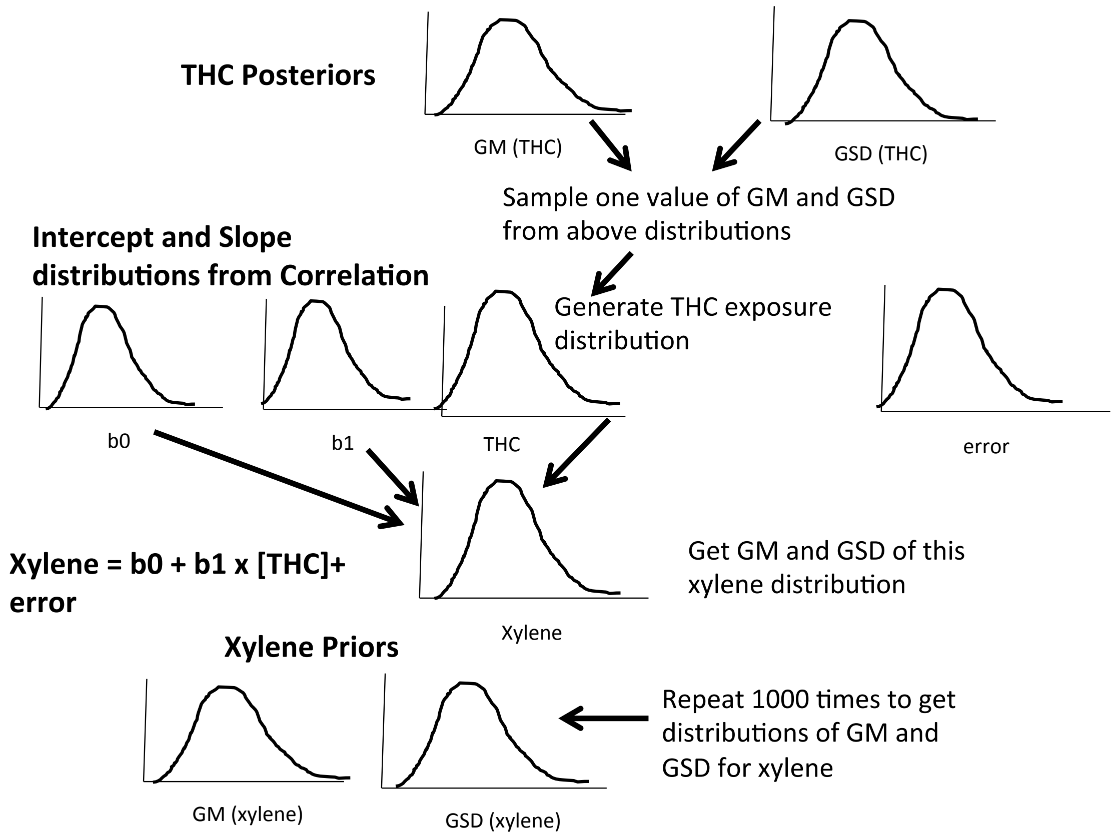

Figure 2: A schematic of the resampling strategy for developing priors for each EGs for the analysis of BTEX chemicals and hexane analysis (xylene is used as an example)

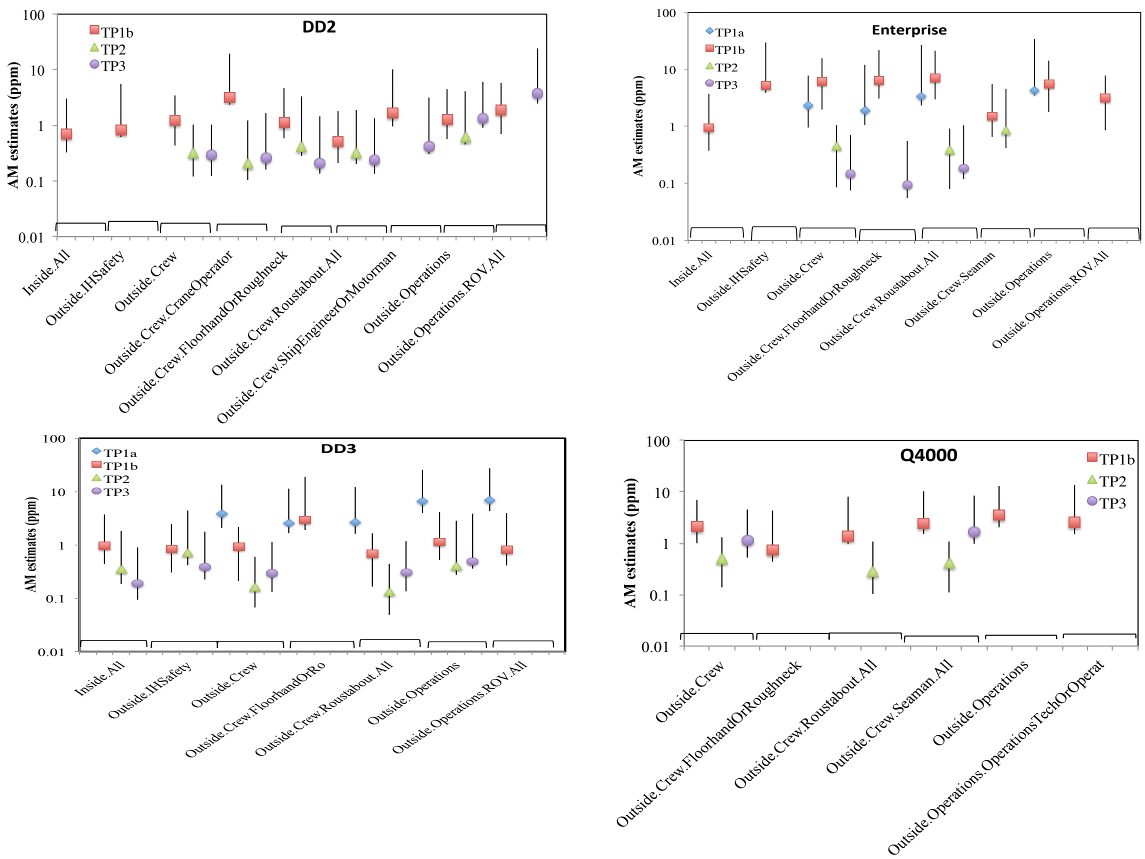

Figure 3: Estimates of the AM for THC exposures (in ppm) for selected EGs on the four rig vessels.

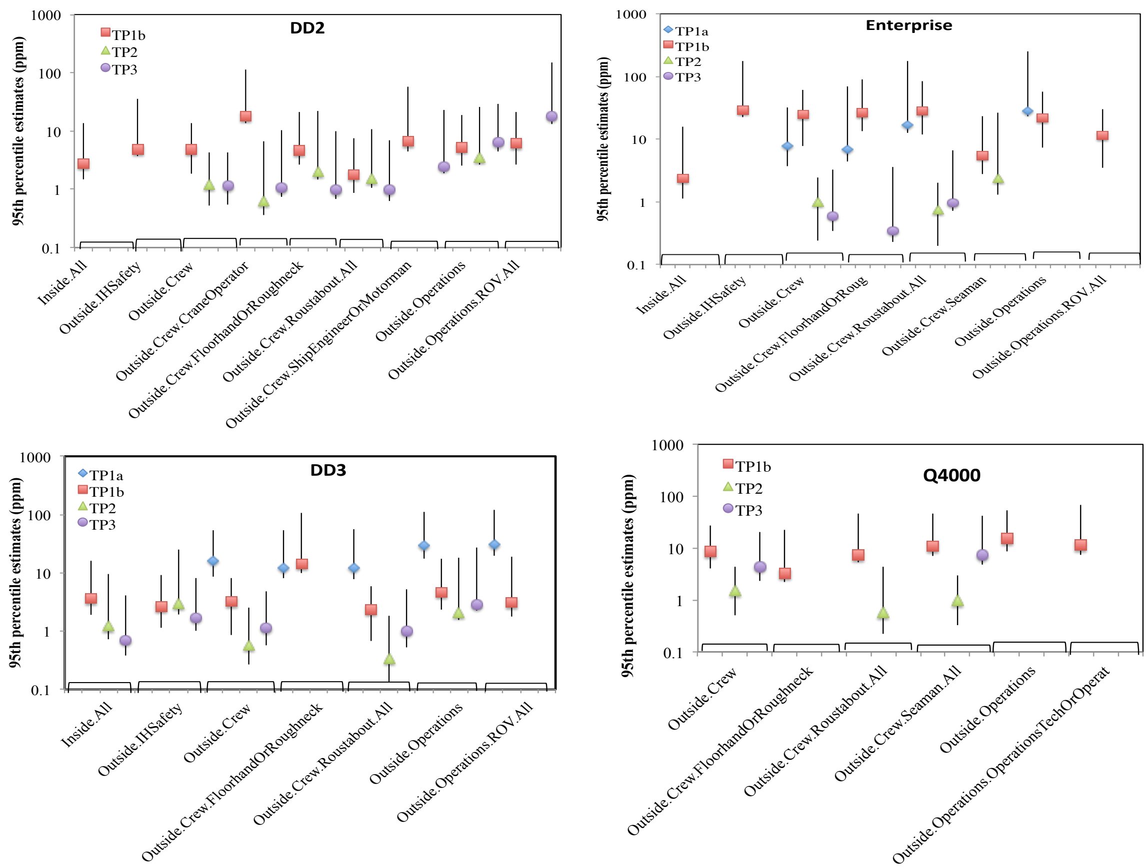

Figure 4: Estimates of the 95th percentile for THC exposures (in ppm) for selected EGs on the four rig vessels.

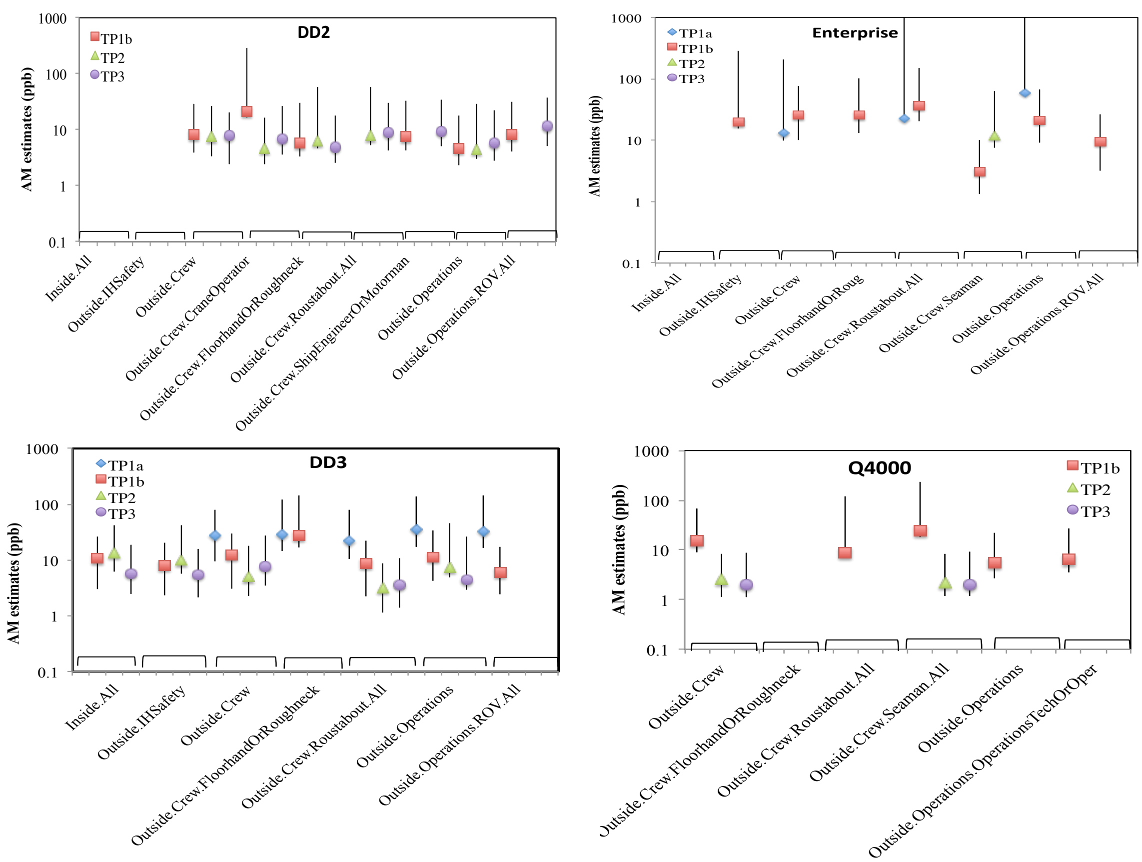

Figure 5: Estimates of the AM for ethylbenzene (in ppb) for selected EGs on the four rig vessels.

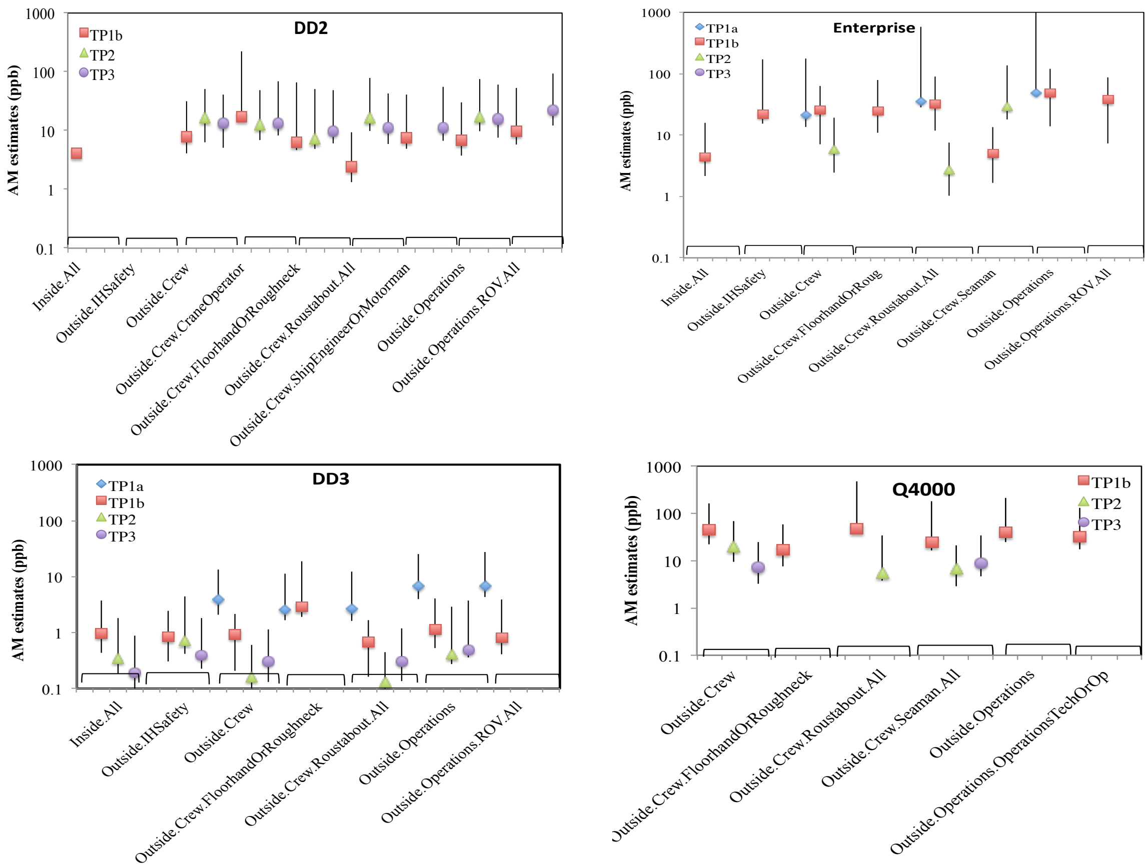

Figure 6: Estimates of the AM for toluene (in ppb) for selected EGs on the four rig vessels.

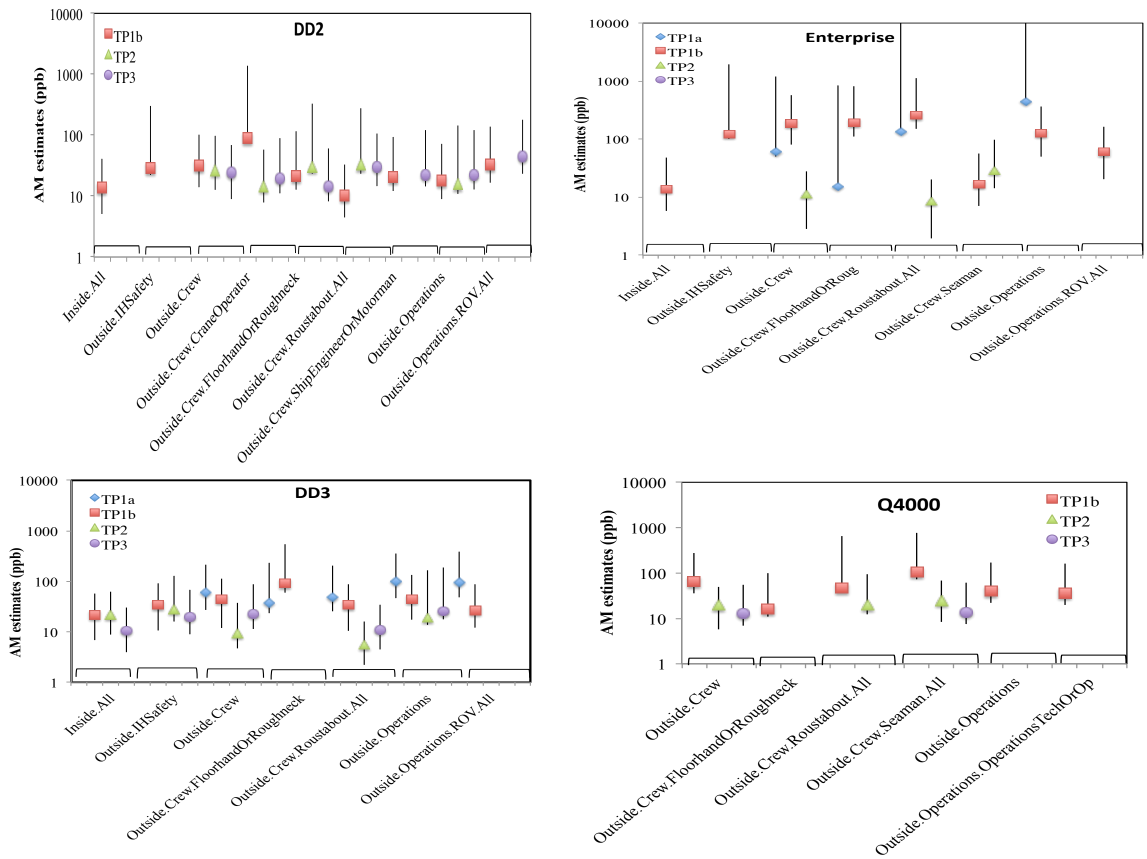

Figure 7: Estimates of the AM for xylene (in ppb) for selected EGs on the four rig vessels.

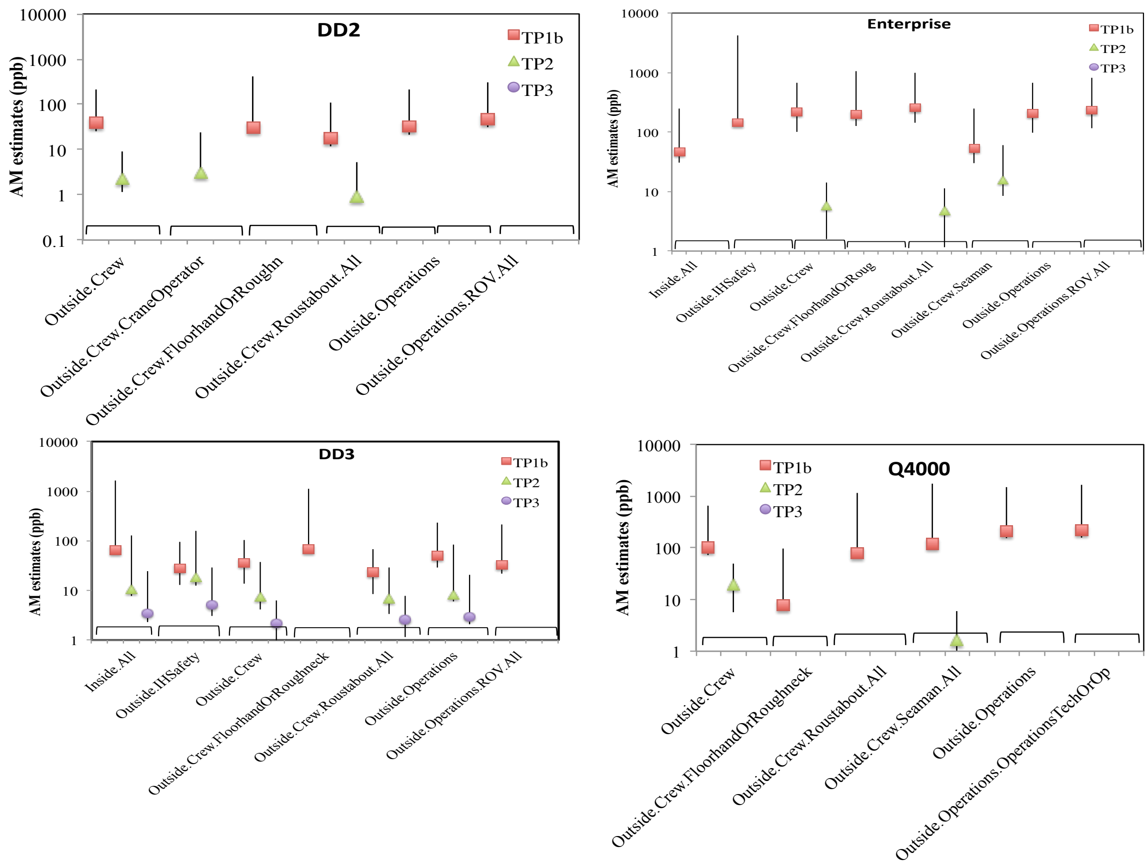

Figure 8: Estimates of the AM for hexane exposures (in ppb) for selected EGs on four rig vessels

Table 1: Number of measurements and percent censoring for each vessel and agent

Vessel	N	% < LOD					N	%<LOD
		THC	Benzene	Toluene	Ethylbenzene	Xylene	Hexane	Hexane
DD2	339	34	96	58	61	50	200	75
Enterprise	436	11	87	36	58	41	274	22
DD3	449	30	95	24	37	34	287	45
Q4000	207	20	92	68	63	39	197	67

Table 2a: Slopes and intercepts between THC and each of the BTEX chemicals and hexane the Inside group for four rig vessels.

		N	Analyte Percent Censored	THC Percent Censored	Median	2.5 Quantile	97.5 Quantile
Benzene							
	Intercept	75	92	33.3	-2.38	-16.12	0.71
	Slope				0.23	-0.44	1.74
Ethylbenzene							
	Intercept	75	54.33	33.33	0.16	-1.35	1.5
	Slope				0.2	-0.03	0.45
Toluene							
	Intercept	75	54.33	33.33	-0.26	-1.89	1.18
	Slope				0.3	0.05	0.57
Xylene							
	Intercept	75	54.33	33.33	-0.58	-1.67	0.4
	Slope				0.52	0.35	0.7
Hexane							
	Intercept	47	59.57	38.3	-7.6	-11.46	-4.86
	Slope				1.53	1.08	2.12

Table 2b: Slopes and intercepts between THC and ethylbenzene for the Outside group on the Enterprise.

Time		N	**Analyte Percent Censored**	**THC Percent Censored**	**Median**	**2.5 Quantile**	**97.5 Quantile**
Time Period 1A	Intercept	22	68.18	4.555	-15.57	-25.02	-8.76
	Slope				2.16	1.36	3.25
Time Period 1B	Intercept	345	51.88	7.54	-6.95	-7.61	-6.33
	Slope				1.12	1.04	1.2
Time Period 2	Intercept	31	83.87	3.23	-10.05	-15.98	-5.65
	Slope				1.68	0.98	2.61
Time Period 3	Intercept	24	100	75	-36.21	-60.28	0.63
	Slope				1.88	-2.67	4.15

Table 2c: Slopes and intercepts between THC and toluene for the Outside group on the Enterprise.

Time		N	Analyte Percent Censored	THC Percent Censored	Median	2.5 Quantile	97.5 Quantile
Time Period							
1A	Intercept	22	54.55	4.555	-10.08	-14.54	-5.81
	Slope				1.59	1.04	2.15
Time Period							
1B	Intercept	345	28.7	7.54	-3.4	-3.93	-2.85
	Slope				0.78	0.71	0.85
Time Period 2	Intercept	31	51.61	3.23	-7.8	-13.12	-3.39
	Slope				1.47	0.73	2.33
Time Period 3	Intercept	24	87.5	75	-11.87	-19.49	-5.29
	Slope				2.23	1.02	3.58

Table 2d: Slopes and intercepts between THC and xylene for the Outside group on the Enterprise.

Time		N	Analyte Percent Censored	THC Percent Censored	Median	2.5 Quantile	97.5 Quantile
Time Period 1A	Intercept	22	68.18	4.555	-17	-27.56	-11.36
	Slope				2.51	1.83	3.73
Time Period 1B	Intercept	345	37.1	7.54	-5.5	-5.97	-4.99
	Slope				1.15	1.09	1.21
Time Period 2	Intercept	31	16.13	3.23	-2.35	-5.15	0.47
	Slope				0.75	0.28	1.21
Time Period 3	Intercept	24	95.83	75	-21.09	-36.96	-5.18
	Slope				3.42	0.92	5.82

Table 2e: Slopes and intercepts between THC and hexane for the Outside group on the Enterprise.

Time		N	Analyte Percent Censored	THC Percent Censored	Median	2.5 Quantile	97.5 Quantile
Time Period							
1B	Intercept	210	12.38	8.1	-5.46	-6.05	-4.86
	Slope				1.26	1.17	1.34
Time Period 2	Intercept	31	25.81	3.23	-3.93	-7.42	-0.39
	Slope				0.89	0.3	1.47
Time Period 3	Intercept	24	100	75	-44.88	-81.34	-29.75
	Slope				3.49	-1.72	11.28

Chapter IV

Conclusions, Implications and Significance of the Research Work, and Future Directions

Overall conclusions

The goal of this research was to develop estimates of occupational inhalation exposures on the four rig vessels to support the GuLF STUDY that investigates the potential adverse health effects associated with oil spill response and clean-up work. The study analyzed a large number personal measurements that contained a high percentage of values below the detection limits. Computer simulation studies were conducted to evaluate classical (or 'frequentist') and Bayesian methods for handling data with detection limits and to identify the most suitable method for analyzing the GuLF STUDY exposure data. The Bayesian method was selected and used to estimate exposures via the inhalation route for workers on the four rig vessels (Enterprise, Q4000, DD2, and DD3) in the hot zone. The following conclusions are drawn from this research:

- The first simulation study (Chapter II) compared classical statistical methods for analyzing left-censored data and found that the β-substitution method performed better than the Maximum Likelihood (ML) method and the Kaplan-Meier (K-M) method under most conditions of our study (including low N, high censoring, high variability, multiple LODs, and mixed distributions). The β-substitution method's accuracy and precision decreased at small and moderate sample sizes ($N \leq 10$), but was still the best of the three methods. Estimates for sample size <5 are likely to be unreliable. The ML method generally did well with large samples sizes and lognormal distributions. The use of the minimum variance unbiased estimator equation to compute the AM from the ML estimates of the GM and the GSD reduced the ML's transformation bias for small to moderate sample sizes. The K-M method was generally less biased at censoring levels $<50\%$. Though very robust, a major limitation of the β-substitution method is the lack of a confidence interval around the mean, whereas confidence intervals can be computed for the ML and the K-M methods. This study suggests that none of the statistical methods evaluated in this paper are recommended for datasets that have a combination of small to moderate sample sizes, high level of censoring, or high variability.

- The second simulation study (Chapter III) compared the β-substitution method and a Bayesian method for estimating the parameters (e.g., AM, GM, GSD, 95^{th} percentile) of exposure distribution from censored data. We have shown that both methods generally delivered accurate point estimates. The β-substitution method was generally less biased and was easier to implement but its measure of uncertainty was less reliable for the AM, and uncertainty could not be estimated for the GM, GSD, and 95^{th} percentile. The Bayesian method may be particularly useful if the practitioner has the computational resources and prior information, as the method generally provides accurate estimates and also provides full distributions for all parameters.
- The third study (Chapter IV) used a Bayesian method to estimate occupational inhalation exposures to total hydrocarbons (THC), benzene, toluene, ethylbenzene, xylene (BTEX chemicals), and hexane on the four rig vessels in the hot zone. THC measurements were least censored compared to other chemicals evaluated. THC exposure trends over time varied by time periods, ships, and exposure groups (EGs). Highest exposures were generally observed during time period before the well was successful top capped. Exposures gradually decreased over time after top capping in selected EGs except a few that might be involved in the decontamination effort. BTEX chemicals and hexane exposures were substantially lower than THC. The variability of the EGs for the GuLF STUDY were generally higher than most occupational exposures seen in typically manufacturing settings, reflecting the non-routine, time-dependent nature of the spill response efforts as well as the challenges of retrospectively constructing exposures for oil spill study.

Implications and Significance of the Research Work

The first simulation study evaluated classical censored data analysis methods over a wider range of conditions/scenarios than previously published studies. It covered smaller sample sizes down to 5, larger GSD (variability) up to 5, and censoring level up to 90%. These scenarios are prevalent in the GuLF STUDY so the findings are directly applicable to the GuLF STUDY and also any studies that have these extreme scenarios. In addition, previous studies compared different sets of methods (e.g., β-substitution method vs. ML method, ML method vs. K-M method, ML vs. other methods) and published their findings resulting some inconsistent recommendations in the literature. This simulation study agglomerated the three promising methods - the β-substitution method, K-M method, and ML method – and compared their performance. While most of the conclusions generally aligned with many findings in the literature, the study also added some new knowledge to the field. For example, the β-substitution method was shown to perform better than the K-M method. It also elucidated the source of inconsistent recommendations in the scientific literature with regard to the ML method which was due to the different equations that was used to estimate the arithmetic mean. And lastly, the study identified the limitations of each of the methods evaluated so that practitioners are better informed when selecting the methods suitable for their needs.

The use of the Bayesian method for analyzing exposure data for risk management has been gaining acceptance and popularity in the field of occupational hygiene. This research extended the applications of Bayesian statistics to analyzing data with detection limits. Since the Bayesian censored data method was found to perform as well as the β-substitution method, the hope is that this finding would encourage more practitioners to explore and incorporate Bayesian statistics in their exposure assessment and risk management. Bayesian statistics offer many attractive features such as the ability to incorporate prior information with the data and providing full distributions for all parameters. These features can be particularly useful for making decisions in risk management.

The methodology and findings from the occupational exposure assessment study for the rig vessels is a unique contribution to the exposure assessment field. As described in the introduction, quantitative exposure studies for oil spill remediation work are few in numbers and none of those studies had the wealth of information and the scientific rigor in the exposure assessment strategy. Since the Bayesian method was used to analyze the data, the study took advantage of the Bayesian features and devoted a significant amount of effort to develop appropriate priors for the analysis. The approach to developing priors based on the correlation of the various oil components presented here may also be used for other studies where such correlation exists.

Lastly, the R-functions for censored data analysis methods evaluated in this dissertation are freely available for practitioners and researchers.

Study limitations

In both simulation studies, for a given dataset (especially a small dataset), the true underlying distribution is often unknown, and the percent censoring from the data does not necessarily correspond to the actual percentile in the true distribution used in our simulations. Hence, the true bias will probably differ from the composite bias obtained from these simulations and thus, the bias reported here cannot be assumed to be the bias of any particular dataset even though it meets the conditions we evaluated. However, these simulations serve as good evaluation tools to compare the methods when subjected to the same conditions. Another limitation is that our simulated conditions that are generally found in typical, routine workplace operations (e.g., GSD for a similarly exposed group is generally less than 3 or 4); the operations in the GuLF STUDY were often non-routine and highly variable, and therefore may not have been covered by our simulation conditions.

In the occupational exposure assessment on the four rig vessels (Chapter IV), the potential for misclassification of workers into exposure groups might be present and this could come from a number of sources including incomplete records of formal tasks or assignments on boards, non-routine nature of spill response work, undocumented tasks, recall bias during the completion of the questionnaire months after the event. However, we are confident that that bias from misclassification was minimized to the fullest extent because our exposure grouping was conducted using the most current and systematic approach and based on a wealth of information from a variety of sources including site visits, public records, interviews with workers on the ships, questionnaires, years of combined professional judgments from research collaborators in industrial hygiene and exposure assessment.

Future directions

Thousands of personal measurements were also collected on large marine vessels that operated the controlled vehicles (ROVs), small and large commercial and private boats that assisted in the skimming and oil burning operation on the water as well as clean-up and decontamination activities on near shore and on land. Similar Bayesian approach can be used to analyze those exposure data.

Bibliography

Aitchison J and Brown, JAC (1969). The Lognormal Distribution, Cambridge: Cambridge University Press.

Aguilera F, Méndez J, Pásaro E, Laffon B. (2010). Review on the effects of exposure to spilled oils on human health. *Journal of applied toxicology : JAT*, *30*(4), 291–301.

Amat-Bronnert A, Castegnaro M, Pfohl-Leszkowicz A. (2007) Genotoxic activity and induction of biotransformation enzymes in two human cell lines after treatment by Erika fuel extract. Environ. Toxicol. Phar- macol. 23: 89–95

Antweiler RC, Taylor HC (2008) Evaluation of statistical treatments of left-censored environmental data using coincident uncensored data sets: I. Summary statistics. Env Sci Technol; 42: 3732–8

Beal, DJ (2010) A macro for calculating summary statistics on left-censored environmental data using the Kaplan-Meier method. Proceedings of the 18th Annual Conference of the Southeast SAS Users Group.

Busschaert P, Geeraerd AH, Uyttendaele M, and Van Impe JF. (2011). Hierarchical Bayesian analysis of censored microbiological contamination data for use in risk assessment and mitigation. Food microbiology, 28(4), 712–9.

Campbell D, Cox D, Crum J, Foster K, Christie P. (1993) Initial effects ofthe grounding of the tanker Braer on health in Shetland. *British Medical Journal*, *307*, 1251–1255.

Campbell, D., Cox, D., Crum, J., Foster, K., Riley, A., Manfredini, R., Gallerani, M., et al. (1994). Later effects of grounding of tanker Braer on health in Shedland. *British Medical Journal*, *309*, 773–774.

Carlin BP, and Louis TA. (2009). Bayesian Methods for Data Analysis, 3rd ed. Boca Raton, FL: Chapman and Hall/CRC Press.

Carrasco, J. M., Lope, V., Pérez-Gómez, B., Aragonés, N., Suárez, B., López-Abente, G., Rodríguez-Artalejo, F., et al. (2006). Association between health information, use of protective devices and occurrence of acute health problems in the Prestige oil spill clean-up in Asturias and Cantabria (Spain): a cross-sectional study. *BMC Public Health*, *6*

Carrasco, J. M., Pérez-Gómez, B., García-Mendizábal, M. J., Lope, V., Aragonés, N., Forjaz, M. J., Guallar-Castillón, P., et al. (2007). Health-related quality of life and mental health in the medium-term aftermath of the Prestige oil spill in Galiza (Spain): a cross-sectional study. *BMC public health*, *7*, 245.
Chaty S, Rodius F, Lanhers M-C, Burnel D, Vasseur P. 2008. Induction of CYP1A1 in rat liver after ingestion of mussels contaminated by Erika fuel oils. Arch. Toxicol. 82: 75–80.
Cleveland, C. (2013). Deepwater Horizon oil spill . Retrieved from http://www.eoearth.org/view/article/161185
Cohen, AC. (1959) Simplified estimators for normal distribution when samples are singly censored or truncated. Technometrics 1, 217-37.
Cohen, AC. (1961) Tables for maximum likelihood estimates: Singly truncated and singly censored samples. Technometrics 3, 535-541.
Cohn, TA (1988) Adjusted maximum likelihood estimation of the moments of log-normal populations from type I censored samples: U.S. Geological Survey Open-File Report 88-350, 34
Crum, J. E. (1993). Peak expiratory flow rate in schoolchildren living close to Braer oil spill. *British Medical Journal*, *307*, 23–24.
Dinse GE, Jusko TA , Ho LA, Annam K, Graubard BI, Hertz-Picciotto I, Miller WF, Gillespie BW, and Weinberg CR* American Journal of Epidemiology Vol. 179, No. 8 DOI: 10.1093/aje/kwu017
Dor, F., Bonnard, R., Gourier-Fréry, C., Cicolella, A., Dujardin, R., & Zmirou, D. (2003). Health risk assessment after decontamination of the beaches polluted by the wrecked ERIKA tanker. *Risk analysis : an official publication of the Society for Risk Analysis*, *23*(6), 1199–208.
Draxler, R.R., Hess G.D. (1998). An overview of the HYSPLIT_V modeling system for trajectories, dispersion, deposition,. Aust. Meteor. Mag, 47, 295-308.
European Food Safety Authority. (2010). Management of left-censored data in dietary exposure assessment of chemical substances. *EFSA Journal*, *8*(3), 1–96. Available at: www.efsa.europa.eu.

Finkelstein MM, and Verma DK. (2001) Exposure estimation in the presence of nondetectable values: another look. Am Ind Hyg Assoc J; 62:195-8.
Fisher RA. (1925) Theory of statistical estimation. Proceedings of the Cambridge Philosophical Society 22, 700-725.
Finney DJ. (1941) On the Distribution of a Variate Whose Logarithm Is Normally Distributed. Journal of the Royal Statistical Society, Supplement 7:155-1 61.
Gallacher, J., Bronstering, K., Palmer, S., Fone, D., & Lyons, R. (2007). Symptomatology attributable to psychological exposure to a chemical incident: a natural experiment. *Journal of epidemiology and community health*, *61*(6), 506–12.
Ganser GH, and Hewett P. (2010). An accurate substitution method for analyzing censored data. *Journal of occupational and environmental hygiene*, *7*(4), 233–44.
Gelfand, AE, Smith, AFM., and Lee, TM. (1992). Bayesian analysis of constrained parameter and truncated data problems using Gibbs sampling. Journal of the American Statistical Association, 87, 523--532.
Gillespie BW, Chen Q, Reichert H et al. (2010) Estimating population distributions when some data are below a limit of detection by using a reverse Kaplan-Meier estimator. Epidemiology; 21: S64–70.
Gill D, Picou J. (1998) Technological disaster and chronic community stress. Soc. Natur. Resour. 11: 795–815
Groth C, Supdipto B, Huynh T, et al., (2014) Correlations of rig ship measurements (in press).
Huynh T, Quick H, Ramachandran G, Banerjee S, Monteiro J, Groth C, Stenzel M, Sandler D, Kwok R, Engel L, Blair A, Stewart PA. “A comparison of the β-substitution method and a Bayesian approach for handling left-censored data” (Chapter III)
Helsel DR. (2005) Nondetects and data analysis. New York: John Wiley & Sons, Inc.
Helsel, D. (2010). Much ado about next to nothing: incorporating nondetects in science. *The Annals of occupational hygiene*, *54*(3), 257–62.

Hewett, P., & Ganser, G. H. (1997). Simple Procedures for Calculating Confidence Intervals around the Sample Mean and Exceedance Fraction Derived from Lognormally Distributed Data. *Applied Occupational and Environmental Hygiene*, *12*(2), 132–142.

Hewett, P., & Ganser, G. H. (2007). A comparison of several methods for analyzing censored data. *The Annals of occupational hygiene*, *51*(7), 611–32.

Hill AB. (1965) The evaluation of disease: association or causation? Proc R Soc Med 1965 58:295-300.

Hornung RW, and Reed LD. (1990) Estimation of average concentration in the presence of non-detectable values. *Appl. Occup. Envir. Hyg.* 5(1):46–51.

Huynh T, Ramachandran G, Banerjee S, Monteiro J, Stenzel M, Sandler DP, Kwok RK, Engel LS, Blair A, Stewart PA. “Comparison of methods for analyzing left-censored data to estimate worker’s exposures for the GuLF STUDY.'” To be submitted to the Annals of Occupational Hygiene (Chapter II)

Ignacio JS, Bullock WH. editors. (2006) A strategy for assessing and managing occupational exposures. 3rd ed. Fairfax, VA: AIHA Press.

Janjua, N. Z., Kasi, P. M., Nawaz, H., Farooqui, S. Z., Khuwaja, U. B., Najam-ul-Hassan, Jafri, S. N., et al. (2006). Acute health effects of the Tasman Spirit oil spill on residents of Karachi, Pakistan. *BMC public health*, *6*, 84.

Kaplan, E. L., & Meier, P. (1958). Nonparametric Estimation from Incomplete Observations. *Journal of the American Statistical Association*, *53*(282), 457.

Krishnamoorthy K, Mallick A, Mathew T. (2009) Model based imputation approach for data analysis in the presence of non-detectable values: Normal and Related Distributions. Ann Occup Hyg; 59: 249–68.

Kroll CN, Stedinger JR. (1996) Estimation of moments and quantiles using censored data. Water Resour Res; 32: 1005–12

Laffon, B., Fraga-Iriso, R., Pérez-Cadahía, B., & Méndez, J. (2006). Genotoxicity associated to exposure to Prestige oil during autopsies and cleaning of oil-contaminated birds. *Food and chemical toxicology : an international journal published for the British Industrial Biological Research Association*, *44*(10), 1714–23.

Leidel, NA, Busch KA, Lynch JR (1977) Occupational Exposure
Sampling Strategy Manual. National Institute for Occupational Safety and Health Pub. No. 77-173 (available from the National Technical Information Service, Pub. No. PB274792)

Lemiere S, Cossu-Leguille, Bispo A, Jourdain MJ, Lanhers MC, Burnel D, Vasseur P. (2005) DNA damage measured by the single-cell gel elec- trophoresis (comet) assay in mammals fed with mussels contami- nated by the 'Erika' oil-spill. Mutat. Res. 581: 11–21.

Lyons, R. Temple, J. M., Evans, D., Fone, D. L., & Palmer, S. R. (1999). Acute health effects of the Sea Empress oil spill. *Journal of epidemiology and community health*, *53*(5), 306–10.

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

Meo, S. A., Al-Drees, A. M., Rasheed, S., Meo, I. M., Al-Saadi, M. M., Ghani, H. a, & Alkandari, J. R. (2009). Health complaints among subjects involved in oil cleanup operations during oil spillage from a Greek tanker "Tasman Spirit". *International journal of occupational medicine and environmental health*, *22*(2), 143–8.

Morita, a, Kusaka, Y., Deguchi, Y., Moriuchi, a, Nakanaga, Y., Iki, M., Miyazaki, S., et al. (1999). Acute health problems among the people engaged in the cleanup of the Nakhodka oil spill. *Environmental research*, *81*(3), 185–94.

Mulhausen, J., and J. Damiano: Chapter 4, Establishing Similar Exposure Groups. In: A Strategy for Managing Occupational Exposures. J.S. Ignacio, W. H. Bullock, (Eds), Third Edition, AIHA Press, Fairfax, VA, p. 33-46 (2006).

National Institute for Occupational Safety and Health (NIOSH) (2010a). *Health Hazard Evaluation of Deepwater Horizon Response Workers*. Retrieved from http://www.cdc.gov/niosh/h.he/reports/pdfs/2010-0115-0129-3138.pdf

National Institute for Occupational Safety and Health. (2010b). Reducing Occupational Exposures while Working with Dispersants During the Deepwater Horizon Response. Retrieved from http://www.cdc.gov/niosh/topics/oilspillresponse/dispersants.html

National Institute for Occupational Safety and Health. (2011) NIOSH Deepwater Horizon Roster Summary Report. Retrieved from http://www.cdc.gov/niosh/docs/2011-175/pdfs/2011-175.pdf

Occupational Safety & Health Administration. (2011) Deepwater Horizon oil spill: OSHA's role in the response. Available at: http://www.osha.gov/oilspills/dwh_osha_response_0511a.pdf.

Palinkas AL, Russell J, Downs AM, Petterson, J. (1992). Ethnic differences in stress, coping, and depressive symptoms after the Exxon Valdez oil spill. *Journal of Nervous and Mental Disease*, *180*(5), 287–295.

Palinkas, L., Petterson, J., Russell, J., & Downs, M. (1993). Community patterns of psychiatric disorders after the Exxon Valdez oil spill. *American Journal of Psychiatry*, *150*(10), 1517–1523.

Palinkas LA, Petterson JS, Russell J, DownsMA (2004) Ethnic differences in symptomsof post-traumatic stress after the Exxon Valdez oil spill.PDM 19: 102–112.

Paulo, MJ, Van der Voet H, Jansen, MJW, Ter Braak, CJF, and Van Klaveren, JD . (2005). Risk assessment of dietary exposure to pesticides using a Bayesian method. *Pest management science*, *61*(8)

Pérez-Cadahía, B., Laffon, B., Pásaro, E., & Méndez, J. (2006). Genetic damage induced by accidental environmental pollutants. *TheScientificWorldJournal*, *6*, 1221–37.

Pérez-Cadahía, B., Lafuente, A., Cabaleiro, T., Pásaro, E., Méndez, J., & Laffon, B. (2007). Initial study on the effects of Prestige oil on human health. *Environment international*, *33*(2), 176–85.

Pérez-Cadahía B, Méndez J, Pásaro E, Lafuente A, Cabaleiro T, Laffon B. 2008a. Biomonitoring of human exposure to Prestige oil: Effects on DNA and endocrine parameters. Environ. Health Insights 2: 83–92.

Pérez-Cadahía B, Laffon B, Porta M, Lafuente A, Cabaleiro T, López T, Caride A, Pumarega J, Romero A, Pásaro E,Méndez J. 2008b. Relationship between blood concentrations of heavy metals and cytogenetic and endocrine parameters among subjects involved in cleaning coastal areas affected by the 'Prestige' tanker oil spill. Chemosphere 71: 447–455.

Pérez-Cadahía B, Laffon B, Valdiglesias V, Pásaro E,Méndez J. 2008c. Cyto- genetic effects inducedby Prestige oil onhumanpopulations:The role of polymorphisms in genes involved in metabolism and DNA repair. Mutat. Res. 653: 117–123.

Martyn Plummer (2003). JAGS: A Program for Analysis of Bayesian Graphical Models Using Gibbs Sampling, Proceedings of the 3rd International Workshop on Distributed Statistical Computing (DSC 2003), March 20–22, Vienna, Austria. ISSN 1609-395X.

R Development Core Team. (2011) R: a language and environment for statistical computing. R Foundation for Statistical Computing: Vienna. http://www.R-project.org.

Ramachandran G. (2005) Occupational exposure assessment for air contaminants. Boca Raton, FL: CRC Press. ISBN: 1-56670-609-2.

Rappaport, S.M. (1991). Assessment of long-term exposures to toxic substances in air – Review. *Ann of Occup Hyg*, 35:61-121.

Robert C, and Casella G. (2005). Monte Carlo Statistical Methods, New York, Springer

Selvin S, Rappaport SM (1989). A Note on the Estimation of the Mean Value from a Lognormal Distribution. Am. Ind. Hyg. ASSOCJ. 50~627-630

Sabucedo, J. M., Arce, C., Ferraces, M. J., & Merino, H. (2009). Psychological impact of the Prestige catastrophe. *International Journal of Clinical and Health Psychology*, *9*, 105–116.

Sandler DP, Kwok RK, Engel LS, Parks C, London SJ, Miller AK, Blair A, Stenzel M, Suárez, B., Lope, V., Pérez-Gómez, B., Aragonés, N., Rodríguez-Artalejo, F., Marqués, F., Guzmán, a, et al. (2005). Acute health problems among subjects involved in the cleanup operation following the Prestige oil spill in Asturias and Cantabria (Spain). *Environmental research*, *99*(3), 413–24.

Seixas, N., Robins, T., Moulton, L. The use of geometric mean and arithmetic mean exposures in occupational epidemiology. *Am J Ind Med* 1988: 14: 465–77.

Shumway RH, Azari RS, and Kayhanian M. (2002). Statistical approaches to estimating mean water quality concentrations with detection limits: Environmental Science and Technology 36, 3345-3353

Succop PA, Clark S, Chen M et al. (2004) Imputation of data values that are less than a detection limit. J Occup Environ Health; 1: 436–41.

Singh A, Maichle R, Lee SE. (2006) On the computation of a 95%upper confidence limit of the unknown population mean based upon data sets with below detection limit observations. Washington, DC: U.S. Environmental Protection Agency EPA/600/R-06/022

SkyTruth. 2010. BP / Gulf Oil Spill - Cumulative Oil Slick Footprints. Retrieved from http://blog.skytruth.org/2010/09/bp-gulf-oil-spill-cumulative-oil-slick.html

Stenzel M. (2014) Development of exposure groups (in press).

Stewart PA, Ramachandran G, Sudipto B, et al. (2014) The NIEHS GuLF STUDY: Overview of the assessment process for estimating exposures to volatile oil-related substances for subjects on the water (in press).

US Coast Guard (USCG). (2011). *On Scene Coordinator Report Deepwater Horizon Oil Spill*. Retrieved from http://www.uscg.mil/foia/docs/dwh/fosc_dwh_report.pdf

U.S. Environmental Protection Agency. *Integrated Risk Information System (IRIS) on Xylene*. National Center for Environmental Assessment, Office of Research and Development, Washington, DC. 2003.

U.S. Environmental Protection Agency. *Integrated Risk Information System (IRIS) on Benzene*. National Center for Environmental Assessment, Office of Research and Development, Washington, DC. 2003.

U.S. Environmental Protection Agency. *Integrated Risk Information System (IRIS) on Toluene*. National Center for Environmental Assessment, Office of Research and Development, Washington, DC. 2005.

U.S. Environmental Protection Agency (USEPA). (2010) . Questions and Answers on Dispersants. Retrieved from http://www.epa.gov/bpspill/dispersants-qanda.html#appl.

U.S. Environmental Protection Agency (USEPA). (2007) ProUCL Version 4.0 Technical Guide. EPA/600/R-07/0

U.S. Coast Guards. On Scene Coordinator Report on Deepwater Horizon Oil Spill(Report). September 2011. Retrieved from http://www.uscg.mil/foia/docs/dwh/fosc_dwh_report.pdf

Zock, J.-P., Rodríguez-Trigo, G., Pozo-Rodríguez, F., Barberà, J. a, Bouso, L., Torralba, Y., Antó, J. M., et al. (2007). Prolonged respiratory symptoms in clean-up workers of the prestige oil spill. *American journal of respiratory and critical care medicine*, *176*(6), 610–6.

Appendix

A. The β-substitution method

1. Derivation of the β-factor for GM for a single LOD (Ganser and Hewett, 2010)

The following derivation of the *β-factor* for GM is provided by the authors

Calculation for β_{GM} -the factor

$x_i =$ lognormal distributed with LOD L

$n =$ sample size

$k =$ number of measurements below LOD

$y_i = \ln x_i$

The y_i that correspond to observed values of x_i has distribution $\dfrac{e^{\frac{-(y-\mu)^2}{2\sigma^2}}}{\int_{\ln L}^{\infty} e^{\frac{-(x-\mu)^2}{2\sigma^2}}dx}$

Note: $\overline{y} = \dfrac{1}{n-k}\sum_{i=1}^{n-k} y_i \Rightarrow E(\overline{y}) = \mu + \sigma f(z), z = \dfrac{\ln L - \mu}{\sigma}, f(z) = \dfrac{e^{\frac{-z^2}{2}}}{\int_z^{\infty} e^{\frac{-x^2}{2}}dx}$

Also $E(e^{ry}) = e^{\mu+\frac{\sigma^2 r^2}{2}} \left(\dfrac{\frac{1}{\sqrt{2\pi}} \int_{\frac{\log L-(\mu+\sigma^2 r)}{\sigma}}^{\infty} e^{-v^2/2}dv}{\frac{1}{\sqrt{2\pi}} \int_{\frac{\log L-\mu}{\sigma}}^{\infty} e^{-v^2/2}dv} \right) = e^{\mu+\frac{\sigma^2 r^2}{2}} X$

So estimate for geometric mean is

$e^{\hat{\mu}} = (x_1 x_2 \cdots x_{n-k})^{\frac{1}{n}} (\beta L)^{\frac{k}{n}} = (e^{\frac{1}{n}\sum_1^{n-k} y_i})(\beta L)^{\frac{k}{n}}$ where β is the factor multiplying LOD to estimate geometric mean.

So $E(e^{\hat{\mu}}) = E(e^{\frac{1}{n}\sum_1^{n-k} y_i})(\beta L)^{\frac{k}{n}} = (e^{\frac{\mu}{n}+\frac{\sigma^2}{2n^2}} \cdot X)^{n-k} (\beta L)^{\frac{k}{n}}$

$$= e^{(\frac{\mu(n-k)}{n} + \frac{\sigma^2(n-k)}{2n^2} + (n-k)\log X + \frac{k}{n}\log(\beta L))} = e^{\mu} e^{(\frac{-\mu k}{n} + \frac{\sigma^2(n-k)}{2n^2} + (n-k)\log X + \frac{k}{n}\log(\beta) + \frac{k}{n}\log L)}$$

$$= e^{\mu}$$

If $\frac{-\mu k}{n}+\frac{\sigma^2(n-k)}{2n^2}+(n-k)\log X+\frac{k}{n}\log(\beta)+\frac{k}{n}\log L=0 \Rightarrow \log\beta=-\sigma z-\frac{\sigma^2(n-k)}{2nk}-\frac{n(n-k)}{k}\log X$

So z is estimated from $z=\frac{\log L-\mu}{\sigma}$ and $\frac{k}{n}\cong\frac{1}{\sqrt{2\pi}}\int_{-\infty}^{z}e^{-v^2/2}dv$

σ is estimated from $\sigma=\frac{\log L-\mu}{z}$ and $\mu=E(\overline{y})-\sigma f(z)\Rightarrow$

$$\sigma=\frac{\log L-E(\overline{y})}{z-f(z)}\cong\frac{\log L-\overline{y}}{z-f(z)}$$

It is now possible to estimate the quantity X as well which gives β. This gives the result that is in the paper.

2. R codes

```
betaSub <- function(obs, lod){
x0=obs
N0=length(obs)
min=min(obs)
max=max(obs)
median=median(obs)
x=log(x0)
llod=log(lod)
bad=bad0=(x<=llod)  #bad = "censored"
#    = FALSE if above LOD
#    = TRUE if below
x[bad]=NA
nmiss=sum(bad) #the number of censored x's
pctcens=round((nmiss/length(obs))*100,2)
good=!bad
```

```
#lod=lod1
llod0=mean(llod[bad])
lod0=exp(llod0)

k=sum(bad)
ybar=mean(x[good])
z=qnorm(k/N0)
fz=dnorm(z)/(1-pnorm(z))
sy=(ybar-llod0) / (fz-z)
fsyz=(1-pnorm(z-sy/N0)) / (1-pnorm(z))

if(k!=0){
 bmean=N0/k*pnorm(z-sy)*exp(-sy*z+sy^2/2)
 am = mean(c(x0[good],bmean*lod[bad]))

 bgm=exp(-(N0-k)*N0/k*log(fsyz) - sy*z - (N0-k)/(2*k*N0)*sy^2)
 gm = prod(x0[good]^(1/N0)) * prod((bgm * lod[bad])^(1/N0))

}else{
 am =(sum(x0[good]))/N0
 gm = prod(x0[good]^(1/N0))
}
if(am>gm){
 sy0=sqrt(2*N0/(N0-1)*log(am/gm))
 gsd=exp(sy0)
}else{
 sy0=0
 gsd=1
}
```

```
x95=exp(log(gm)-sy0^2/(2*N0) + qnorm(.95) * sy0)
result <- data.frame(N0, nmiss, pctcens, min, max, median, am, gm, gsd, x95)
names(result) <- c("N", "N.Cens", "PctCens","Min", "Max", "Median" , "Mean",
"GM","GSD","X95th")
result <- round(result, 4)
return(result)
}
```

B. Maximum Likelihood Method

```
mle <- function(init_par, obs, lod)
{
cens=(obs==lod)*1

minus_loglik <- function(par, obs, cens){
  mu = par[1]
  sigma = exp(par[2])
  logliks <- ifelse(cens == 0, dlnorm(obs, mu, sigma, log=TRUE),
 log(plnorm(obs, mu, sigma)))
  return(-sum(logliks))
}

opt_obj <- optim(init_par, minus_loglik,
                 obs = obs, cens = cens)
#nocon=opt_obj$con

est <- exp(c(opt_obj$par[1], exp(opt_obj$par[2])))    ## antilog of average
 of estimate to find GM and GSD
```

```
GM <- est[1]
GSD <- est[2]
AM.ml=exp(log(GM)+.5*log(GSD)^2)  # old parenthesis

n=length(obs)
g = (log(GSD)^2/2)
Yfunc = (1+(n-1)*g/n + (n-1)^3*g^2/(n^2*(n+1)*factorial(2)) + (n-
 1)^5*g^3/(n^3*(n+1)*(n+3)*factorial(3)) + (n-
 1)^7*g^4/(n^4*(n+1)*(n+3)*(n+5)*factorial(4)))

  AM.mvue = GM*Yfunc

#SD=sqrt((AM)^2*exp(log(GSD)^2-1));  # SD for MLE not quite correct.
 Look page 251<-

X95th <-exp(log(GM) + 1.645*log(GSD))
result <- cbind(AM.ml, AM.mvue, GM,GSD,X95th)

return(result)
}
```

C. Kaplan-Meier Method

1. K-M equations for computing the cumulative distribution function

The equations for the K-M method can be found in the EPA ProUCL Technical Guide (EPA, 2007). Beal (2009) also described the method and provided SAS macro for it.

Let x_1, x_1... $x_n$represent the n concentrations (either detected concentrations or non-detects) obtained from environmental samples. The n concentrations are assumed to be statistically independent and representative samples from the environmental population being measured. Let y_1, y_2... y_p denote the p distinct values at which detects are observed so that p$\leq$ n. For j = 1, 2, …, p, let m_j denote the number of detects at y_j and let n_j denote the cumulative number of $x_j \leq y_j$. Define F(x) in Eqn. 1.

$$F(x) = 1 \qquad\qquad x \geq y_p$$

$$F(x) = \prod_{j \ni y_j}^{p} \frac{n_j - m_j}{n_j} \qquad y_j \leq x \leq y_{p-1}$$

$$F(x) = F(y_j) \qquad\qquad x_j \leq x \leq y_j$$

$$F(x) = 0 \qquad\qquad 0 \leq x \leq x_j$$

An estimate of the population mean μ using the KM method is shown in Eq. 2,

$$\widehat{AM} = \sum_{j=1}^{p} y_i \left[F(y_j) - F(y_{j-1})\right]$$

2. R codes

```
km <- function(obs, cens)
{
o <- array(obs)
detects <- obs[cens==0]
nondetects <- obs[cens==1]
tempCount <- 0
tempIndex <- 1
nextNumber <- -1
cc <- array() # CumNum same as N
for (n in 1:dim(o)[1]) {
```

```
  currentNumber <- o[n]
  currentState <-  cens[n]
  if (n < length(o)) {
    nextNumber <- o[n+1]
  } else {
    nextNumber <- -1
  }

  if(currentState == 0 && currentNumber != nextNumber) {
    cc[tempIndex] = n
    tempIndex <- tempIndex + 1
  }
}
uniqueTable <- as.data.frame(table(as.numeric(detects)))
uniques <- as.numeric(as.character(uniqueTable$Var1))
numUnique <- as.numeric(as.character(uniqueTable$Freq))
Fy <-  1
FyArray <- array()  #create empty array
len <- length(cc)
for ( n in 1:length(cc))
{
FyArray[n] <- Fy
Fy  <- Fy * (cc[len] - numUnique[len]) / cc[len]
len <-  len - 1
}
FyArray <- sort(FyArray)
aArray = array()
a <- 0
FyLength <- length(FyArray) - 1
for (n in 1:FyLength)  {
```

```
uNext <- uniques[n+1]
u <- uniques[n]
Fx <- FyArray[n]

a <- (a + ((uNext - u) * Fx))



aArray[n] <- a


}
# aArray

 # aJ
ajArray <- array()

for (n in 1:length(aArray))
{
  aValue <-  aArray[n]
nextNumUnique <-  numUnique[n+1]
nextN <-  cc[n + 1]
ajArray[n] <-  aValue * aValue * nextNumUnique / (nextN* (nextN -nextNumUnique));

}

# y
yArray <- array()

for(n in 1:length(FyArray))
{
  u <- uniques[n]
```

```
fyValue <-  FyArray[n]
if(n == 1)
  {fyPrevious = 0}
else
 { fyPrevious = FyArray[n - 1]}

yArray[n] <-  u * (fyValue - fyPrevious)

}
meanEst <-  sum(yArray)
# std dev
stdDevArray <- array()
for (n in 1:length(FyArray))
{ u = uniques[n]
fyValue <-  FyArray[n]
if (n == 1)
  {fyPrevious = 0}
else
  {fyPrevious <-  FyArray[n - 1]}

stdDevArray[n] <- ((u - meanEst) * (u - meanEst)) * (fyValue - fyPrevious)
}
stdDev <-  sqrt(sum(stdDevArray))
oLen <- length(o)
cLen <- length(nondetects)
stdErrMean <-  sqrt((oLen - cLen) / (oLen - cLen - 1) * sum(ajArray))
# compute X95
N= length(obs)
if (N > 19)
{
```

```
x = sort(obs)
i=floor(0.95*(N+1))
X95th = x[i]+(0.95*(N+1) - i) * (x[i+1]-x[i])
}
else
{X95th = NA}

x=sort(rnorm(20,2,4))
result <- data.frame(meanEst,stdDev, X95th)
 return(result)
}
```

D. Bayesian censored model

1. Codes for fitting Bayesian censored data model

WinBUGS codes

```
model{
# likelihood function
 for(i in 1:N){
 Y[i] ~ dnorm(mu, tau)C(,Y.cen[i])  # Y.cen = log LODs
}
# prior
 tau <- 1/(sigma*sigma)
sigma ~ dunif(log(1.01), log(12))
mu ~ dunif(log(0.025), log(500))
}
```

JAGS codes

```
  model {
  for (i in 1:N) {
  above.lod[i] ~ dinterval( x[i] , llodVec[i] )  # llodVec = vector of log LODs
  x[i] ~ dnorm( mu , tau)
  }
```

```
  mu ~ dunif(log(0.025), log(500))
  tau <- 1/(sigma*sigma)
  sigma ~ dunif(log(1.01), log(12))
}
```