

Exposure Assessment Techniques Applied to the Highly Censored *Deepwater Horizon* Gulf Oil Spill Personal Measurements

Mark R. Stenzel^{1,*}, Caroline P. Groth², Sudipto Banerjee³, Gurumurthy Ramachandran⁴, Richard K. Kwok^{5,6}, Lawrence S. Engel^{5,7}, Dale P. Sandler⁵ and Patricia A. Stewart⁸

¹Exposure Assessment Applications, LLC, 6045 N. 27th. St., Arlington, VA 22207, USA; ²Department of Epidemiology and Biostatistics, WVU School of Public Health, West Virginia University, One Medical Center Drive, Morgantown, WV 26506, USA; ³Department of Biostatistics, UCLA Fielding School of Public Health, University of California—Los Angeles, 650 Charles E. Young Drive, Los Angeles, CA 90095, USA; ⁴Department of Environmental Health and Engineering, Bloomberg School of Public Health, Johns Hopkins University, 615 N Wolfe St, Baltimore, MD 21205, USA; ⁵Epidemiology Branch, National Institute of Environmental Health Sciences, 111 T.W. Alexander Drive, MD A3-05, P.O. Box 12233, Research Triangle Park, NC 27709, USA; ⁶Office of the Director, National Institute of Environmental Health Sciences, 9000 Rockville Pike, Bethesda, MD 20892, USA; ⁷Department of Epidemiology, University of North Carolina at Chapel Hill, 35 Dauer Drive, Chapel Hill, NC 27599, USA; ⁸Stewart Exposure Assessments, LLC, 6045 N. 27th. St., Arlington, VA 22207, USA

*Author to whom correspondence should be addressed. Tel: 703-303-7585; e-mail: mark_stenzelEAApps@hotmail.com

Submitted 19 October 2020; revised 7 June 2021; editorial decision 4 July 2021; revised version accepted 13 July 2021.

Abstract

The GuLF Long-term Follow-up Study (GuLF STUDY) is investigating potential adverse health effects of workers involved in the *Deepwater Horizon* (DWH) oil spill response and cleanup (OSRC). Over 93% of the 160 000 personal air measurements taken on OSRC workers were below the limit of detection (LOD), as reported by the analytic labs. At this high level of censoring, our ability to develop exposure estimates was limited. The primary objective here was to reduce the number of measurements below the labs' reported LODs to reflect the analytic methods' true LODs, thereby facilitating the use of a relatively unbiased and precise Bayesian method to develop exposure estimates for study exposure groups (EGs). The estimates informed a job-exposure matrix to characterize exposure of study participants. A second objective was to develop descriptive statistics for relevant EGs that did not meet the Bayesian criteria of sample size ≥ 5 and censoring $\leq 80\%$ to achieve the aforementioned level of bias and precision. One of the analytic labs recalculated the measurements using the analytic method's LOD; the second lab provided raw analytical data, allowing us to recalculate the data values that fell between the originally reported LOD

and the analytical method's LOD. We developed rules for developing Bayesian estimates for EGs with >80% censoring. The remaining EGs were 100% censored. An order-based statistical method (OBSM) was developed to estimate exposures that considered the number of measurements, geometric standard deviation, and average LOD of the censored samples for $N \geq 20$. For $N < 20$, substitution of $\frac{1}{2}$ of the LOD was assigned. Recalculation of the measurements lowered overall censoring from 93.2 to 60.5% and of the THC measurements, from 83.1 to 11.2%. A total of 71% of the EGs met the $\leq 15\%$ relative bias and $< 65\%$ imprecision goal. Another 15% had censoring >80% but enough non-censored measurements to apply Bayesian methods. We used the OBSM for 3% of the estimates and the simple substitution method for 11%. The methods presented here substantially reduced the degree of censoring in the dataset and increased the number of EGs meeting our Bayesian method's desired performance goal. The OBSM allowed for a systematic and consistent approach impacting only the lowest of the exposure estimates. This approach should be considered when dealing with highly censored datasets.

Keywords: Bayesian methods; censored data; *Deepwater Horizon*; exposure assessment

Introduction

After the explosion of the *Deepwater Horizon* (DWH) on 20 April 2010, ~4.9 million barrels (780 000 m³) of oil were released from the Macondo oil well (Lehr *et al.*, 2010) into the Gulf of Mexico. The release of oil continued until 15 July 2010, when the damaged blowout preventer (BOP) was replaced with a functioning BOP that mechanically capped the well. Approximately 55 000 (National Institute for Occupational Safety and Health (NIOSH), 2011) workers and volunteers were identified as having worked in the oil spill response and cleanup (OSRC).

The National Institute of Environmental Health Sciences is investigating potential adverse health effects among workers involved in the OSRC for the Gulf Long-term Follow-up Study (GuLF STUDY) (Kwok *et al.*, 2017). A crucial component of this study was the estimation of exposures experienced by the 32 608 study participants. The goal of the study's exposure assessment was to develop descriptive statistics of exposures (e.g. arithmetic means [AMs]) from personal air measurements taken for various exposure groups (EGs), the members of which were expected to have had similar distributions of exposure (Stenzel, Arnold *et al.*, 2021). The exposure estimates were linked to the study participants' DWH OSRC work histories to estimate various measures of intensity and cumulative exposure (Stewart, Groth *et al.*, 2021).

In support of the OSRC, the Responsible Party (RP) of the spill (as designated by the US government) hired industrial hygiene and safety contractors, who collected ~28 000 personal air samples between 22 April 2010 and 30 June 2011, the GuLF STUDY exposure assessment period. Each sample was analyzed for several oil-related

compounds and in some cases, 2-butoxyethanol (2-BE), a component of one of the dispersants used in the cleanup effort, and D-limonene, a degreasing solvent, for a total of ~160 000 measurements. Over 93% of these measurements were below the limit of detection (LOD) as reported by the analytic labs. This high level of censoring would have limited our ability to explore exposure–response relationships in the study.

This paper describes the procedure used to reduce the censoring in the RP data by recalculating the measurement results to reflect the analytic methods' LODs rather than the higher reported values. A Bayesian method was selected for statistical analyses of the measurement data after a simulation study showed that it had an acceptable relative bias of $\leq 15\%$ and root mean square error (rMSE) of $< 65\%$ (together referred to here as the method performance goal) (Huynh *et al.*, 2014, 2016). Meeting the performance goal required ≥ 5 measurements with $\leq 80\%$ censoring per EG. A substantial number of EGs, however, continued to have >80% censoring even after the recalculation to the analytic LOD. The paper then describes the methods used to develop descriptive statistics of exposure for the remaining EGs.

An overview of the entire exposure assessment is available in Stewart, Groth *et al.*, (2021). The development of EGs is described in Stenzel, Arnold *et al.*, (2021). The measurement descriptive statistics for the oil-based chemicals of interest (total hydrocarbons [THC], benzene, toluene, ethylbenzene, xylene, and *n*-hexane [BTEX-H]) are described in Huynh *et al.*, (2021a,b,c) and Ramachandran *et al.* (2021). Other exposure estimates are found in Arnold *et al.* (2021) and Stenzel, Groth *et al.* (2021) (dispersant aerosols and vapors, respectively), Pratt *et al.* (2021) (PM_{2.5}), and Stewart, Groth *et al.* (2021) (oil mist). Dermal

exposure estimation is described in [Gorman Ng et al., \(2021\)](#) and [Stewart, Gorman Ng et al., \(2021\)](#).

Methods

Background

In the GuLF STUDY, participants' exposures were estimated by linking their detailed DWH OSRC work history information obtained via telephone interview to EGs through a job-exposure matrix. To develop the EGs, we identified determinants of exposure we deemed to be important: job/activity/task/vessel (e.g. roustabout, skimming); location (near shore, Louisiana); and dates (as time periods, e.g. the 2-month period during which dispersant was injected into the plume of the leaking oil at the wellhead) ([Stenzel, Arnold et al., 2021](#)). These three determinants (job/activity/task/vessel, location, time period) were assigned to each measurement result and formed the basis of unique EGs.

Data collection and exclusions

We used the ~28 000 personal air samples provided by the RP as the basis of our exposure assessment for the epidemiologic study because that database comprised the largest set of measurements collected during the OSRC of the DWH disaster. The samples generally represented full-shift exposures (≥ 4 - and ≤ 18 h) and covered the widest spectrum of jobs, activities, and tasks performed by the OSRC workers of all available databases. The RP contractors collected the measurements to assess compliance with established workplace standards and guidelines. All the measurements were collected before the GuLF STUDY began.

The RP predominantly used two companies to provide industrial hygiene and safety services for the OSRC effort. One primarily supported offshore [≥ 3 nautical miles (nmi) = 5.56 km from shore] operations (referred to here as lab 1) and the other, near shore (< 3 nmi from shore) and land operations (lab 2). Most samples were collected using passive dosimeters (3M 3500 or 3520, Assay Technology 521, SKC 575).

[Table 1](#) describes the analytes and number of measurements for the 28 000 samples (~160 000 total measurements) collected between 22 April 2010 and 30 June 2011. A number of these measurements were excluded from analysis for the reasons specified below. THC, measured as total petroleum hydrocarbons; benzene, toluene, ethylbenzene, *ortho* (*o*-), *meta* (*m*-), and *para* (*p*-) xylene, and *n*-hexane (BTEX-H) were the primary focus of the epidemiologic study. The isomers of xylene were combined and expressed as total xylene or xylene.

2-BE, as a component of one of the dispersants sprayed, was also of interest. Other chemicals analyzed in the measurements were not considered in the study, but may be considered in future evaluations.

A total of 10 032 measurements were excluded for durations outside the general work-shift duration of 4–18 h ($n = 6824$) or where sample documentation indicated that the sample was not valid (e.g. handling/storage reasons, such as the sample not being properly capped, a sample transmittal form was missing), or the sample documentation was missing or was so limited that it was not apparent what activity was measured ($n = 3208$).

Typical lab analyses

Typically, to determine LODs of analytic methods, labs prepare standards at various analyte concentrations in the method's desorption solvent to reflect the expected range of exposures to be measured. Each standard is then analyzed and a calibration curve prepared to identify the relationship between the concentrations in the standards and the area counts (ACs) of the analytic instrument [here, a gas chromatograph (GC)] readout. The range of concentrations used when compliance is of interest is generally 0.1–2 times the target concentration ([OSHA, 2005](#)). The detection limit of the analytical procedure (DLAP), in micrograms (μg), is determined then from the calibration curves ([OSHA, 2005](#)). For any sample at or below the DLAP, the DLAP is divided by the sample air volume to obtain the sample LOD. Because sampling duration varies by sample, a LOD is calculated for each sample, considering its duration. Typically, the DLAP corresponds to an AC that is three times the average AC of the blank samples. Thus, this true LOD is based on the capability of the method, i.e. the lowest concentration in a sample that can be quantified with the desired precision and accuracy and can be distinguished from the background concentrations observed in blank samples.

The calibration curve is expressed by equation (1).

$$(1) \quad AC_a = MS_a \times m_a + b_a$$

where 'a' is the analyte of interest (e.g. THC), AC_a is the measured GC AC of the sample corresponding to the analyte, MS_a represents the mass of analyte in the standard (μg), m_a is the slope of the calibration curve corresponding to analyte, and b_a is the intercept of the calibration curve corresponding to analyte.

Equation (1) can be rearranged to calculate the mass of the analyte corresponding to the observed AC:

$$(2) \quad MS_a = [(AC_a - b_a) / m_a]$$

Table 1. Number of measurements collected by the RP contractors between 22 April 2010 and 30 June 2011 before exclusions (see text).

Analyte measured	Number of measurements collected in the study period
2-Butanone	5
2-Butoxyethanol	1029
Acetone	9
Benzene	27 861
Cyclohexane	3722
Ethylbenzene	27 861
Limonene	244
<i>n</i> -Heptane (C7)	3722
<i>n</i> -Hexane (C6)	3722
Petroleum distillates	38
Tetrahydrofuran	3722
THC, calculated as <i>n</i> -hexane ^a	27 861
Toluene	27 861
Trimethylbenzenes, total	3722
Xylenes, total (<i>m</i> -, <i>o</i> -, and <i>p</i> -isomers)	27 861
Total measurements	159 240

RP, Responsible Party of the spill; THC, total hydrocarbons, analyzed as total petroleum hydrocarbons.

^aSee text for details on the *n*-hexane calculation.

After the calibration curve is developed, a known volume of sample desorption solution is injected into the GC. This desorption solution may be diluted to optimize the instrument's sensitivity. Equation (3) represents the dilution adjustment (DA).

$$(3) \quad MS_a = [(AC_a - b_a) / m_a] \times DA$$

The DA is the inverse of the dilution amount. Thus, if the desorption solvent is diluted in half, the DA equals 2 to reflect that the analyte concentration in the original desorption solution would have been twice as large if the sample had not been diluted.

The AC corresponding to each analyte peak in the sample is entered in equation (3). The mass is then adjusted for the volume of air sampled by:

$$(4) \quad \text{Sample air volume}_a = [\text{sampling duration} \times (\text{dosimeter sampling rate}_a / 1000 \text{ ml/L})]$$

where sample air volume_a is the air volume sampled by the dosimeter in liters (L) associated with the specific analyte in the specific sample, sample duration_a is the time (min) that the dosimeter sampled the air, and dosimeter sampling rate_a is the analyte's specific sampling rate provided by the dosimeter manufacturer (ml min⁻¹).

The calculated mass of analyte_n in the sample (equation 3) obtained from the calibration curve is then

divided by the air volume sampled (equation 4) to obtain the personal sample concentration (C_a) expressed in units of mg m⁻³. Note that µg L⁻¹ = mg m⁻³.

$$(5) \quad C_a = MS_a / \text{sample air volume}_a$$

The sample concentration expressed in mg m⁻³ is converted to a concentration in ppm using equation (6).

$$(6) C_a(\text{ppm}) = C_a \times 24.45 / (\text{analyte}_a\text{'s molecular weight})$$

The RP labs' analyses

The two labs used by the RP for the analysis of the measurements were certified by the American Industrial Hygiene Association (AIHA) Laboratory Accreditation Program, LLC administered by the AIHA and NIOSH. The sample documentation indicated that the labs analyzed the samples using NIOSH method 1501 (hydrocarbons, aromatic), 1500 (hydrocarbons), 1550 (naphthas), and in some cases, OSHA 7 method (GC-FID solvents on charcoal) (NIOSH, 1994a,b,c; OSHA, 2000). The labs initially analyzed the samples for THC (measured as total petroleum hydrocarbons), benzene, toluene, ethylbenzene, total xylene (*o*-, *m*-, and *p*-isomers) (BTEX), and in some cases, 2-BE. Analysis of *n*-hexane (H) was added by lab 1 beginning 26 May 2010. Lab 2 did not analyze any of their samples for *n*-hexane.

THC in this study is a composite of the volatile components of total petroleum hydrocarbons and is

Table 2. Censoring in the original RP personal measurements (after exclusions, see text) and after recalculation.^a

	Benzene	Cyclohexane	EB	<i>n</i> -Heptane	<i>n</i> -Hexane	Toluene	THC	TMBs	Xylenes, total ^b	2-BE	D-Limonene	Grand total
Personal samples (N) ^c	26 161	3361	26 161	3361	3361	26 161	26 161	3361	26 161	941	221	14 5411
Lab 1	4615	3361	4615	3361	3361	4615	4615	3361	4615	941	0	37 460
Lab 2	21 546	0	21 546	0	0	21 546	21 546	0	21 546	0	221	107 951
Personal samples censored prior to recalculation (N)	25 403	3036	25 564	2906	2730	24 985	21 738	3094	25 188	752	89	135 485
Lab 1	4541	3036	4198	2906	2730	4061	2225	3094	3877	752	0	31 420
Lab 2	20 832	0	21 366	0	0	20 924	19 513	0	21 311	0	89	104 065
Percentage of censored samples prior to recalculation	97.1	90.3	97.7	86.5	81.2	95.5	83.1	92.1	96.3	79.9	40.3	93.2
Lab 1	98.4%	90.3%	91.0%	86.5%	81.2%	88.0%	48.2%	92.1%	84.0%	79.9%		83.9%
Lab 2	96.8%	0.0%	99.2%	0.0%	0.0%	97.1%	90.6%	0.0%	98.9%	0.0%	40.3%	96.4%
Personal samples censored after recalculation (N)	23 210	2354	22 523	1946	1716	17 544	2920	1517	13 702	548	28	88 008
Lab 1	4039	2354	2576	1946	1716	1701	1031	1517	1765	548	28	19 193
Lab 2	19 171	0	19 947	0	0	15 843	1889	0	11 937	0	0	68 815
Percentage of censored samples after recalculation	88.7	70.0	86.1	57.9	51.1	67.1	11.2	45.1	52.4	58.2	12.7	60.5
Lab 1	87.5%	70.0%	55.8%	57.9%	51.1%	36.9%	22.3%	45.1%	38.2%	58.2%	0.0%	51.2%
Lab 2	89.0%	0.0%	92.6%	0.0%	0.0%	73.5%	8.8%	0.0%	55.4%	0.0%	12.7%	63.8%
Reduction in percentage censoring	8.4%	20.3%	11.6%	28.6%	30.2%	28.4%	71.9%	46.9%	43.9%	21.7%	27.6%	32.7%
Lab 1	10.9%	20.3%	35.1%	28.6%	30.2%	51.1%	25.9%	46.9%	45.8%	21.7%	0.0%	32.6%
Lab 2	7.9%	0.0%	6.6%	0.0%	0.0%	23.6%	81.8%	0.0%	43.5%	0.0%	27.6%	32.7%

^a2-BE, 2-butoxyethanol; EB, ethylbenzene; RP, Responsible Party; THC, total hydrocarbons; TMBs, trimethylbenzenes.

^b*m*-, *o*-, and *p*-isomers.

^cLab 1 primarily collected and analyzed offshore measurements; lab 2 primarily collected near shore and land measurements.

expressed, as is standard practice, as *n*-hexane. This means that all the components of THC were assumed to have the same sampling rate, sensitivity factor on the GC detector, and molecular weight as those of *n*-hexane. Once the concentrations were calculated to mg m⁻³, they were converted to ppm using *n*-hexane's molecular weight (86.16 grams mole⁻¹) and equation (6).

Focused on assessing compliance, the labs calibrated the GCs with prepared calibration standards relative to the target concentrations selected by the RP for compliance. The labs reported the concentration that corresponded to the lowest calibration standard as the LOD. We refer to that here as the reported LOD (RLOD) to distinguish it from the analytic method's LOD. Our original review of the measurements found that the RLODs were much higher than would have been observed based on the published expected performance of the analytical methods (OSHA, 1998, 1999, 2001).

Thus, although the labs referred to their censored data as being at or below the LOD, in actual practice they reported a measurement as censored if the result was below their calibration range, resulting in over 93% of the measurements being below the RLOD (Table 2). This high amount of censoring makes the development of stable, accurate, and reliable exposure estimates challenging. It also would have reduced the ability of the epidemiologic study to investigate exposure-response relationships because of the large number of study participants who would have been in the lowest exposure category.

Recalculation of the air samples

After realizing that the RLOD levels were unexpectedly high, we contacted the RP. We learned, as indicated above, that the labs had developed the RLODs based on compliance concerns rather than the analytic methods' capabilities. At our request, the RP asked both labs to recalculate the measurement data to reflect the analytical method's capabilities. The recalculation of the data did not involve re-analyzing the samples, but rather simply recalculating the measurement results using the already generated calibration curves. For those measurements below the RLOD, it was assumed that the instrument response was linear between the lowest calibration standard and the analytical method's LOD. This is a reasonable assumption (OSHA Methods 1005, 111, and 1002).

Lab 1 provided a new dataset with measurement results that reflected the analytical method's capabilities and set the LOD at three times the instrument signal associated with the blank samples. Lab 2 stated that they could not provide recalculated measurement data

because doing so would violate their written policy and procedures. Instead, they provided the raw lab data that allowed us to recalculate the measurement results. The raw data provided were: sample identification number, collection date, peak name (i.e. the analyte), sample dilution amount, ACs associated with each analyte of interest, and slope and intercept of each analyte's calibration curve. Lab 2 indicated they diluted all samples by a factor of two. This was possibly done to reduce the likelihood of saturating the GC detector, which otherwise could have resulted in a non-linear response.

We used equations (1–6) to recalculate the concentrations of all of lab 2's measurements. In some cases, the recalculated measurement result was less than the method's LOD due to AC uncertainty and calibration curve variability. In these cases, we set the measurement result at the method's average LOD derived from measurements with the same sample duration. For samples where the original measurement result was non-censored, the recalculated value was compared to the value originally provided by the lab to assure they were the same.

Xylene's LOD is twice as high as that of the other BTEX chemicals because 'xylene' corresponds to 'total xylene', which includes the *ortho* (*o*-), *meta* (*m*-), and *para* (*p*-) isomers. The *m*- and *p*-xylenes elute as one peak on the GC and the *o*-xylene as a second peak. The analytic LODs of the xylene peaks were 0.0023 and 0.0024 ppm, respectively. If both peaks were censored, we reported a LOD for total xylene as the sum of the LODs from the two peaks. If one peak was censored and the other non-censored, we added the non-censored result to the censored result and considered the total xylene measurement to be non-censored.

For both labs, we compared the RLOD to the method LOD to determine the impact this recalculation had on the analytical results, i.e. the reduction factor, by dividing the RLOD by the LOD.

Censoring issues within exposure groups

With the new measurements reflecting the recalculated sample concentrations, we developed our descriptive statistics for all job/work groups (Huynh *et al.*, 2021 a,b,c; Ramachandran *et al.*, 2021) with $N \geq 5$ and censoring <80% using the Bayesian method to achieve the average relative bias of $\leq 15\%$ and rMSE of <65% (Huynh *et al.*, 2016, Groth *et al.*, 2017, 2018, Table 3, Case 1). (Note: those authors used the term 'job group' or 'work group' as the basis for the measurement descriptive statistics for conditions $N \geq 5$ and censoring $\leq 80\%$. These job/work groups are a subset of the GuLF STUDY EGs. Many of the other EGs had larger N with greater censoring (discussed below) or incorporated

measurements from other EGs (Stewart, Groth *et al.*, 2021). Thus, the references below are to EGs because the measurements had >80% censoring.)

A substantial number of EGs did not meet the $\leq 80\%$ censoring and ≥ 5 measurements criteria. We extended the simulation work of Huynh *et al.* (2016) to identify other combinations of N and censoring with the same approximate level of performance. These are identified as Cases 3, 5, 7, 9, and 11 (Table 3). We allowed censoring up to 98% with sample N 's of 250–1000. These extended criteria allowed the Bayesian method to estimate the descriptive statistics while maintaining the same study level of accuracy and precision.

The remaining EGs did not meet our Bayesian criteria, but we used the Bayesian method if there was at least one non-censored measurement, which is identified as the 'High censoring Bayesian method' (Cases 2, 4, 6, 8, and 10, Table 3). In addition, when there were at least 1000 samples in an EG and at least one non-censored sample, we used the Bayesian method (Case 12, Table 3). In these cases, the confidence level of the estimate was reduced by one as described in Stewart, Groth *et al.* (2021).

In cases where all the measurements in the EGs were censored, we used an order-based statistical method (OBSM) to estimate the arithmetic average (AM), geometric mean (GM), and 95th percentile. The OBSM is based on the premise that the probability of observing a specific number of samples all below the LOD provides a way to estimate the AM, GM, and 95th percentile. For

example, if five measurements are collected, and all are observed to be below the LOD, it is expected that the GM and the AM are less than or equal to the LOD. We used this method when $N \geq 20$ and censoring was 100% (Case 13, Table 3). See the Supplementary Material (available at *Annals of Work Exposures and Health* online) for the underlying theory of this method. Finally, when sample size was 5–20 (groups with <5 measurements were not considered, see Stewart, Groth *et al.*, 2021, for treatment of these groups) and censoring was 100%, we used the substitution method to assign $\frac{1}{2}$ of the LOD (Case 14, Table 3).

We did not find a recognized statistical method to evaluate the OBSM estimates (Huynh *et al.* 2014, 2016) that considered variability due to study major determinants of location, time period, and activity/task. In the case of the BTEX-H chemicals, the priors did include information related to location and time period. A comparison study of a subset of our data (described in the Supplementary Material, available at *Annals of Work Exposures and Health* online) was conducted to determine how well the OBSM estimates predicted the Bayesian method estimates. We required, across two substances, a minimum of 100 pairs of EG estimates that did not meet our bias and precision performance goal but had at least one non-censored value (Cases 2, 4, 6, 8, 10, 12, Table 3). We selected benzene and toluene as the chemicals (Supplementary Table S1a, available at *Annals of Work Exposures and Health* online—benzene; Supplementary Table S1b, available at *Annals of Work Exposures and*

Table 3. Criteria used to identify the estimation method and the calculation of confidence in EG datasets.

Case No.	Number of measurements (N)	Censoring	Estimation method	Confidence level
1	≥ 5 but < 10	$\leq 80\%$	Bayesian estimate	No reduction
2	≥ 5 but < 14	$> 80\%$ but $< 100\%$	High censoring Bayesian method	Reduce by 1
3	≥ 14 but < 50	$> 80\%$ but $\leq 85\%$,	Bayesian estimate	No reduction
4	≥ 10 but < 50	$> 85\%$ but $< 100\%$	High censoring Bayesian method	Reduce 1
5	≥ 50 but < 100	$> 80\%$ but $\leq 90\%$	Bayesian estimate	No reduction
6	≥ 50 but < 100	$> 90\%$ but $< 100\%$	High censoring Bayesian method	Reduce 1
7	≥ 100 but < 250	$> 80\%$ but $\leq 95\%$	Bayesian estimate	No reduction
8	≥ 100 but < 250	$> 95\%$ but $< 100\%$	High censoring Bayesian method	Reduce 1
9	≥ 250 but < 1000	$> 80\%$ but $\leq 98\%$	Bayesian estimate	No reduction
10	≥ 250 but < 1000	$> 98\%$ but $< 100\%$	High censoring Bayesian method	Reduce 1
11	≥ 1000	$> 80\%$ but ≥ 5 non-censored measurements	Bayesian estimate	No reduction
12	≥ 1000	> 1 and < 5 non-censored measurements (i.e. $< 100\%$ censored)	High censoring Bayesian method	Reduce 1
13	≥ 20	100%	Order-based statistical method	Reduce 1
14	≥ 5 and < 20	100%	$\frac{1}{2}$ of the LOD for the specific analyte	Reduce 1

Health online—toluene). The Bayesian estimates (GM, AM, and 95th percentile) were plotted as the independent variables with the OBSM corresponding parameters as the dependent variables on the natural (normal) scale. Coefficients of determination (R^2) were calculated using linear regression for both chemical datasets of all EGs with $N = 1$ and with $N > 1$ non-censored measurements. The 93 benzene comparisons evaluated AMs [overall and by various subsets of location, time period, and geometric standard deviations (GSDs)], GMs, and 95th percentile. Due to the smaller numbers ($N = 20$), the toluene evaluation was limited to all data of non-censored measurements =1 and >1.

Results

Recalculation of personal samples

Table 2 shows the number of samples analyzed and the percentage of censoring by lab before and after the recalculation effort. THC had the largest decrease in percentage censoring of 83 to 11%. The other analytes had final censoring percentages between 45 and 70%, except benzene and ethylbenzene, which continued to be heavily censored (85–90%).

Table 4 presents the LODs for each analyte prior to and after recalculation and the reduction factors based on a 720-min sample. Lab 1's reduction factors were 3–23; lab 2's factors were 4–49. Benzene, THC, and 2-BE had the lowest factors for lab 1, whereas *n*-hexane had the highest factor. In contrast, for lab 2, benzene had the lowest reduction factor and THC had the highest reduction factor.

Censoring issues within exposure groups

Overall, 71% of the estimates for the EGs (43–95% by analyte) met the Bayesian performance goal of $\leq 15\%$ and $< 65\%$ average relative bias and rMSE, respectively (Table 5). THC had the highest percentage of estimates that met our performance goal (95%), whereas benzene, ethylbenzene, and *n*-hexane had the lowest percentage (43, 56, and 55%, respectively). The percentage of estimates for which we used our Bayesian method despite not meeting our performance goals was 15% (5–27% by analyte). The percentage of EG estimates for which we used the OBSM ranged from 0 to 8% by analyte (overall, 3%). Finally, we assigned $\frac{1}{2}$ of the LOD to 0–22% of the estimates by analyte (overall, 11%).

OBSM/Bayesian comparison study

Table 6 presents the results of the OBSM/Bayesian comparison study. The R^2 s between the benzene Bayesian

AM estimates and those calculated by the OBSM ranged between 0.64 and 0.91. When there was only one non-censored measurement for the EG, the R^2 for the AMs was 0.64–0.77. When there was more than one non-censored measurement, the R^2 s were higher (0.90–0.91). The same pattern was essentially seen for measurements taken on land and on water and by time period or using the GSD generated by the Bayesian method, rather than a single GSD for water (GSD = 6) or land (GSD = 8). The correlation decreased when the GM and the 95th percentile were calculated (0.71–0.92). For toluene EGs the R^2 s ranged from 0.47 to 0.50.

Discussion

We describe our efforts to reduce the percentage of censored data by recalculating the measurement data to reflect the analytic methods' actual LODs. The work was done on 145 462 measurements covering 11 analytes. As a result, the overall percentage of censoring dropped from 93 to 60%. This was important for the epidemiologic results in that it allowed greater differentiation among the study participants. It also provided a purer 'low exposed' subgroup of the study population to which higher exposed participants could be compared than had this recalculation not been done. The drop in the percentage censoring also increased the number of EGs with descriptive statistics with acceptable levels of relative bias and rMSE. This reduction was particularly important for THC. To estimate THC we used a moderately informed uniform Bayesian prior based on the THC measurement distribution within overarching groups of measurements (e.g. all land measurements) (Groth *et al.*, 2021). For the BTEX-H chemicals, however, we used the correlations between THC and each of the BTEX-H analytes as priors (Groth *et al.*, 2017, 2018).

The proportion of detectable measurements followed expected patterns (Table 2). Lab 1 samples were primarily collected offshore where exposure levels were expected to be higher due to the proximity to the well site where the fresh crude oil was being released. Thus, the percentage of detectable samples went from 16% (83.9% non-detectable) to 49% (51.2% non-detectable) across analytes. In contrast, the measurements analyzed by lab 2 were generally near shore (< 3 nmi from shore) and land samples and likely reflected greater amounts of weathering and lower concentrations of volatiles. Because the analytes of interest were some of the more volatile components of THC, a larger percentage of the BTEX-H had already disappeared, resulting in the near shore and land samples having lower levels of analytes and higher

Table 4. Reported LODs^a before and after recalculation.

Analyte	Lab 1 ^b			Lab 2		
	LOD prior to recalculation	LOD after recalculation	LOD reduction factor ^c	LOD prior to recalculation	LOD after recalculation	LOD reduction factor ^c
Benzene	0.0120	0.0030	4.0000	0.0100	0.0024	4.1667
Ethylbenzene	0.0340	0.0027	12.5926	0.0600	0.0023	26.0870
Toluene	0.0344	0.0030	11.4667	0.0600	0.0023	26.0870
Total hydrocarbons	0.3910	0.1139	3.4328	0.4900	0.0100	49.0000
Total xylene	0.0697	0.0057	12.2281	0.1900	0.0047	40.4255
Cyclohexane	0.0391	0.0026	15.0385	NA ^d	NA	NA
<i>n</i> -Heptane	0.0365	0.0024	15.2083	NA	NA	NA
<i>n</i> -Hexane	0.0625	0.0027	23.1481	NA	NA	NA
Trimethylbenzenes	0.0944	0.0064	14.7500	NA	NA	NA
2-Butoxyethanol	0.0460	0.0151	3.0464	NA	NA	NA
D-Limonene	NA	NA	NA	0.2700	0.0067	40.2985

LOD, limit of detection.

^aSample duration of 720 min.

^bLab 1 primarily collected and analyzed offshore measurements; lab 2 primarily collected near shore and land measurements.

^cReduction factor = reported LOD/LOD after recalculation.

^dAnalyte not analyzed (NA) by lab.

Table 5. Number (*N*) and percentage of EGs by estimation method and by analyte.

Analyte	Estimation method				Total
	Bayesian ^a (N, %)	High censoring Bayesian ^b (N, %)	OBSM (N, %)	Substitution ^c	
Total hydrocarbons	2541 (95)	125 (5)	0	7 (0)	2673
Benzene	1147 (43)	727 (27)	211 (8)	588 (22)	2673
Toluene	2338 (87)	216 (8)	9 (0)	110 (4)	2673
Ethylbenzene	1507 (56)	480 (18)	139 (5)	547 (20)	2673
Xylene	2418 (90)	144 (5)	4 (0)	107 (4)	2673
<i>n</i> -Hexane	1462 (55)	715 (27)	142 (5)	354 (13)	2673
Total	11 413 (71)	2407 (15)	505 (3)	1713 (11)	16 038

^aBayesian ($\leq 80\%$ censored, $N \geq 5$; or higher censoring at greater values of N as indicated in Table 3, Cases 1, 3, 5, 7, 9, 11).

^bHigh censoring Bayesian (indicated in Table 3, Cases 2, 4, 6, 8, 10, 12).

^cSubstitution method = LOD/2.

levels of censoring. The percentage of detectable samples went from 4% (96.4% non-detectable) to 36% (63.8% non-detectable) across analytes. The percentage of EGs by analyte with high censoring followed the same pattern as the percentage of measurements below the LOD.

Some differences were observed between the labs' data. The LOD reduction factor for the various analytes ranged from 3.0 to 23.1 for lab 1 and from 4.2 to 49 for lab 2. These differences may have occurred for several reasons. First, because the recalculation method for both labs was basically the same, lab 2's higher reduction factors may be due to the latter's using higher original calibration standards. Second, lab 2's recalculated LOD for THC (0.01 ppm) was somewhat lower than lab 1's (0.11 ppm). As a result, lab 2's THC reduction factor was considerably higher than the BTEX-H chemicals reduction factors (see the following paragraph). Third, lab 1 may have run the THC analysis on a less sensitive GC detector channel to avoid overloading the detector, which would have resulted in a higher observed THC LOD.

Censoring was reduced the most for THC (83.1 to 11.2%) and the least for benzene (97.1 to 88.7%) and ethylbenzene (97.7 to 86.1%). THC is a composite of the volatile chemicals in crude oil, primarily aromatics, and alkanes, all having approximately the same sensitivity on the GC hydrogen flame ionization detector used in the analytical method. Thus, the THC method's LOD was expected to be comparable to that of the BTEX-H chemicals (~ 0.01 ppm). Both labs, prior to recalculation, reported their THC LOD ~ 0.5 ppm and therefore, there was a reduction factor of nearly 50 for lab 2's recalculation (LOD = 0.01 ppm), but ~ 5 for lab 1 (LOD = 0.11 ppm), resulting in many of lab 2's THC measurements becoming detectable despite the lower levels (Huynh *et al.*, 2021a,b,c).

In contrast, the reduction factors for benzene and ethylbenzene were small. There are several likely reasons for this benzene finding. First, because of benzene's low occupational exposure limit (OEL) (0.5 ppm), the calibration standards were prepared at much lower levels than the other analytes' standards (with OELs >20 ppm), this meant there was a smaller difference between the RLOD and the analytic method's LOD for benzene as compared to the other analytes. Also, benzene's percentage concentration in the crude oil was low, i.e. 0.23% versus 0.65% for toluene, 0.70% for the xylenes, and 1.40% for *n*-hexane (Reddy *et al.*, 2011). Additionally, benzene is more soluble in water than are the other TEX-H chemicals; its solubility is ~ 1800 ppm, compared to toluene (~ 500 ppm) and ethylbenzene, xylene, and *n*-hexane (latter 3 being below 200 ppm). This is important because the oil, released near the Gulf floor, traveled through 5000 ft (1524 m) of water before reaching the water surface, allowing a larger percentage of the benzene to be absorbed as it rose to the surface than the other chemicals of interest. Ethylbenzene's low reduction in censoring is likely due to its low concentration in the crude oil ($<0.1\%$). Also the calibration standards for ethylbenzene were also very low to begin with [the LOD was 0.034 (lab 1) and 0.06 ppm (lab 2)], the same as those observed for toluene, and the OEL considered for ethylbenzene was 100 ppm versus 20 ppm for toluene. As mentioned earlier, the labs developed calibration standards that reflected exposure limits. In the case of ethylbenzene, the calibration standards were likely artificially low as compared to its exposure limit.

Although the recalculation effort substantially lowered the amount of censoring, several EGs still had $>80\%$ censoring, the minimum threshold recommended by Huynh *et al.* (2016). In their simulation study, they varied the sample size from 1 to 100, the censoring

Table 6. Summary of the Bayesian/OBSM comparison study.

Analyte	Comparison #	Number of EGs	Metric	Time period	Description	GSD derivation ^a	Slope	Intercept	R ²
Benzene	1a	93	AM	All	All data	OBSM	1.44	0.21	0.76
Benzene	1b	66	AM	All	Non-censored, N > 1	OBSM	1.49	0.06	0.90
Benzene	1c	27	AM	All	Non-censored, N = 1	OBSM	2.05	0.26	0.74
Benzene	2a	70	AM	All	All land	OBSM	1.47	0.18	0.78
Benzene	2b	52	AM	All	All land, non-censored, N > 1	OBSM	1.52	0.04	0.90
Benzene	2c	18	AM	All	All land, non-censored, N = 1	OBSM	1.96	0.32	0.77
Benzene	3a	23	AM	All	All water	OBSM	1.30	0.35	0.67
Benzene	3b	14	AM	All	All water, non-censored, N > 1	OBSM	1.32	0.22	0.91
Benzene	3c	9	AM	All	All water, non-censored, N = 1	OBSM	2.41	0.07	0.70
Benzene	4a	93	AM	All	All data	Bayesian	1.44	0.21	0.76
Benzene	4b	66	AM	All	Non-censored, N > 1	Bayesian	1.62	0.03	0.90
Benzene	4c	27	AM	All	Non-censored, N = 1	Bayesian	1.89	0.35	0.64
Benzene	5a	44	AM	TP1b and TP2 ^b	All data	OBSM	1.35	0.27	0.72
Benzene	5b	49	AM	TP3-6	All data	OBSM	1.54	0.17	0.79
Benzene	6a	30	AM	TP1b and TP2	Non-censored, N > 1	OBSM	1.41	0.08	0.91
Benzene	6b	36	AM	TP3-6	Non-censored, N > 1	OBSM	1.60	0.03	0.90
Benzene	7a	93	GM	All	All data	OBSM	1.24	0.04	0.71
Benzene	7b	66	GM	All	Non-censored, N > 1	OBSM	1.29	0.12	0.82
Benzene	8a	93	95th %-ile	All	All data	OBSM	1.30	0.74	0.80
Benzene	8b	66	95th %-ile	All	Non-censored, N > 1	OBSM	1.32	0.28	0.92
Toluene	9a	20	AM	All	All data	OBSM	1.37	0.33	0.49
Toluene	9b	13	AM	All	Non-censored, N > 1	OBSM	1.81	-0.18	0.50
Toluene	9c	20	AM	All	All data	Bayesian	1.29	0.38	0.47

95th %-ile, 95th percentile; AM, arithmetic mean; GM, geometric mean; GSDs, geometric standard deviations; OBSM, order-based statistical method.

^aOBSM indicates that a GSD of 6.0 was used for water EGs, and 8.0 was used for land EGs. Bayesian indicates that the GSD generated by the Bayesian analysis was used in the OBSM calculation.

^bTP1a was not included because there were no EGs that met the study criteria.

5–90%, and GSDs from 2 to 5 using weakly informed uniform priors. We observed that we could achieve our same goal for relative bias and imprecision for groups with even greater censoring if we had a larger sample size.

We established rules to develop Bayesian exposure estimates for censoring >80%. Based on [Huynh et al.'s \(2016\)](#) study, we identified N and percentage censoring conditions that met the study bias and rMSE goals ([Table 3](#)). While for $N = 5$ and censoring $\leq 80\%$, there can only be 1 non-censored value, at <85% censoring for N between 14 and 50 there are at least two non-censored measurements, whereas at <90% with $N = 50$ –100 there are at least five non-censored measurements. From this, we inferred that at greater censoring levels, if there were at least five non-censored measurements, Bayesian methods would produce estimates that meet the study bias and precision performance goals. With $N \geq 1000$, we considered Bayesian methods to be acceptable for any degree of censoring as long as there were at least five non-censored measurements. With these added rules, we were able to achieve 71% of our EG estimates meeting our performance goals.

We also used Bayesian methods for those EGs with fewer than five non-censored values but at least one non-censored measurement. While bias and imprecision likely increased because the EG was more highly censored, the Bayesian model was still informative because observed values were actually present, although with higher uncertainty (i.e. wider 95% credible intervals). This uncertainty affected THC more than the BTEX-H chemicals, because the BTEX-H model had information from both the 1–4 non-censored measurement(s) and from the corresponding THC measurements, and additionally, the THC:BTEX-H correlations were used as priors. This approach allowed us to develop Bayesian estimates for another 5–27% (by analyte) of the EGs.

Some EGs had 100% censoring. In these cases, a Bayesian strategy would rely primarily on prior information because there is only information in the data that these measurements are censored. In addition, the prior would incorporate measurement information related only to the location and time period and not the activity, which is similar to the overarching priors we used for other estimates ([Groth et al., 2021](#)), this procedure would result in a wider intervals than we are comfortable with.

We chose not to drop these entirely censored EGs because epidemiologic analyses often use low-exposed people as a comparison group. Alternatively, we could have accepted these very wide intervals and decreased our confidence in the exposure estimates further. Instead, we developed estimates without credible intervals, since these estimates were below the LOD. We considered

using simple substitution, i.e. of $\frac{1}{2}$ of the LOD or the LOD divided by the square root of 2 ([Hornung and Reed, 1990](#)) but only used this approach for small datasets for reasons described in the [Supplementary Material](#) (available at *Annals of Work Exposures and Health* online).

Thus, for larger N , we used an OBSM to estimate descriptive statistics for EGs with 100% censoring. The OBSM is affected by two observed factors that impact the exposure statistics, i.e. the number of measurements collected and the samples' durations. The OBSM is dependent on the Z-score, which varies with sample size. To illustrate, consider a sampling strategy for benzene where an industrial hygienist collected for three groups, 20, 50, and 80 measurements, all 100% censored and the same sampling duration. If in each case, the LOD was 2.4 ppb and the GSD = 6, the AM estimate would be 0.60 ppb, 0.30 ppb, and 0.21 ppb, respectively, even though the measurements were 100% censored. [Supplementary Table S3](#) (available at *Annals of Work Exposures and Health* online) includes eight different cases including the cases used to generate the three concentrations in the prior sentence. [Supplementary Table S3](#) (available at *Annals of Work Exposures and Health* online) was included to illustrate the impact on the OBSM related to dataset size, number of censored measurements, LOD and GSD. It should also be noted that the table in an EXCEL format can be used to calculate GM, AM, and 95th percentile of varying sizes for datasets of interest to the reader. The equation numbers are included in the appropriate columns to allow the reader to configure the table in an EXCEL format.

Based on the results of our comparison study, for comparable size datasets with $N \geq 20$, the OBSM provided similar estimates of the AM, GM, and GSD [although slightly higher, as indicated by the slope ([Supplementary Table S2](#), available at *Annals of Work Exposures and Health* online)] to those generated by the Bayesian method with ≥ 1 non-censored measurement. The R^2 s were 0.50–0.90 over the subsets evaluated (land, water, time period, Bayesian GSDs). We deemed from this study that the method was reasonable.

The statistics described here were assigned to each EG. When we developed our EG exposure estimates, we assigned a relative level of confidence to reflect how well the participants' reported exposure determinant values matched the measurement determinant values (i.e. job/activity/task, etc. on a scale of 1–5, 5 being the highest) ([Stenzel, Arnold et al., 2021](#); [Stewart, Groth et al., 2021](#)) to allow stratification of the participants in the epidemiologic analyses by confidence in the estimates. On the basis of the work described here, we modified these confidences to reflect the expected average relative

bias and rMSE (Table 3). The original confidence was lowered by 1 if the sample size and censoring likely resulted in a higher level of bias and rMSE than our performance goal or if we used the OBSM or substitution method.

The limitations of this work include the use of Bayesian methods when the sample size/censoring criteria were not met, i.e. were worse than what was needed to achieve our goal of $\leq 15\%$ bias and $< 65\%$ rMSE. To minimize the impact of this procedure we assigned a lower confidence to these estimates, such that study participants could be excluded from the epidemiologic analysis if this deviation was of concern. A second limitation is the dependency of the OBSM on the sample size and the duration of the samples. However, by using the simple substitution method for $N \geq 5$ and < 20 , the impact of this dependency was diminished. In addition, for convenience, we used a single GSD for water EGs and a single GSD for land EGs. If the true GSDs were different this could influence the results. However, these values were chosen based on GSDs observed in our data, and since we cannot observe the true underlying GSD, this was the best option.

Our comparison of the OBSM estimates with Bayesian estimates, however, found R^2 s of 0.64–0.90, suggesting this procedure had little impact on the estimates. In addition, the simple substitution method results in no variability, but because the levels are below the LOD, there should be little impact on the epidemiologic analyses.

Strengths include the fact that most of the estimates (71%) had a low average relative bias ($\leq 15\%$) and rMSE ($< 65\%$). BTEX-H estimates were likely to have even less bias and rMSE because those estimates were developed from priors based on the THC: analyte correlations. Second, the estimates using less desirable methods (i.e. the OBSM and simple substitution) occurred only at very low exposure levels (100% censored) and affected only 14% of the estimates. In addition, the OBSM estimates compared well with Bayesian estimates in our comparison study. Also, by separating out the lower exposed groups (due to using the two methods), there is likely to be sharper contrast across study participants, so that if an exposure–response exists, there will be an increased likelihood of finding an association. Finally, we have provided transparency to our approach for treating censored data.

The high censoring in this database should have little impact on the epidemiologic analyses. If the epidemiologic study develops ordinal exposure categories, all

study participants assigned a OBSM or simple substitution estimate would likely be assigned a value that is equal to the non-exposed workers (i.e. those trained but never worked) or a value that is above the non-exposed group but below the LOD (0.01 ppm), minimizing any effect of overestimation of exposures or lack of variability. Thus, there should be no impact. If the epidemiologic analyses use the estimates as continuous variables, the participants assigned an OBSM or substitution estimate could have higher estimates than truth, but they would still be ranked higher than non-workers (assessed as 0.01 ppm) and lower than higher exposed groups with levels $> \text{LOD}$. This approach, then, is likely to have little impact on exposure disease risk estimates.

In today's workplaces, as exposures are becoming more controlled, highly censored datasets are likely to be encountered more frequently, particularly if the analytic lab is focused on compliance. The approaches described here to account for censored data should be considered by others who want to quantify low measurements that were analyzed to reflect compliance level LODs and not analytic methods' LODs or who want to develop exposure descriptive statistics for measurements of low-exposed subgroups that are characterized by measurements that were very highly or completely censored.

Conclusion

Recalculating measurements below the originally RLOD to reflect the analytic methods' LODs substantially strengthened the study exposure assessment effort by greatly decreasing the percentage censoring of the original measurements, especially for THC. This added information on THC increased the breadth of the EGs estimated by the Bayesian statistical methods, which were shown to have acceptable bias and imprecision in the EG exposure estimates. In addition, because the BTEX-H chemicals were informed by their relationship with THC, the recalculation increased the number of EGs with acceptable bias and precision for the BTEX-H chemicals. The various approaches presented in this paper which were based on dataset size and censoring, allowed the estimation of descriptive exposure statistics for all EGs developed for the GuLF STUDY. The better differentiation among EGs due to the methods described here allows for sharper contrast among individuals in exposure–response analyses and increases the likelihood of finding associations if they exist. Scientists with highly censored datasets may want to consider our approach.

Supplementary Data

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

Funding

This study was funded by the National Institutes of Health (NIH) Common Fund and the Intramural Research Program of the NIH, National Institute of Environmental Health Sciences (ZO1 ES 102945). Dr. Sudipto Banerjee acknowledges funding from the following grants: NIH/NIEHS RO1ES027027-01; NIH/NIESH R01ES030210-01; NSF DMS-1513654, NSF IIS1562303, NSF DMS-1916349.

Acknowledgements

We thank Wendy McDowell and Caitlin Roush of McDowell Safety and Health Services, LLC for their assistance in compiling the study information. We thank BP for providing the data used in these analyses.

Conflict of interest

The authors declare no conflict of interest relating to the material presented in this article. Its contents, including any opinions and/or conclusions expressed, are solely those of the authors.

Data availability

The data underlying this article will be shared on reasonable request, consistent with protections for the privacy of study participants and existing multi-party agreements. Requests should be made following instructions on the study website <https://gulfstudy.nih.gov>.

References

Arnold S, Stewart PA, Pratt GC *et al.* (2021) Estimation of aerosol concentrations of oil dispersants COREXIT™ EC9527A and EC9500A during the *Deepwater Horizon* oil spill response and clean-up operations. *Ann Work Expo Health*; **66**: i188–i201.

Gorman Ng M, Cherrie JW, Sleuwenhoek A *et al.* (2021) GuLF DREAM: a model to estimate dermal exposure among oil spill response and clean-up workers. *Ann Work Expo Health*; **66**: i218–i233.

Groth C, Banerjee S, Ramachandran G *et al.* (2017) Bivariate left-censored bayesian model for predicting exposure: preliminary analysis of worker exposure during the *Deepwater Horizon* oil spill. *Ann Work Expo Health*; **61**: 76–86.

Groth CP, Banerjee S, Ramachandran G *et al.* (2018) Multivariate left-censored Bayesian modeling for predicting exposure using multiple chemical predictors. *Environmetrics*; **29**(4): e2505.

Groth CP, Huynh TB, Banerjee S *et al.* (2021) Linear relationships between total hydrocarbons and benzene, toluene,

ethylbenzene, xylene, and *n*-hexane during the *Deepwater Horizon* response and clean-up. *Ann Work Expo Health*; **66**: i71–i88.

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

Huynh TB, Groth CP, Ramachandran G *et al.* (2021a). Estimates of occupational inhalation exposures to six oil-related compounds on the four rig vessels responding to the *Deepwater Horizon* oil spill. *Ann Work Expo Health*; **66**: i89–i110.

Huynh TB, Groth CP, Ramachandran G *et al.* (2021b) Estimates of inhalation exposures to oil-related components on the supporting vessels during the *Deepwater Horizon* oil spill. *Ann Work Expo Health*; **66**: i111–i123.

Huynh TB, Groth CP, Ramachandran G *et al.* (2021c) Estimates of inhalation exposures among land workers during the *Deepwater Horizon* oil spill clean-up operations. *Ann Work Expo Health*; **66**: i124–i139.

Huynh T, Quick H, Ramachandran G *et al.* (2016) A comparison of the β -substitution method and a Bayesian method for analyzing left-censored data. *Ann Work Expo Health*; **65**: 56–73.

Huynh T, Ramachandran G, Banerjee S *et al.* (2014) Comparison of methods for analyzing left-censored occupational exposure data. *Ann Occup Hyg*; **58**: 1126–42.

Kwok RK, Engel LS, Miller AK *et al.*; GuLF STUDY Research Team. (2017) The GuLF STUDY: a prospective study of persons involved in the *Deepwater Horizon* oil spill response and clean-up. *Environ Health Perspect*; **125**: 570–8.

Lehr B, Bristol S, Possolo A. (2010) Oil budget calculator [WWW document]. Available at https://www.restorethegulf.gov/sites/default/files/documents/pdf/OilBudgetCalc_Full_HQ-Print_111110.pdf. Accessed 1 June 2020.

NIOSH (National Institute for Occupational Safety and Health). (1994a) NIOSH method 1501 (hydrocarbons, aromatic). NIOSH manual of analytical methods. 4th edition. Available at <https://www.cdc.gov/niosh/docs/2003-154/pdfs/1501.pdf>. Accessed 28 July 2021.

NIOSH. (1994b) NIOSH method 1500 (hydrocarbons, BP 36°–216°C). NIOSH manual of analytical methods. 4th edition. Available at <https://www.cdc.gov/niosh/docs/2003-154/pdfs/1500.pdf>. Accessed 28 July 2021.

NIOSH. (1994c) NIOSH method 1550 (naphthas). NIOSH manual of analytical methods. 4th edition. Available at <https://www.cdc.gov/niosh/docs/2003-154/pdfs/1550.pdf>. Accessed 28 July 2021.

NIOSH. (2011) NIOSH *Deepwater Horizon* roster summary report. Available at <http://www.cdc.gov/niosh/docs/2011-175/pdfs/2011-175.pdf>. Accessed 28 July 2021.

OSHA. (1998) OSHA method 111 – toluene. Document number T-111-FV-01-9804-M. Available at <https://www.osha.gov/dts/sltc/methods/organic/org111/org111.pdf>. Accessed 28 July 2021.

OSHA. (1999) OSHA method 1002 – xylenes (*o*-, *m*-, *p*-isomers), ethylbenzene. Available at <https://www.osha.gov/dts/sltc/methods/mdt/mdt1002/1002.html>. Accessed 28 July 2021.

OSHA. (2000) OSHA method 7 (organic vapors). OSHA sampling and analytical methods. Available at

- mdcampbell.com/niosh/oshameth/org007/org007.html. Accessed 29 July 2021.
- OSHA. (2001) OSHA method 1005 benzene. Document number T-1005-FV-01-0111-M. Available at <https://www.osha.gov/dts/sltc/methods/validated/1005/1005.pdf>. Accessed 29 July 2021.
- OSHA. (2005) Evaluation guidelines for air sampling methods utilizing spectroscopic analysis. Document number T-0012-02-0510-M. Available at <https://www.osha.gov/dts/sltc/methods/spectroguide/spectroguide.html>. Accessed 2 June 2020.
- Pratt GC, Stenzel MR, Kwok RK *et al.* (2021) Modeled air pollution from *in situ* burning and flaring of oil and gas released following the *Deepwater Horizon* disaster. *Ann Work Expo Health*; **66**: i172–i187.
- Ramachandran G, Groth CP, Huynh TB *et al.* (2021) Using real-time area VOC measurements to estimate total hydrocarbons exposures to workers involved in the *Deepwater Horizon* oil spill. *Ann Work Expo Health*; **66**: i156–i171.
- Reddy CM, Arey JS, Seewald JS *et al.* (2011) Composition and fate of gas and oil released to the water column during the *Deepwater Horizon* oil spill. *Proc Natl Acad Sci USA*; **109**: 20229–34. doi:10.1073/pnas.1101242108.
- Stenzel MR, Arnold SF, Ramachandran G *et al.* (2021) Estimation of airborne vapor concentrations of oil dispersants COREXIT™ EC9527A and EC9500A, volatile components associated with the *Deepwater Horizon* oil spill response and clean-up operations. *Ann Work Expo Health*; **66**: i202–i217.
- Stenzel MR, Groth CP, Huynh TB *et al.* (2021) Exposure group development in support of the NIEHS GuLF STUDY. *Ann Work Expo Health*; **66**: i23–i55.
- Stewart P, Groth CP, Huynh TB *et al.* (2021) Assessing Exposures from the *Deepwater Horizon* Oil Spill Response and Clean-up. *Ann Work Expo Health*; **66**: i3–i22.
- Stewart PA, Gorman Ng M, Cherrie JW *et al.* (2021) Estimation of dermal exposure to oil spill response and clean-up workers after the *Deepwater Horizon* disaster. *Ann Work Expo Health*; **66**: i234–i246.