Assessing Exposures from the *Deepwater Horizon* Oil Spill Response and Clean-up

Patricia Stewart[1,*], Caroline P. Groth[2], Tran B. Huynh[3], Melanie Gorman Ng[4], Gregory C. Pratt[5], Susan F. Arnold[5], Gurumurthy Ramachandran[6,7], Sudipto Banerjee[7], John W. Cherrie[8], Kate Christenbury[9], Richard K. Kwok[10,11], Aaron Blair[12], Lawrence S. Engel[10,13], Dale P. Sandler[10] and Mark R. Stenzel[14]

[1]Stewart Exposure Assessments, LLC, 6045 N. 27th. St., Arlington, VA 22207, USA; [2]Department of Epidemiology and Biostatistics, School of Public Health, West Virginia University, One Medical Center Drive, Morgantown, WV 26506, USA; [3]Department of Environmental and Occupational Health, Dornsife School of Public Health, Drexel University, 3215 Market St., Philadelphia, PA 19104, USA; [4]School of Population and Public Health, Faculty of Medicine, 3rd Floor, 2206 East Mall, Vancouver, BC V6T 1Z3 Canada; [5]Division of Environmental Health, University of Minnesota, School of Public Health, 420 Delaware St. S.E., Minneapolis, MN 55455, USA; [6]Department of Environmental Health and Engineering, Bloomberg School of Public Health, Johns Hopkins University, 615 N. Wolfe St., Baltimore, MD 21205, USA; [7]Department of Biostatistics, Suite: 51-254 CHS. UCLA Fielding School of Public Health, 650 Charles E. Young Drive South, Los Angeles, CA 90095-1772, USA; [8]Insitute of Occupational Medicine, Research Avenue North, Riccarton, Edinburgh, Midlothian EH14 4AP, UK; [9]Public Health Sciences, Social and Scientific Systems Inc., a DLH Holdings Company, 4505 Emperor Blvd, Suite 400, Durham, NC 27703, USA; [10]Epidemiology Branch, National Institute of Environmental Health Sciences, National Institutes of Health, 111 T.W. Alexander Drive – MD A3-05, Research Triangle Park, NC 27709, USA; [11]Office of the Director, National Institute of Environmental Health Sciences, 9000 Rockville Pike, Bethesda, MD 20892, USA; [12]National Cancer Institute, 9609 Medical Center Drive, Building 9609 MSC 9760, Bethesda, MD 20892-9760, USA; [13]Department of Epidemiology, Gillings School of Global Public Health, University of North Carolina, 35 Dauer Drive, Chapel Hill, NC 27599, USA; [14]Exposure Assessment Applications, LLC, 6045 N. 27th. St., Arlington, VA 22207, USA

*Author to whom correspondence should be addressed. Tel: +0/703-534-2956; e-mail: trish_stenzeleaapps@hotmail.com

Submitted 27 September 2021; revised 31 October 2021; editorial decision 31 October 2021; revised version accepted 10 November 2021.

Abstract

The GuLF Study is investigating adverse health effects from work on the response and clean-up after the *Deepwater Horizon* explosion and oil release. An essential and necessary component of that study was the exposure assessment. Bayesian statistical methods and over 135 000 measurements of total hydrocarbons (THC), benzene, ethylbenzene, toluene, xylene, and *n*-hexane (BTEX-H) were used to estimate inhalation exposures to these chemicals for >3400 exposure groups (EGs) formed from three exposure determinants: job/activity/task, location, and time period. Recognized deterministic models were used to estimate airborne exposures to particulate matter sized 2.5 µm or less ($PM_{2.5}$) and dispersant aerosols and vapors. Dermal exposures were estimated for these same oil-related substances using a model modified especially for this study from a previously published model. Exposures to oil mist were assessed using professional judgment. Estimated daily THC arithmetic means (AMs) were in the low ppm range (<25 ppm), whereas BTEX-H exposures estimates were generally <1000 ppb. Potential 1-h $PM_{2.5}$ air concentrations experienced by some workers may have been as high as 550 µg m^{-3}. Dispersant aerosol air concentrations were very low (maximum predicted 1-h concentrations were generally <50 µg m^{-3}), but vapor concentrations may have exceeded occupational exposure excursion guidelines for 2-butoxyethanol under certain circumstances. The daily AMs of dermal exposure estimates showed large contrasts among the study participants. The estimates are being used to evaluate exposure–response relationships in the GuLF Study.

Keywords: *Deepwater Horizon*; total hydrocarbons; $PM_{2.5}$; dispersant; exposure assessment

Introduction

The *Deepwater Horizon* (*DWH*) explosion in 2010, resulting in the largest marine oil spill in US history, killed 11 workers and released ~5 million barrels (780 000 m^3) of oil into the Gulf of Mexico. The >55 000 people rostered by the US National Institute for Occupational Safety and Health (NIOSH) as having worked on some aspect of the spill (NIOSH, 2011) reported multiple adverse health symptoms, including respiratory irritation, heat-related disorders, mental and physical fatigue, headache, and nausea (King and Gibbins, 2011). The US National Institute of Environmental Health Sciences (NIEHS) initiated an epidemiologic study (the GuLF Long-term Follow-up Study) to investigate these and other acute and chronic health effects among 32 608 people who worked or were trained to work on the oil spill response and clean-up (OSRC) (Kwok *et al.*, 2017).

Although several epidemiologic studies have observed adverse health effects among oil spill workers, few evaluated exposure–response relationships (Laffon *et al.*, 2016). We had access to a large air monitoring database and an extensive amount of other exposure-related information on the OSRC. We used these data to estimate airborne exposures to oil-related components [total hydrocarbons (THC), benzene, toluene, ethylbenzene, xylene, and *n*-hexane (BTEX-H)]. Due to the lack of measurements, we modeled air concentrations of particulate matter sized 2.5 µm or less ($PM_{2.5}$) and dispersants and their components and subjectively evaluated oil mist exposures. We also modeled dermal exposures to the same oil-related substances. This report describes this exposure assessment process (Table 1), the results of which support the GuLF Study in the investigation of exposure–response relationships.

Background

The *Deepwater Horizon* oil rig explosion occurred in the Gulf of Mexico on 20 April 2010. When the oil rig sank 2 days later, it severed the riser pipe that connected the well to the rig, releasing oil into the Gulf of Mexico waters. Within about two weeks, two oil rigs arrived on the scene; the *Discoverer Enterprise* (*Enterprise*) to stop the oil release and the *Development Driller III* (*DDIII*) to drill a relief well. Shortly after that, two more rig vessels arrived, the *Helix Q4000* (*Q4000*) to support the *Enterprise* and the *Development Driller II* (*DDII*) to drill a second relief well. All four rigs were supported by 14 vessels piloting remotely operated vehicles (ROVs), referred to here as ROV vessels, and by a large number of marine vessels (MVs). Multiple efforts to stop the oil release were unsuccessful until 15 July 2010. The well was permanently sealed 10 August 2010. We call this component the *response effort*.

The *clean-up effort* included >9000 vessels that had been deployed by September 2010 (U.S. Coast Guard, 2011) to skim the water of surface oil, burn the surface oil, deploy boom to contain the oil, search for oil and oiled wildlife, and decontaminate (decon) the outsides of oiled vessels to prevent contamination around the ports and docks. The effort also comprised clean-up activities on land that occurred across four US Gulf coastal states, including patrolling beaches for oil, tar, and contaminated wildlife; cleaning beaches, manmade structures and marshes of oil; deconning vessels, equipment and booms; and wildlife rehabilitation. A large support staff included administrative support, security, cooks, housekeepers, material handlers, fuelers, and pilots.

Most of the OSRC work had been completed by 31 December 2010, although beach, jetties, and marsh

Table 1. Overview of the exposure assessment process

Steps	Reference	Comment
Questionnaire development		
Reviewed air measurement data for determinant information to develop exposure groups that could link to questions in the questionnaire	Stenzel, Groth, Huynh *et al.*, 2021	Due to limited job title information in measurement database, interview questionnaire focused on work activities
Developed dermal exposure questions	Stewart *et al.*, 2021	Input data for GuLF DREAM (below)
Exposure groups (EGs)		
Reviewed and coded measurement data for exposure determinants that were identifiable for study participants	Stenzel, Groth, Huynh *et al.*, 2021	Based on exposure determinants, primarily on job-activity-task, location, and time (Supplementary Tables S1 and S2, available at *Annals of Work Exposures and Health* online).
Exposure estimation		
THC and BTEX-H		
Investigated effect of censorship of measurements, small sample sizes, high variability, and multiple LODs and distributions on various statistical approaches. Selected estimation model	Huynh *et al.*, 2014, 2016	Selected Bayesian methods to develop exposure statistics: selected $N \geq 5$ and % censoring ≤80 as criteria for exposure estimation of an EG
Recalculated the personal monitoring data to reflect the analytic LOD	Stenzel, Groth, Banerjee *et al.*, 2021	Overall censoring decreased from 93 to 60%
Excluded inappropriate personal measurements and measurements of other non-study related chemicals	Stenzel, Groth, Banerjee *et al.*, 2021	Decreased total number of measurements (160 000; 143 000 THC and BTEX-H) to 135 000 THC and BTEX-H measurements
Developed methodology to predict exposures using Bayesian univariate and bivariate models from personal measurements	Groth *et al.*, 2017, 2018	Used Bayesian univariate model to develop THC exposure statistics; the bivariate model with THC to develop BTEX-H exposure statistics
Estimated priors for Bayesian analyses	Groth, Huynh *et al.*, 2021	Used correlations between THC and BTEX-H personal measurements based on high-level exposure determinants for priors
Developed personal exposure estimates	For rig workers (Huynh *et al.*, 2021a); other water workers (Huynh *et al.*, 2021b); and land workers (Huynh *et al*, 2021c)	For all EGs with $N \geq 5$ and % censoring ≤80, developed AMs, GMs, GSDs, and 95%iles and their 95% CIs for a JEM for each of THC and BTEX-H

Table 1. Continued

Steps	Reference	Comment
Developed methods to estimate EGs with measurements that did not meet estimation criteria	Stenzel, Groth, Banerjee *et al.*, 2021, Stewart *et al.*, this paper, SM	1) Relaxed rule of <80% censoring if number of measurements substantially exceeded 5 for AMs, GMs, GSDs, and 95%iles and their 95% CIs 2) Further relaxed rules when censoring and N criteria were not met 3) Where N was <5, used the concept of sister ships, sister states and sister time periods 4) Used order-based statistical method (probability Z-scores) if $N \geq 20$ and censoring = 100% for AMs, GMs, and GSDs and 95%iles 5) Used a substitution method if $N \geq 5$ to <20 and censoring=100% for AMs, GMs, GSD, and 95%iles Inserted into JEM
Developed methods to estimate EGs with 0 to <5 measurements	Stenzel, Groth, Banerjee *et al.*, 2021	Used similar EGs ('sister' rigs, broad rig job groups, 'sister' states, 'sister' time periods) for AM, GM, GSD and 95%ile estimates. Inserted into JEM
Reviewed >26 000 000 area VOC measurements and summarized	Groth, Banerjee *et al.*, 2021	Reduced number of measurements to ~22,000 VOCs hourly vessel estimates
Developed THC full-shift equivalent estimates from VOCs hourly vessel estimates based on THC:VOCs relationship. Calculated THC means from original THC measurements and converted THC estimates by vessel-time period. Applied bivariate method to develop BTEX-H estimates	Ramachandran *et al.*, 2021	Increased the number of vessel-days available for estimation by 60% for ROV and response marine vessels
PM2.5		
Estimated emissions from information in literature; modeled air concentrations from emissions, AERMOD and other input parameters; estimated air concentrations by averaging air concentrations across broad areas of the Gulf and days	Pratt *et al.*, 2021	Quantitative $PM_{2.5}$ air concentration estimates of AMS and GSDS for 18 areas across the Gulf of Mexico and 4 Gulf coastal states for TP1b for JEM
Dispersants		
Analyzed dispersant-related exposure situations. Applied AgDisp model with input parameters for aerosols from aerial and vessel spraying	Arnold *et al.*, 2021	Air concentration estimates of AMs, GMs, GSDs dispersant aerosol exposures. Analysis suggested low probability of exposure, generally low levels and low duration of exposure.
Modeled vapor exposures using two-box and Plume models with Monte Carlo simulations for dispersant handling or being in an area with dispersants	Stenzel, Arnold *et al.*, 2021	Air concentration estimates of dispersant vapor exposures for AMs GSDs, and 95th percentiles, and their 90% confidence intervals for the JEM. Generally low levels.

Table 1. Continued

Steps	Reference	Comment
Oil mist		
Applied professional judgment by 2 industrial hygienists and achieved consensus for pressure washing or wave action sources	This paper	Ordinal estimates of oil mist exposure (scale 0–4) for JEM
Dermal Exposures (THC, BTEX-H, PAHs, dispersants)		
Reviewed dermal estimation models and recent dermal exposure studies to modify previously published dermal assessment model	Gorman Ng *et al.*, 2021	Developed GuLF DREAM model
Applied study participant responses and professional judgment by industrial hygienists as input parameters to model	Stewart *et al.*, 2021	Participant-specific quantitative estimates of dermal exposures

N, number; CI, credible interval; JEM, job-exposure matrix; SM, supplementary materials. TP1b, 15 May, 2010–15 July 2010.

clean-up work continued on an as needed basis until after 30 June 2011, the last date of the exposure assessment for the NIEHS study.

OSRC workers comprised federal, state, and local workers, as well as employees of contractors or subcontractors to the Responsible Party (RP, as identified by the US government). Some workers had previous work experience in their jobs (e.g. drilling crews, marine vessel crews, US Coast Guard employees), but many were fishermen who were banned from fishing during the OSRC effort, while many others were hired from the general population (e.g. beach clean-up workers). The latter two groups typically were hired on a temporary basis to perform expected short-duration work (e.g. months), so many workers did not have formal job titles and reported performing a number of activities with varying start and end dates.

The unique characteristics of this oil spill, i.e. the occurrence of inhalation and dermal exposure to a variety of oil-related substances; the large water and land area involved; the thousands of vessels and hundreds of employers; the temporary nature and lack of detailed information on the work force, including the lack of descriptive job titles; the performance of multiple activities; the large number of measurements for oil-related substances; and the lack of useful measurements on other substances of interest meant that a creative approach was needed to develop a comprehensive set of estimates that would allow epidemiologists to investigate exposure–response relationships.

Methods

Daily inhalation and dermal exposures for six oil-related substances were identified for assessment: THC (measured as total petroleum hydrocarbons) and BTEX-H, based on their acute and chronic toxicities; the likely size of the exposed population; the level of effort required for the estimation; and the availability of monitoring data. Also assessed were daily aerosol and vapor air concentrations from spraying COREXIT™ EC9527A and EC9500A chemical dispersants on the water. $PM_{2.5}$ from the burning of oil and gas was assessed, as was oil mist aerosol primarily generated from medium- and high-pressure spraying used to decontaminate vessels, equipment, and booms of oil. Table 1 identifies the major components of the exposure assessment effort with the associated reports.

A structured telephone interview was administered between March 2011 and May 2013 to all study participants (N = 32 608) to collect detailed information on OSRC work and other data (Kwok *et al.*, 2017). A second in-home interview administered to a subset (N = 11 193 participants) in their homes collected further information. (the questionnaires may be found at https://gulfstudy.nih.gov/en/fr_researchers/fr_studyquestionnaires.html.) Because we had been unable to interview OSRC workers prior to questionnaire development, we used the available personal measurement data collected by the RP of the spill and by government agencies to identify job titles/activities/tasks performed over the study period. These job/activities/tasks were the basis for the occupational component of the questionnaire. For questions on dermal exposures, we used generally recognized parameters of dermal exposure (Vermeulen *et al*., 2002).

Exposure groups

We used a job-exposure matrix (JEM) approach for inhalation exposures, with exposure groups linking study participants' job/activity/task data provided in the questionnaire to the measurement data of each of the evaluated substances. We define an exposure group (EG) as a group of workers performing similar tasks who were expected to have a similar distribution of exposures due to their being characterized by similar exposure determinants. We reviewed the extensive literature and measurement documentation on the *DWH* event to identify likely exposure determinants (Table 2; Supplementary Appendix A and Supplementary Tables S1 and S2, available at *Annals of Work Exposures and Health* online; and Stenzel, Groth, Huynh *et al*., 2021).

A major determinant was job (asked of workers on the rig vessels) or, to compensate for the lack of job titles for many of the remaining workers, activity, or task (Table 2). A second determinant was the weathering of oil. Weathering changes the concentrations of the oil components in the oil over time and space due to wave action, evaporation, dispersion, emulsification, dissolution, other natural processes and/or application of dispersants. Because, however, study participants would not have been able to answer questions on the amount of weathering the oil had undergone, we used the proxies of location (specific vessel names for the 4 rig vessels, 14 ROV vessels, 3 burner fire control vessels, and 33 research vessels; one of 7 areas of the Gulf; and one of 4 US states) and time (one of 7 time periods to reflect likely step changes in exposure levels). Each EG was a unique combination of job-activity-task/location/time period. More detail is provided in Stenzel, Groth, Huynh *et al.* (2021).

In several of the accompanying papers (Huynh *et al*., 2021a, b, c; Ramachandran *et al*., 2021), the term of 'job group' or 'work group' was used for the group of workers for which exposure estimates were developed. These terms were used to distinguish them from the EGs in this article. In those papers, the groups were formed using the same unique job-activity-task/location/time period determinant combinations as the EGs described here. Exposure statistics [e.g. the arithmetic means (AMs), etc.], however, were reported only for groups for which there were ≥5 measurements and ≤80% censoring [see Estimation of oil-related exposures (THC and BTEX-H) from measurements below, for the basis of these criteria]. Here, for the epidemiologic study, we assigned those same statistics to the EGs that corresponded to the job/work groups in Huynh *et al.* (2021a, b) and Ramachandran *et al.* (2021) (i.e. job/work groups = EGs). There were, however, many (29%) other EGs that needed exposure estimates, but which did not have measurements that met those criteria and so were not reported in those papers. For these remaining EGs, we assigned exposure statistics using other criteria (see Estimates of Inhalation Exposures below).

The same EGs were considered for all inhalation exposures, although depending on the substance, many EGs were considered unexposed.

For dermal exposures, estimates were subject specific. The concept of a JEM and, for the most part, the same EGs (from the same determinants) was used, however, to replace missing data (see estimation of modeled dermal exposures below).

Measurements

Personal air measurements were taken by industrial hygienists, and analyzed by two laboratories, contracted to the RP. Passive organic vapor dosimeters (3M 3500 or 3520; Assay Technology 521; and SKC 575) were used to collect >28 000 full-shift personal exposure samples (~143,000 measurements) analyzed for 5 oil-related constituents: THC (measured as total petroleum hydrocarbons) and BTEX, and, for some samples, n-hexane. From these, >135,000 THC and BTEX-H measurements were used that met study inclusion criteria for the exposure assessment (Stenzel, Groth, Banerjee *et al*., 2021). Over 93% of the measurements were reported as below the laboratories' limits of detection (LOD). We learned that the laboratories had prepared calibration standards to investigate compliance, rather than the methods' actual limits of detection. Upon request to the RP, the two laboratories either recalculated the measurements, or provided the original data to allow us to recalculate the reported LODs, to the methods' LODs (Stenzel, Groth, Banerjee *et al*., 2021). The level of censoring fell to 60%.

The documentation of all samples was reviewed, and each measurement of each sample was coded for the same job-activity-task/location/time period exposure determinants as was done for the EGs to ensure an efficient estimation of exposures.

Table 2. Values of the exposure group determinants[a]

Determinant	Values	Definition
Job	35 jobs across rigs (+ 3 broad 'jobs', 'Crew', 'Operations', 'All jobs')	Actual job titles identified in the measurement data and reported by study participants
Activity-task[b]	• 13 vessel activities for thousands of other unnamed vessels	• Vessel activities include skimming water for oil and putting out boom
	• 6 worker water activities	• Personally skimmed and personally put out boom
	• 2 water tasks	• Cleaned oil pools
	• 37 worker land activities	• Put out boom
	• 9 land tasks	• Maintained pumps/tanks or pumped
Type of vessel	• Rig	• Rig: large drilling platforms or ships
	• ROV	• ROV: marine vessels piloting remotely operated vehicles (ROVs)
	• Fire control	• Fire control: supported flaring by rig vessels
	• RV	• RV: research vessels
	• Other	• Other: barges, crew boats, shrimpers, fishing vessels, recreational vessels, jon/draft/air boats
Vessel	Specified rig vessels (n = 4), specified ROVs (n = 14 + 'All ROVs'), specified RVs (n = 33 + 'All RVs')	NA
Area of the Gulf of Mexico	• Hot zone	• ≤1 nmi radius around wellhead
	• Hot zone/source	• ≤5 nmi radius around wellhead
	• Offshore	• >5 nmi from wellhead and >3 nmi from shoreline
	• Near shore	• ≤3 nmi from shoreline
US state	Louisiana, Mississippi, Alabama, Florida	NA
Time period (TP)[c]	TP1a (22 April - 14 May, 2010)	Oil being released, dispersant sprayed on water, water operations slowly increasing
	TP1b (15 May–15 July 2010)	Oil being released, dispersant sprayed on water, burning occurring, substantial water and land operations
	TP2 (16 July–10 August 2010)	Well sealed, dispersant and burning operations ended, other water operations decreasing somewhat, land operations still substantial
	TP3 (11 August–30 September 2010)	Water operations decreasing and vessels started being decontaminated and decommissioned, land operations starting to decrease
	TP4 (1 October–31 December 2010)	Water operations essentially over: most vessels decontaminated and decommissioned; land operation decreasing
	TP5 (1 January–31 March 2011)	Land operations decreasing
	TP6 (1 April–30 June 2011)	Land operations decreasing, warmer weather compared with TP5

NA, not applicable.

[a]See Stenzel, Groth, Huynh *et al.* (2021) for more detail.

[b]The difference between worker activity and worker task is not rigorous.

[c]The dates associated with the time periods are not rigid with regards to the events occurring within the time period and some activities may have occurred outside the designated time periods for a relatively small number of days or by a relatively small number of people.

In addition, over 26,000,000 area measurements of volatile organic chemicals (VOCs) had been collected using direct-reading instruments on 38 of the large ships near the wellhead (Groth, Banerjee *et al.*, 2021).

No measurements of $PM_{2.5}$ or of aerosolized dispersants were collected on workers. Few relevant measurements of dispersant vapors or oil mist were available. No dermal measurements were collected.

Estimation of oil-related exposures (THC and BTEX-H) from measurements

To address the high amount of censoring (i.e. 60%) across the measurements, we evaluated several methods that dealt with censoring (Huynh *et al.*, 2014, 2016). Bayesian methods provided the lowest relative average bias and imprecision as estimated by the root mean squared error (rMSE) (Huynh *et al.*, 2016). In addition, the method provided an estimate of coverage, based on credible intervals (CI, similar to confidence intervals). We chose as our performance goal an average relative bias of <15% and an average relative rMSE of <65%. At this level of performance, a measurement sample size of ≥5 and censoring ≤80% per EG was required.

To estimate THC descriptive statistics in ppm, accounting for measurements below LOD, we modeled THC measurements for each EG using a univariate Bayesian Method (Groth *et al.*, 2017; Huynh *et al.*, 2016). Then, we modeled those same descriptive statistics to BTEX-H with THC as the predictor and each chemical of interest (BTEX-H) as the response variable using a bivariate Bayesian method accounting for measurements below the LOD in both the response and predictor. In this process, BTEX-H measurements below the LOD and measurements that had been analyzed for THC/BTEX but not for *n*-hexane (Supplementary Appendix B, available at *Annals of Work Exposures and Health* online) were imputed (Groth *et al.*, 2017, 2018, Groth, Banerjee *et al.*, 2021). As priors for the Bayesian method, we used the correlations of THC: each BTEX-H chemical from overarching groups of measurements identified by high-level determinants (Groth *et al.*, 2018, Groth, Huynh *et al.*, 2021). In all cases, the natural log of each chemical was modeled to preserve the normality assumptions of these methods. Details of the method are in Huynh *et al.*, 2021a, b, c; Groth *et al.* (2017, 2018), Groth, Huynh *et al.*, 2021, and Supplementary Appendix B (available at *Annals of Work Exposures and Health* online).

There were many EGs, however, that did not meet our criteria of N ≥ 5 and ≤80% censoring. We reviewed the Huynh *et al.* (2016) data and found that with larger numbers of measurements, we could accept higher levels of censoring and still meet our goal for relative average bias and imprecision. We therefore developed rules as to the minimum number of measurements required for various higher levels of censoring (e.g. 80–85% censoring was accepted if N ≥ 14 but <50) (Stenzel, Groth, Banerjee *et al.*, 2021). For situations where the study criteria were not met but there was at least 1 non-censored measurement, our High censoring Bayesian method was used. For EGs with 100% censored measurements, we used an order-based statistical method if N ≥ 20, and we used a substitution method (i.e. ½ the LOD) if N ≥ 5 and <20 (Stenzel, Groth, Banerjee *et al.*, 2021). Even after these steps, there remained many EGs without estimates, because the EGs had <5 measurements. For these, we combined measurements across EGs using measurements from the EG's 'sister' rig vessel (*Enterprise* and the *Q4000, Development Driller II* and the *Development Driller III*); 'sister' state [Louisiana (LA) and Mississippi (MS), Alabama (AL) and Florida (FL)]; or 'sister' time period (TP) (TP1a and 1b, TP2–4, TP5–6). More detail is provided in Supplementary Appendix C (available at *Annals of Work Exposures and Health* online).

Workers on the 4 rig vessels, 14 ROV vessels, and several response MVs had some of the highest estimated exposures in the study, as they were located in the hot zone and source areas (Table 2). In addition, the response effort activities on these vessels were dynamic, resulting in high day-to-day variability in exposures. It was, therefore, important to have measurement data over a sufficient number of days to ensure accurate estimation. Sufficient numbers of personal sampling days were available for the rig vessels, but coverage of measured days was lower on the ROV vessels and other response MVs. To augment the personal sampling data, we took advantage of >26 000 000 direct-reading, approximately 1 min in duration, area measurements of VOCs collected on 38 vessels located near the wellhead. Although VOCs are not exactly the same as THC (100 ppm VOCs was roughly equivalent to 80 ppm THC in our study), VOCs comprise the same aromatic chemicals of interest to this study. The locations of the instruments (median number of instruments per vessel = 7) on the vessel were not reported, however, so we calculated hourly averages across all instruments on a vessel using a Bayesian model that accounted for censored data (Groth *et al.*, 2017, Groth, Banerjee *et al.*, 2021). From these we developed full-shift daily VOC averages on each vessel day and estimated the linear relationship between those full-shift averages and THC daily averages from the personal measurements (Ramachandran *et al.*, 2021). This relationship was used to impute THC for days without THC measurements. We validated this method using rig vessel data (Ramachandran *et al.*, 2021). We then estimated BTEX-H levels on these ships by using the bivariate Bayesian method described above with overarching groups (Groth *et al.*, 2017, 2018, Groth, Banerjee *et al.*, 2021).

Every inhalation THC and BTEX-H estimate associated with an EG was assigned a confidence rating to reflect our relative confidence in the estimate. If the measurements' determinants matched those of the EG (i.e. the same job-activity-task/location/time period), a rating of 5 was assigned. If any two of the three determinants matched, a rating of 4 was assigned, and if only one of the determinants matched, a 3 was assigned. No matches resulted in a 2 being assigned. If the censoring did not meet our Bayesian performance goal for relative bias and rMSE or there was 100% censoring of the measurements, the previously developed confidence was lowered by 1.

We applied Bayesian methods using Monte Carlo methods to develop 25 000 estimates of each descriptive statistic developed. For each EG described in this section, we estimated a posterior (i.e. the modeled) median AM, geometric mean (GM), geometric standard deviation (GSDs), 95th percentile (95%ile), and their corresponding 95% CIs. We also identified the determinants on which the EG was based and our confidence for each of our six substances.

For presentation purposes, we developed broad groups of jobs, all jobs on 'All rigs', 'All ROVs', All burner fire control vessels', 'All research vessels', 'All other water operations', and 'All land operations' using non-overlapping 95% CIs of the AMs to determine credible differences. We describe these as notable or credible because Bayesian analyses do not rely on statistical significance (i.e. *P*-values) but instead denote such differences based on the overlap of CIs and the non-inclusion of 0. Analyses were conducted in JAGS (Just Another Gibbs Sampler) (Plummer, 2003) and R (R Core Team). Analyses using the >26 million VOC observations were conducted using supercomputing through the Minnesota Supercomputing Institute.

Estimation of modeled $PM_{2.5}$ air concentrations

Oil and gas flared by two of the rig vessels and a single MV and oil burned *in situ* on the water surface by teams of smaller vessels in TP1b only (Table 2) created $PM_{2.5}$. Our goal here was to estimate maximum air concentrations to $PM_{2.5}$ arising from these sources, so as to inform future responders of oil spills. We applied AERMOD (US EPA, 2017), a recognized air dispersion model, to estimate hourly $PM_{2.5}$ air concentrations (μg m^{-3}). For most participants, we had no information on their location other than the general area they worked [i.e. our areas (Table 2)]. We therefore developed $PM_{2.5}$ estimates by area. First, *DWH*-specific information on the amount, duration, dates, location, and meteorology, along with emission data from the published literature, was used to estimate hourly $PM_{2.5}$ air concentrations at each of 3432 model receptors (points of intersections across an imaginary grid system of 10 × 10 km^2 squares used by the model) across the Gulf for each day of burning (Pratt *et al.*, 2021). From these concentrations, we estimated for each day the maximum 1- and 12 h and the average 24-h concentrations per receptor, resulting in three concentrations per day at each receptor. After averaging each 1-, 12-, or 24-h value across all of the receptors in each of our 7 Gulf areas per day (Table 2), we then derived AMs, GMs, and GSDs across all days of TP1b, by area for each duration. A third source of $PM_{2.5}$ came from the combustion of gas and diesel by vessel engines on the water. As few data were available on the type and location of vessels, we developed rough estimates for the time period of the highest and the lowest number of vessels. We had no information on mechanical equipment on land and so did not estimate $PM_{2.5}$ concentrations on land.

Estimation of modeled dispersant air concentrations

Dispersants COREXIT™ EC9527A (9527A) and COREXIT™ EC9500A (9500A) were sprayed onto the water by plane offshore, by vessel in the hot zone and source, and injected by wand into the oil plume directly over the wellhead, the latter deemed to have no aerosol generation, resulting in possible aerosol and vapor exposures.

We used AgDISP (Bird *et al.*, 2002) to provide estimates of direct (being near or under the spray) and indirect (from spray drift) total aerosol air concentration estimates (Arnold *et al.*, 2021). Known or estimated input data to the model for aerial applications were aircraft design specifications, spray characteristics, weather conditions, and topographical characteristics. For vessel applications, the input data were vessel position relative to the wind and to other vessels in the area, nozzle characteristics, spray composition, and meteorological conditions. We estimated 1-h (the shortest duration allowed by the model) concentrations of total aerosol (μg m^{-3}), reflecting various plane types, vessel deck heights above the water, wind speeds, at right angles to the flight path on the same horizontal plane downwind from the spraying. We conducted a cross-validation using AERMOD (US EPA, 2017) for two specific plane sorties using the same variables as used in AgDISP (Arnold *et al.*, 2021).

Exposures to vapors could have occurred from handling dispersant-related equipment (connecting/disconnecting lines, transferring dispersant and maintaining pipes, pumps, and tanks); cleaning up spills; collecting for research purposes water samples from dispersant-contaminated Gulf waters; and being in the area immediately after dispersant had been sprayed on the water. 2-BE (in 9527A) and propylene glycol (PG, in both 9527A and 9500A) were two components of interest in the dispersant vapors. Few measurements and little descriptive information were available on possible dispersant vapor exposures. Two-box (for indoor operations; Nicas, 2009) and plume (for outdoor operations; Armstrong, 2009) models were used to estimate air concentrations from activities that could have generated vapor exposures (Stenzel, Arnold *et al.*, 2021), and Monte Carlo simulations provided estimates of uncertainty due to varying wind speeds, surface areas, and air changes per hour. Estimates of the AMs, GMs, GSDs, and 95%iles and the 90% confidence intervals were developed.

The total aerosols and the 2-BE and PG vapor estimates were assigned to the appropriate EGs for TP1a and TP1b only (the time periods of dispersant use).

Estimation of ordinal oil mist exposures

Two industrial hygienists independently reviewed the activities performed by each EG and assigned a level of none (0), very low (1), low (2), medium (3), or high (4) to likely oil mist exposure levels. Estimates were finalized after consensus. No differentiation in exposure levels was made by time period within an exposure group.

Estimation of modeled dermal exposures

The oil-related substances and the dispersants also were of interest for skin exposure, either systemically or topically. Lacking measurements, we modified a previously published estimation model (van Wendel de Joode *et al.*, 2003) to suit our data, which we called GuLF DREAM (Gorman Ng *et al.*, 2021). The GuLF DREAM model considers chemical and physical properties and the frequency and intensity of three exposure pathways (emission, deposition, and surface transfer), as well as use of personal protective equipment (PPE) and contact with sea water. Estimates to THC, BTEX-H, and polycyclic aromatic hydrocarbons (PAHs) in oil and in tar, and THC and xylene (both representing petroleum distillates, hydrotreated light, a component of 9500A), were modeled.

Dermal exposures were study participant specific, as input came from responses to a series of questions in the interview questionnaires about contact with chemicals and sea water on clothing or on the skin and about PPE use (Stewart *et al.*, 2021). The same set of dermal questions was asked in reference to each job-activity-task reported (above, Methods, *Exposure groups*). Data for the other model variables were imputed by the study industrial hygienists (e.g. chemical and physical properties that reflected the degree of weathering the oil had undergone and the percent contamination of each body part). Missing responses were imputed from other respondents with the same job-activity-task/location/time period. The output for this model was a dimensionless 'GuLF DREAM unit' (GDU).

We reviewed exposure studies to validate our model but found only two that we considered relevant (Cavallari *et al.*, 2012; Christopher *et al.*, 2011). Using the documentation in those studies, we assessed exposures to oil and tar, respectively, using GuLF DREAM and compared the estimates to the measurements (Gorman Ng *et al.*, 2021).

For presentation purposes, we calculated AMs, GMs, GSDs, and 95% confidence intervals for broad groups of participants by time period and considered non-overlapping 95% confidence intervals to indicate statistical significance. Values < 0.02 GDUs were deleted from these calculations.

Assigning exposure estimates to participants

A JEM of airborne exposure statistics was developed for each substance of interest by EG (i.e. by job-activity-task/location/time period). Exposure statistics were assigned to the study participants through the participants' reported information that reflected the appropriate job-activity-task/location/time period EGs. The subject-specific dermal estimates were assigned directly to each study participant's reported job-activity-task/location/time period. The exposure assessment was done blind to any health outcome or subject-specific information.

Results

A total of 3420 possible EGs for each inhalation exposure was developed and considered for estimation. The actual number of EGs with an exposure estimate varied depending on the substance being estimated. For example, because virtually all the burning occurred in TP1b, estimates were developed for only that time period.

THC, BTEX-H

THC concentrations are expressed in ppm; BTEX-H are in parts per billion (ppb).

The posterior median AMs of the THC estimates ranged from a low of 0.01 ppm for 'Offsite driver' (All states, All time periods) to a high of 22.4 ppm [All workers, *Boa Sub C* (an ROV vessel), TP1a] (not shown). Only 4% of the estimated EGs exceeded 3 ppm and only 15% exceeded 1 ppm.

For THC, there were notable exposure differences across the broad groups of vessels and activities (Fig.

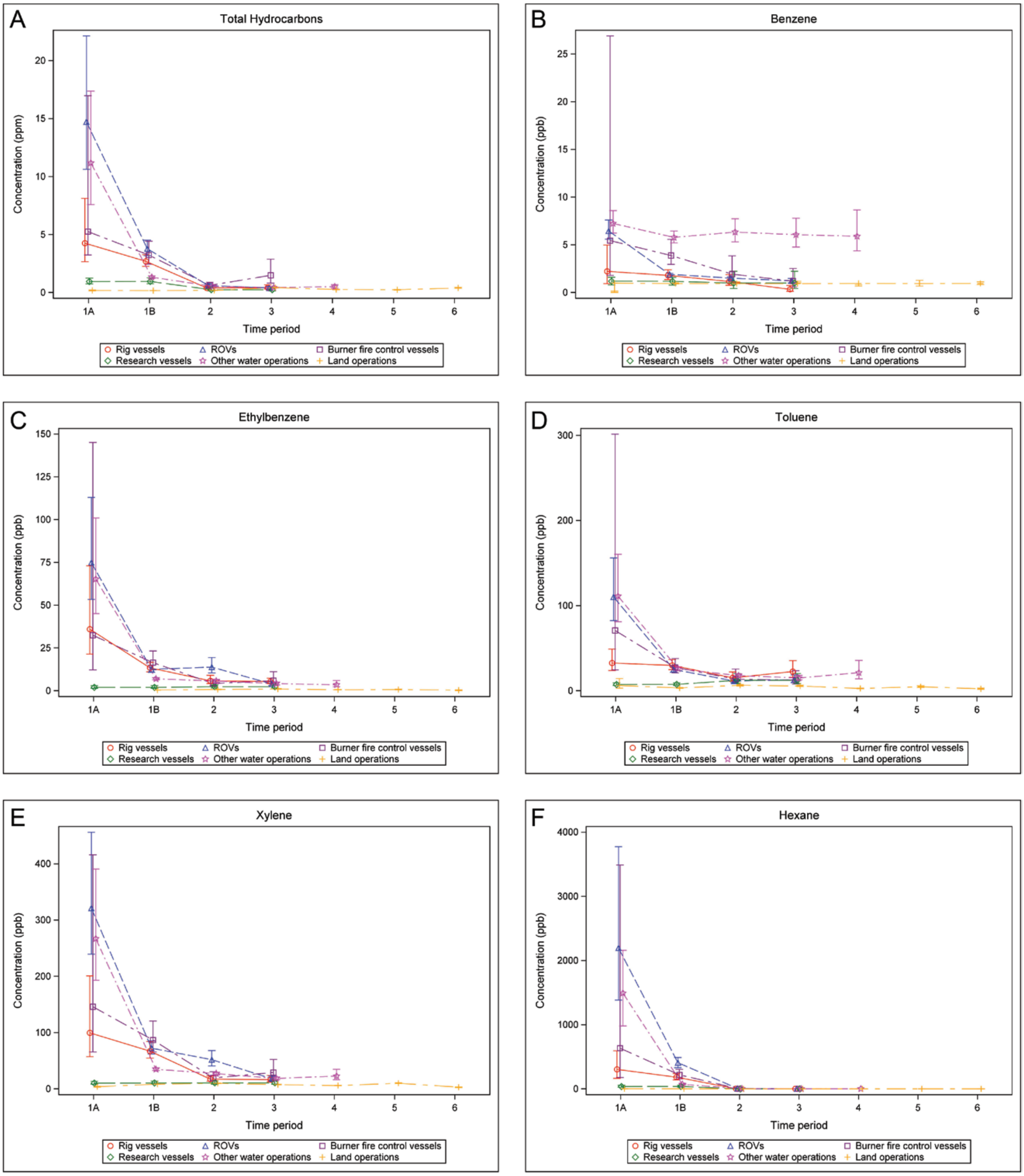

Figure 1. Inhalation exposure estimates for total hydrocarbons (THC), benzene, ethylbenzene, toluene, xylene and *n*-hexane (BTEX-H) by time period (THC = ppm; BTEX-H = ppb) for 'All rig vessels', 'All ROVs', 'All fire control vessels', 'All research vessels', 'All Other Water Operations', and 'All Land Operations'. ROV vessels: vessels that piloted remotely operated vehicles (ROVs). The benzene graph showing the higher AMs for 'All Other Water Operations' is due to an artifact of the methods used by the labs to calculate the limits of detection.

1 and Supplementary Table S3, available at *Annals of Work Exposures and Health* online). For example, in TP1b, the posterior median AMs for 'All rigs', 'All ROVs', and 'All burner fire control vessels' were greater than the posterior median AM for 'All other water operations', which was greater than the comparable AM for 'All land operations'. The patterns for the other time periods showed differences similar to those in TP1b.

Table 3. Estimated air concentrations to particulate matter$_{2.5}$ ($PM_{2.5}$), dispersant aerosols, dispersant vapors, and oil mist

Substance (unit of measurement)	Exposure group	Air concentration
1-/12-h $PM_{2.5}$ (μg/m^3) (all estimates for TP1b only)	Workers in the hot zone	545.0/96.9
	Workers in the hot zone/source	177.3/28.7
	Workers who conducted *in situ* burns	67.0/10.4
	Workers located on water offshore (all states)	8.8/1.3
	Workers located near shore (all states)	1.8/0.3
	Workers on land (all states)	1.5/0.2
Dispersant aerosol (μg m^{-3}) [all estimates for TP1a (plane only) and TP1b (plane and vessel) only]	By-stander concentration near application by plane (152 m from flight path)	9527A and 9500A: 0.4 to ~50 (TP1a-1b only)
	By-stander concentration near application by vessel (10 m from nozzle)	(9527A not applied by vessel) 9500A: 0.3–0.33 (TP1a–1b)
Dispersant vapor (ppb) [all estimates for TP1a (2-BE and PG) and TP1b (PG only) only]	Maintained pumps/tanks or dis/connected anything	Pumps: 2-BE: 15; PG: 2 Tanks: 2-BE: 495; PG: 53 Dis/connect: 2-BE: 85; PG: 9
	Pumped dispersant	2-BE: 77 PG: 8
	Took samples on a research vessel	2-BE: 25 PG: 3
	By-stander concentration near application by plane[a]	2-BE: 11 700 PG: 1240
	By-stander concentration near application by vessel	PG: 35

TP1a, 22 April–14 May, 2010; TP1b, 15 May–15 July, 2010; 2-BE, 2-butoxyethanol; PG, propylene glycol.
[a]See Stenzel, Arnold *et al*. (2021) for more information on interpretation of this scenario.

In addition, TP1a and TP1b generally had substantially higher posterior median AMs compared to those of TP2–6. Notable differences were observed for broad groups of jobs on the rig vessels (e.g. 'Outside Crew' and 'Outside operations', but generally not among specific jobs; Huynh *et al.*, 2021a). We also saw notable differences among the posterior median AMs for the individual activities performed on the water and on land (Huynh *et al.*, 2021b, c). Differences in the AMs were observed among specific areas of the Gulf waters (e.g. near the wellhead vs. near shore) and among the states (generally LA versus MS, AL, and FL) (not shown).

Figure 1 (and Supplementary Table S3, available at *Annals of Work Exposures and Health* online) also presents the mean daily exposures to BTEX-H for the same broad groups. The posterior median AM estimates for benzene ranged from <0.01 to 62.52 ppb (not shown). Only 5% of the EG AMs exceeded 10 ppb and 16% exceeded 3 ppb. Ethylbenzene posterior median AM values were between <0.01 ppb and 137.02 ppb. Of possible EGs, 24% were >3 ppb and 8% exceeded 10 ppb. For toluene, the posterior median AM estimates ranged from 0.02 ppb to 187.98 ppb. The percent of EGs >10 ppb was 29 and >30 ppb was 8. Xylene posterior median AM estimates ranged from 0.37 to 445.35 ppb. About 43% of the EGs had AMs > 10 ppb, whereas about 12% had AMs > 30 ppb. Finally, the lowest posterior median AM estimate for *n*-hexane was 0.02 ppb and the highest was 2441.0 ppb. About 17% of the EG AM estimates were >10 ppb and 9% were >30 ppb. Notable differences among the median AMs for the BTEX-H chemicals occurred less frequently than for THC; likely, in part, due to the higher censoring associated with these chemicals than with THC.

$PM_{2.5}$

The average maximum 1- and 12-h averages $PM_{2.5}$ air concentrations in the hot zone were 545.03 and 96.93 μg m^{-3}, respectively (Table 3 and Supplementary Table S4, available at *Annals of Work Exposures and Health* online) (Pratt *et al.*, 2021). The equivalent values for the combined hot zone/source areas (i.e. those assigned to the ROV, fire control, and other MVs near the wellhead) were 177.29 and 28.70 μg m^{-3}, respectively. Air concentrations at the *in situ* burns were estimated to have average daily maximum levels of 67.01 and 10.4 μg m^{-3}, respectively. In contrast, air concentrations closer to and on land were generally <1–9 μg m^{-3}, respectively. Estimates of average daily air concentrations from engine exhaust from vessels in the Gulf ranged from 0.17 to 14.3 μg m^{-3}, depending on the time period and location in the Gulf (not shown).

Dispersant aerosol and vapors

Air concentrations from direct (being near or under the spray) exposure to dispersants were deemed to be unlikely. Average predicted 1-h estimates of total aerosol concentrations resulting from aerial spray drift at 152–762 m (500–2500 ft) at right angles to the flight path on the same horizontal plane were similar for the two dispersants ranging from about 0.4 to ~50 μg m^{-3} (Table 3 and Supplementary Table S5a, available at *Annals of Work Exposures and Health* online) (Arnold *et al.*, 2021). Total aerosol estimates from vessel spray drift were 0.001–0.33 μg m^{-3} at horizontal distances of 10–500 m, respectively (Table 3 and Supplementary Table S5b, available at *Annals of Work Exposures and Health* online).

Air concentrations from two aerial sorties were derived using both AgDisp and AERMOD. The maximum 1-h AgDisp estimates for the two sorties were 420 and 365 μg m^{-3} at 0–20 m (Arnold *et al.*, 2021). The corresponding AERMOD predictions were 427 and 174 μg m^{-3}.

We deemed that no OSRC workers were located indoors or in protected areas, but we developed estimates for these situations for other possible oil spills (see Stenzel, Arnold *et al.*, 2021). Study participants with possible dispersant vapor exposures were likely to have been outdoors, and thus, estimates from the plume model were considered to be relevant. The average air concentration for 2-BE estimated for 'Maintained/worked on pumps/tanks' was 15 ppb for pumps and 495 ppb for tanks; in contrast, the estimate for 'Handled/pumped dispersant' was 77.3 ppb (Table 3 and Supplementary Table S6, available at *Annals of Work Exposures and Health* online) and that for dis/connects was 85 ppb. Working on a research vessel taking water samples resulted in a predicted air concentration of 25.4 ppb. Downwind 2-BE air concentrations when in an area with dispersant on the water were estimated to be as high as 11 700 ppb at 3 m, 710 ppb at ~9 m, and 30 ppb at 50 m under possible, but extremely unlikely worst-case conditions (see Stenzel, Arnold *et al.*, 2021 for more information on this scenario). The corresponding PG estimates were 0.11 times the 2-BE estimates. Thus, corresponding air concentrations for PG were all <10 ppb except for maintaining tanks (53 ppb) and being in an area recently sprayed at 3 m (1240 and 35 ppb). See Stenzel, Arnold *et al.* (2021) for further information on the concentrations in recently sprayed areas.

Oil mist

The EGs with the highest assigned estimated oil mist exposures were 'Deconned vessels/land', 'Deconned other equipment/land', 'Deconned booms/land', and 'Deconned All/Land' in all states and for all time periods

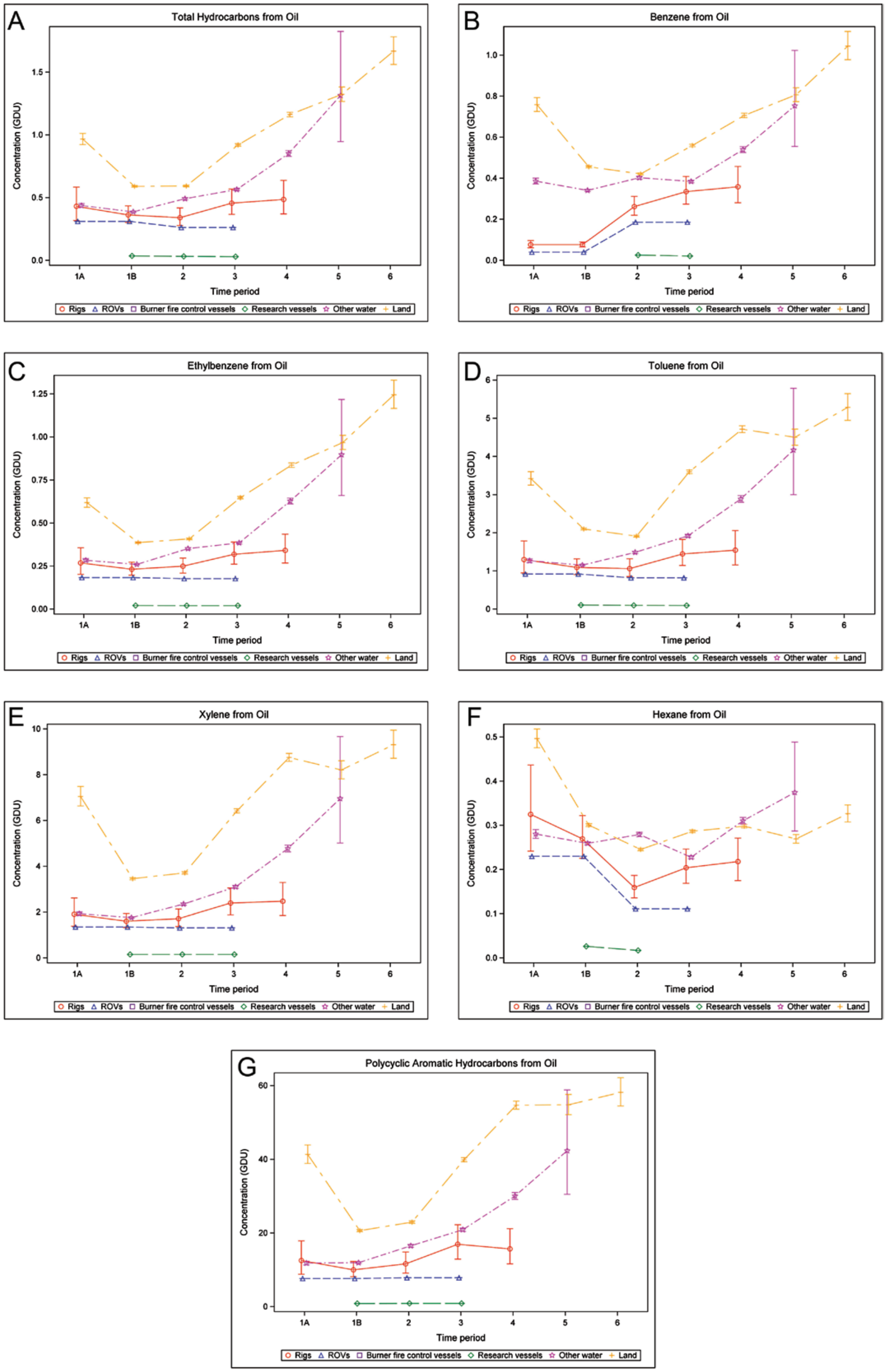

Figure 2. Dermal exposure estimates for total hydrocarbons, benzene, ethylbenzene, toluene, xylene and *n*-hexane from oil by time period (Gulf Dream Units, GDUs) for 'All Rigs', 'All ROVs', 'All Research Vessels', 'All Water Operations', and 'All Land Operations'. ROV vessels: vessels that piloted remotely operated vehicles (ROVs). These vessels left the area after TP3, so no estimates were developed. Burner fire control vessels were present only in TP1b and had no dermal exposures to any oil-related substance.

when deconning was being done (Supplementary Table S7, available at *Annals of Work Exposures and Health* online). About 15–20% of the EGs were categorized as high, as medium, as low, and as very low. The remaining one-third of the EGs were estimated to have had no oil mist exposure.

Dermal exposures

Estimates were developed for components of oil, tar, and dispersants (Supplementary Table S8a and b, available at *Annals of Work Exposures and Health* online) (Stewart *et al.*, 2021). The patterns for dermal exposures by broad groups over time differed from those seen for inhalation for the oil-related substances. THC estimates from oil exposure ranged from AMs of <0.02 GDUs to 5.5 GDUs (not shown). Equivalent tar AMs were <0.02–142.14 GDUs. The AMs for the THC dermal estimates across study participants associated with the same broad groups of EGs that were presented for inhalation are displayed in Fig. 2. Significant differences were seen primarily among the other water and land workers across several time periods. For example, the oil AMs increased over time for both other water workers and land workers, after TP3, with the AMs for land workers generally being significantly higher than the those for other water workers, which were significantly higher than those of rig and ROV workers within each time period. Patterns for tar were similar to those of oil.

Minimum AM values across study participants for the BTEX-H chemicals in oil was <0.01 GDUs (not shown). Benzene AMs for oil and for tar across study participants reached 12.77 and 3.69 GDUs, respectively; for ethylbenzene, the respective maximum AMs were 12.17 GDUs and 11.65 GDUs. Maximum AMs for toluene were 17.45 GDUs; for tar, the values rose to 42.37 GDUs. For xylene, the respective values were 36.77 GDUs and 88.18 GDUs. For n-hexane, the maximum AM for oil was 2.22 GDUs and for tar 5.56 GDUs. The maximum PAHs AMs were 219.31 GDUs for oil and 587.98 GDUs for tar. The patterns seen for each of these chemicals across the broad groups of study participants were similar to the pattern seen for THC (Fig. 2 and Supplementary Tables S8a and b, available at *Annals of Work Exposures and Health* online).

The two components evaluated for dispersants were THC and xylene, both to represent petroleum distillates, hydrotreated light in 9500A. The AM of the estimates for participants on land for THC were 25.39 and 16.74 GDUs and for xylene, 0.21 and 0.29 GDUs in TP1a and TP1b, respectively, the only time periods in which dispersants were used. For participants on the water the AMs were for THC, 25.33 and 56.88 and for xylene, 0.44 and 0.99 GDUs, respectively. Workers on the rig vessels, the ROV vessels and the burner fire control vessels were not considered exposed to dispersants.

Our evaluation of the GuLF DREAM model for hand exposure using previously published data on heavy fuel oil and asphalt found moderate correlation ($R^2 = 0.59$) for hands between measured and modeled estimates (Gorman Ng *et al.*, 2021). Insufficient measurements were available to analyze other body parts.

Discussion

The *Deepwater Horizon* explosion resulted in tens of thousands of workers being exposed to chemicals at potentially harmful levels. Questionnaires administered to GuLF Study participants collected information on jobs-activities-tasks performed during the 14-month study period of the *Deepwater Horizon* response and clean-up efforts. The responses were linked to inhalation exposures estimates using EGs and JEMs. Questionnaire responses were linked directly to study participants for modeling of dermal exposures.

An extensive air sampling database with measurements collected by the RP for oil-related substances (THC and BTEX-H) (Huynh *et al.*, 2021a, b, c; Groth, Banerjee *et al.*, 2021; Groth, Huynh *et al.*, 2021; Ramachandran *et al.*, 2021; Stenzel, Groth, Banerjee *et al.*, 2021) was used to estimate exposures to THC and BTEX-H. Low exposures to these substances were generally found when compared to their occupational exposure limits (Huynh *et al.*, 2021a, b, c; Ramachandran *et al.*, 2021). In the absence of other air monitoring data, we used recognized mathematical/deterministic models to estimate air concentrations to other airborne spill-related substances of interest [$PM_{2.5}$ (Pratt *et al.*, 2021), and dispersant aerosols (Arnold *et al.*, 2021) and vapor (Stenzel, Arnold *et al.*, 2021)]. Potentially high levels were estimated for $PM_{2.5}$ that could have exceeded the general population 24-h standards. Dispersant aerosol levels were generally low, but dispersant vapor air concentration levels could have exceeded the occupational exposure limits under certain circumstances. Airborne oil mist estimates were not quantitative, and the dermal estimates could not be related to concentrations on the skin, but the latter resulted in a wide range of exposures (Stewart *et al.*, 2021). Despite limitations, the level of data available provided the opportunity to develop detailed exposure metrics that far exceed any other efforts on oil spills to date.

Work histories

The large number of employers made collecting work histories from company records infeasible. Exposure and work histories therefore were obtained from the study participants via a telephone interview and for a subset, an in-home interview. In addition, monitoring data generally did not identify useful job titles. Because, however, the monitoring data identified activities and tasks being performed, we asked about activities-tasks in the questionnaires. To ensure we had not missed any important job-activities-tasks, we also asked open-ended questions on job-activities-tasks and reviewed the extensive documentation on the event. Relying on study participants to provide information on job titles-activities-tasks should not be a major source of error. Teschke *et al.* (2002) showed that study subjects have an accuracy rate of 70–90% for reporting employer, job classification, person-years in a job, and start and termination dates. Because we asked about specific activities 1–3 years after the event, recall bias or error should be small. The number of reported activities per study participant, however, was not small (median = 6), which may have increased reporting error. In addition, study participants may not have understood the terminology or the intent of some of the interview questions.

Exposure groups

We were unable to observe most of the jobs-activities-tasks. By the time the exposure assessment began, almost all jobs-activities-tasks had been completed. Given that there were thousands of vessels and tens if not hundreds of land work sites across four states with different employers, it is unlikely that uniform work practices were followed consistently across all work sites. Reviews of both the extensive measurement data and the many reports available on the disaster allowed us to identify the key exposure determinants of job-activity-task, location, and time. The determinants appeared to be appropriate, at least in part, as we found many notable differences in the AMs across the EGs formed from these determinants. That is not to say there is not error in our EG estimates. Many of the EGs had high GSDs, which could signal misclassification, although the high GSDs could also be due to the dynamic, non-routine, and time-dependent work being performed outdoors. Even if multiple distributions were the cause of the high GSDs, the level of bias and imprecision associated with our Bayesian methods should not have been affected because we considered the presence of mixed distributions in the computer simulations when estimating bias and imprecision (Huynh *et al.*, 2016).

Another error source, however, could be that other important determinants for some EGs were not identified because of the need to keep the questionnaire of manageable length. The occupational component alone covered 82 activities, with each activity often having 4–5 additional questions on time spent on the activity. Administration of this component took an average of 20 min of questioning out of a total average interview time of one hour. Extending the interview further to obtain more determinant information would have added an undue burden onto the study participants and likely would have resulted in some participants terminating the interview prematurely, which would have presented difficulties regarding other components of the questionnaire. Moreover, for many specific activities, we did not ask about the geographic location where individuals worked. Workers typically moved around even within a day, although more so on water than on land, which would have made location difficult for study participants to report. Also few records of the participants' locations were available. Yet, we found substantial differences in air concentrations across locations in the Gulf. To compensate for this lack of information, we included two questions in the questionnaire to identify the general location in the Gulf waters where participants spent their time [i.e. <10 nmi (nautical miles) of the wellhead, i.e. the hot zone/source and <3 nmi of the shoreline (near shore)]. Offshore was the default area if responses to both questions was 'no'. This information was used to supplement the estimates from participants' activities-tasks by our considering these locations as additional EGs, which raised or lowered participants' overall exposure, depending on the area. The location of work on land (i.e. state) was assumed to be the same state as where the participants resided. These procedures allowed some discrimination among the exposure levels for the same activity.

THC and BTEX-H

We relied on data collected by the RP as it was the largest collection of measurements and had the greatest coverage of activities performed by the study participants. Reliance on the RP's data should not be a major limitation as it has been shown that even experienced industrial hygienists often cannot accurately identify high or low exposed workers by observation (Arnold *et al.*, 2015). That is, if industrial hygienists cannot accurately identify high or low exposed workers, it would be unlikely that the sampling strategy could be biased to a particular outcome. We compared the RP's VOCs area data on two of the rig vessels with the THC personal data and found a correlation of determination (R^2) of 0.73 (95% CI, 0.60, 0.82), similar to what we found for the for the VOCs:THC data on the ROV vessels ($R^2 = 0.61$) (Ramachandran *et al.*, 2021). In addition, the R^2 of 0.61 was associated with a slope of 1.01 (95% confidence intervals: 0.88, 1.15) for TP1a and TP1b measurements, suggesting little overall bias.

One strength is that the recalculation of the measurement data to reflect the true analytic LOD from the reported LOD greatly enhanced our ability to develop exposure estimates, increasing the percent of non-censored data from 7% to 40%. In particular, the percent of THC measurements rose to 89%, allowing us to use information from these THC measurements to inform the Bayesian statistics when estimating exposures to BTEX-H.

In addition, the Bayesian method we used to account for censored data has <15% relative average bias and <65% relative rMSE. The bias and imprecision were likely even lower for BTEX-H, as the Bayesian estimates for those chemicals were strengthened further from the correlations between THC and each of the BTEX-H chemicals. These methods resulted in 71% of the EGs with AMs that achieved the bias and rMSE level of performance (Stenzel, Groth, Huynh *et al.*, 2021), with 15% of the estimates derived from our high censoring Bayesian method and 14% from the order-based statistical or substitution methods. Many of the remaining EG estimates were very low (<LOD). In the absence of sufficient measurement data we borrowed information from the nearest appropriate neighbor with available data to develop estimates from measurements with the most similar conditions. We assigned a confidence to each estimate to allow sensitivity analyses that exclude study participants with more uncertain estimates in the epidemiologic analyses.

From our data, various exposure metrics for study participants can be calculated for THC and BTEX-H, including maximum, cumulative and average exposure estimates and from the 95% percentile, full-shift peak exposures, accounting for variability using the GSDs or the 95% CIs. Thus, the estimates generated for the epidemiologic analyses will support a variety of analytic approaches to explore relationships between exposure and health outcomes of interest. For example, investigators may consider maximum exposures or average exposures within a time period or for a minimum number of days worked.

$PM_{2.5}$

Our goal in estimating $PM_{2.5}$ air concentrations was to provide possible concentration estimates for consideration when evaluating mitigation options in future oil spills. As such, we found levels of $PM_{2.5}$ may have exceeded the 24-h general population US National Ambient Air Quality Standards for some workers on some days (Pratt *et al.*, 2021).

The $PM_{2.5}$ estimates relied on AERMOD and the quality of the model input data. Although taken from published results and data compiled during the *DWH in situ* burns and flares, emissions traditionally are difficult to estimate (Pratt *et al.*, 2012). We also were unable to account for emissions from vessel exhaust or exhaust from land equipment because we had no information on where in the Gulf most vessels were located or on the numbers and locations on land of mechanical equipment with combustion engines. Engine emissions were as high, or higher, than the $PM_{2.5}$ estimates from burning and flaring developed for workers on land, near shore and offshore. For this reason, engine emissions should be considered a possible confounder or analyses should generally be limited to (higher exposed) participants performing activities in the specific areas where *in situ* burning and flaring were carried out and where engine emissions likely contributed less to the overall $PM_{2.5}$ air concentrations.

Another source of error was the skewness of the estimates (i.e. most air concentrations across the Gulf were estimated to have been very low), so that calculating average air concentrations would have been non-informative. We therefore calculated maximum estimates. Because of the large area of the Gulf with minimal estimated $PM_{2.5}$ air concentrations (i.e. <1 µg m^{-3}), no single individual was likely to have been exposed to his/her assigned concentration level every day, and on no day were all individuals likely to have experienced the levels we estimated.

Strengths include the use of a recognized model and model assumptions and input data that were taken from measurements in comparable studies or from values reported or estimated from videos and photographs of the actual *DWH* burns. We present potential air concentrations levels across large areas of the Gulf from *in situ* burning and flaring separately as well as from the combined effect of both sources (Pratt *et al.*, 2021). Finally, this is the first study to estimate $PM_{2.5}$ air concentrations from *in situ* burning and flaring operations. Nonetheless, because of the limitations of the estimates, disease risk findings from these data should be interpreted with caution.

Dispersants

AgDisp was used to predict total aerosols generated from spraying of dispersants on the water surface by plane or by vessel (Arnold *et al.*, 2021). Total aerosol levels from spray drift were in the µg m^{-3} range. There are no occupational exposure limits for total dispersant aerosols. One limitation of the aerosol estimation process is that we had limited information for some of the input data of the model and resorted to using default values of the model. The model also developed point estimates rather than distributions of air concentrations. Previously published comparison with real data (albeit of pesticide air concentrations on land, Bird *et al.*, 2002) and our comparison with results from AERMOD of 2 sorties suggested good agreement.

Vapor estimates also were developed from recognized models (Nicas, 2009; Armstrong, 2009). Although we did not observe the activities and had no information on the working conditions where vapor exposure could have occurred, we based our input data on expected conditions, given our knowledge of the activities and used expected ranges of input values, along with Monte Carlo methods, to provide estimates of uncertainty (Stenzel, Arnold *et al.*, 2021). Estimates of air concentrations under expected conditions were substantially below the relevant occupational limits for most situations.

Dermal exposures

Due to the lack of dermal exposure measurements, we modified a previously published model (van Wendel de Joode *et al.*, 2003) to fit our needs (Gorman Ng *et al.*, 2021). The dermal exposure estimates developed here are likely to be the estimates with the greatest error in the GuLF Study for several reasons. First, the model output is in GDUs, which is not directly relatable to actual skin exposure. Second, the high correlation (Stewart *et al.*, 2021) among many of the THC and BTEX-H substances may make it difficult to identify the putative agent associated with any risk estimate obtained from an epidemiologic analysis. The model did, however, allow us to rank study participants for each substance evaluated, so it may be possible to relate an adverse health effect with dermal exposure to oil-related chemicals in general, given the large contrast of exposure levels among study participants. Third, while our comparison with data in previously published studies suggested reasonable hand correlations, we were unable to evaluate other body parts. Fourth, there were over 90 input variables, some reported by the study participants, some entered by the study industrial hygienists. It may have been difficult for workers to accurately respond to questions on the frequency of contact with nine body parts and to differentiate between oil and tar. Industrial hygienists provided the information that participants could not, albeit without observing the operations. Instead, the hygienists reviewed the substantial documentation on the spill operations and had access to hundreds of photographs taken during the response and clean-up operations. Thus, although there are several potential sources of error, the magnitude and direction is unclear.

One major source of error was reduced, however, in that results were individual-specific, rather than JEM, values. Additional strengths include the use of responses from similar participants' (performing the same activity in the same state in the same time period) to impute missing data and the use of a model that had relatively good agreement with measurement data (at least for hands). We found different trends for dermal and inhalation exposure, as has been observed by others (Vermeulen *et al.*, 2002). Moreover, our study is the first to develop dermal exposure estimates resulting from an oil spill. Epidemiologic analyses based on exposure categories (e.g. low, medium, and high) rather than the actual GDUs due to the uncertainty in the estimates should help minimize error.

General

We did not evaluate within- and between-worker variability for our exposure measures several reasons. First, information on workers' names and other personal identifying information was not consistently documented in the measurement database. Second, anecdotal information indicated that families of workers were often hired, and sometimes, the same name was held by different people. Also, the person identified in the monitoring database may have been the member of a family or group of workers who spoke English or had specific personal identifying information, such as a Social Security number rather than the person actually monitored. Third, the median number of activities reported was 6; thus, workers may have been performing multiple activities on the same day (e.g. deploying boom, inspecting boom, retrieving boom), making it difficult to identify unique individuals to include in a within- and between-worker analysis. In any case, we expected high variability of exposures because of the critical and dynamic nature of the event, the multiple employers and lack of standard procedures, and the *ad hoc* and outdoor nature of much of the work, at least initially.

As with all JEMs, there is expected error among the inhalation exposures assigned to the study participants because in a JEM the same value is assigned to all members of the EG. Epidemiologic analyses could incorporate variability information on the CIs or the GSDs in the analysis to evaluate the effect of this error. In addition, for THC and BTEX-H we provided a relative level of confidence. This allows exclusion of study participants with low confidence in sensitivity analyses.

Overall strengths of the exposure assessment work include the estimation of inhalation exposure to 10 substances and of dermal exposure to 8 substances for a wide range of activities never before evaluated in an oil spill epidemiologic study. We applied the same determinants across all EGs and considered the same basic set of EGs across all assessments and found notable differences in the oil-related AMs across EGs. For THC and BTEX-H, we had a large number of measurements on which to base our estimates. We used a Bayesian method with relatively low bias and imprecision. From our data, various exposure metrics for study participants can be calculated for THC and BTEX-H, including maximum, cumulative, and average exposure estimates. Thus, the estimates generated for the epidemiologic analyses will support a variety of analytic approaches to explore relationships between exposure and health outcomes of interest. The exposure assessment component was one of the study's major focuses. For this reason, a substantial component of the interview was devoted to occupational exposures and considerable resources were devoted to developing exposure estimates as precise as the available data allowed.

Conclusions

Estimates of inhalation exposures were developed for total hydrocarbons and benzene, toluene, ethylbenzene, xylene and hexane, as well as for air concentrations of $PM_{2.5}$, dispersant aerosols and vapors, and oil mist. These estimates were linked to the study subjects via responses to a telephone questionnaire through exposure groups in a JEM. Other than to $PM_{2.5}$, exposures were generally low compared to occupational limits. Dermal exposure estimates suggested a wide range of exposures. The detailed exposure estimates allow for a variety of analytic approaches to explore relationships between specific and combined exposures and adverse health effects resulting from the oil spill clean-up and response.

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 IIS-1562303, NSF DMS-1916349. Dr Tran Huynh was also supported, in part, by CDC/NIOSH award K01OH011191 award.

Acknowledgements

We thank Wendy McDowell and Kaitlyn Rousch of McDowell Safety and Health Services, Inc. and Matthew Curry, Braxton Jackson, John McGrath and Anna Jones of Social and Scientific Systems, Inc. for the tremendous help they provided on this study. We thank Harrison Quick for his work in applying his downweighting methodology in the statistical methods. We also thank the workers for their participation in this study. We also thank Lakes Environmental Software for providing a complimentary copy of their AERMOD View software for use in this study. We thank BP for providing the data for these analyses.

Conflict of Interest

Prof Cherrie is currently undertaking consulting work related to the *Deepwater Horizon* disaster. All of his involvement with this paper was prior to any potential conflict of interest arising.

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 at https://gulfstudy.nih.gov.

References

Armstrong TW. (2009) Air dispersion plume models. In Keil CESCB, Anthony TR, editors. *Mathematical models for estimating occupational exposures to chemicals*. Fairfax, VA: AIHA Press.

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.

Arnold S, Stenzel M, Drolet D *et al*. (2015) Using checklists and algorithms to improve qualitative exposure assessment accuracy. *J Occ Environ Hyg*; **13**: 159–168.

Bird SL, Perry SG, Ray SL *et al*. (2002) Evaluation of the AgDISP aerial spray algorithms in the AgDRIFT model. *Environ Toxicol Chem*; **21**: 672–81.

Cavallari JM, Osborn LV, Snawder JE *et al*. (2012) Predictors of dermal exposures to polycyclic aromatic compounds among hot-mix asphalt paving workers. *Ann Occup Hyg*; **56**: 125–37.

Christopher Y, Van Tongeren M, Urbanus J *et al*. (2011) An assessment of dermal exposure to heavy fuel oil (HFO) in occupational settings. *Ann Occup Hyg*; **55**: 319–28.

Gorman Ng M, Cherrie JW, Sleeuwenhoek 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*; **81**: 76–86.

Groth C, Banerjee S, Ramachandran G *et al*. (2018) Multivariate left-censored Bayesian Modeling for predicting exposure using multiple chemical predictors. *Environmetrics*; **29**: 1–16.

Groth CP, Banerjee S, Ramachandran G *et al*. (2021) Methods for the analysis of 26 million VOC area measurements during the *Deepwater Horizon* oil spill clean-up. *Ann Work Expo Health*; **66**: i140–i155.

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.

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 approach for handling left-censored data. *Ann Occup Hyg*; **60**: 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.

King BS, Gibbins JD. (2011) Health hazard rvaluation of Deepwater Horizon response workers. Health Hazard Evaluation Report HETA 2010-0115 and 2010-0192-3138 (2011). Available at http://www.cdc.gov/niosh/hhe/reports/pdfs/2010-0115-0129-3138.pdf. Accessed 19 September 2018.

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.

Laffon B, Pásaro E, Valdiglesias V. (2016) Effects of exposure to oil spills on human health: updated review. *J Tox Environ Health, Part B*; **19**:3–4, 105–128.

Nicas M. (2009). The Near field/far field (two-box) model with a constant contamination emission rate. In Keil CESCB, Anthony TR, editors. *Mathematical models for estimating occupational exposures to chemicals*. Fairfax, VA: AIHA Press.

NIOSH. (2011) NIOSH Deepwater Horizon Roster Summary Report. Available at https://www.cdc.gov/niosh/updates/upd-12-19-11.html. Accessed 16 June 2018.

Plummer M. (2003). JAGS: a program for analysis of Bayesian graphical models using Gibbs sampling. In Proceedings of the 3rd International Workshop on Distributed Statistical Computing (DSC 2003), March 20–22, Vienna, Austria.

Pratt GC, Dymond M, Ellickson K *et al*. (2012) Validation of a novel air toxic risk model with air monitoring. *Risk Anal*; **32**: 96–112.

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.

R Core Team. *R: a language and environment for statistical computing*. Vienna, Austria: R Foundation for Statistical Computing. Available at http://www.R-project.org. Accessed 16 June 2015.

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, Banerjee S *et al*. (2021) Exposure assessment techniques applied to the highly censored *Deepwater Horizon* Gulf oil spill personal measurements. *Ann Work Expo Health*; **66**: i56–i70.

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 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.

Teschke K, Olshan AF, Daniels JL *et al*. (2002) Occupational exposure assessment in case–control studies: opportunities for improvement. *Occ Env Med*; **59**: 575–594.

U.S. Coast Guard. (2011) On Scene Coordinator Report Deepwater Horizon Oil Spill. Available at https://repository.library.noaa.gov/view/noaa/283. Accessed 11 November 2021.

US EPA 2017. 40 CFR 51 Appendix W, Guideline on Air Quality Models. U.S. Federal Register; 82: 5182–235. Available at https://www.ecfr.gov/current/title-40/part-51/appendix-Appendix%20W%20to%20Part%2051.

van Wendel de Joode B, Brouwer DH, Vermeulen R *et al*. (2003) DREAM: a method for semi-quantitative dermal exposure assessment. *Ann Occup Hyg*; **47**(1): 71–87.

Vermeulen R, Stewart P, Kromhout H. (2002) Dermal exposure assessment in occupational epidemiologic research. *Scand J Work Environ Health*; **28**: 371–85.