



# Low reproductive success rates of common bottlenose dolphins *Tursiops truncatus* in the northern Gulf of Mexico following the Deepwater Horizon disaster (2010–2015)

Nicholas M. Kellar<sup>1,\*</sup>, Todd R. Speakman<sup>2</sup>, Cynthia R. Smith<sup>3</sup>, Suzanne M. Lane<sup>2</sup>, Brian C. Balmer<sup>2,4</sup>, Marisa L. Trego<sup>1</sup>, Krista N. Catelani<sup>1</sup>, Michelle N. Robbins<sup>1</sup>, Camryn D. Allen<sup>1</sup>, Randall S. Wells<sup>4</sup>, Eric S. Zolman<sup>2</sup>, Teresa K. Rowles<sup>5</sup>, Lori H. Schwacke<sup>2</sup>

<sup>1</sup>Marine Mammal and Turtle Division, Southwest Fisheries Science Center, National Marine Fisheries Service, National Oceanic and Atmospheric Administration, 8901 La Jolla Shores Drive, La Jolla, CA 92037, USA

<sup>2</sup>Hollings Marine Laboratory, National Centers for Coastal Ocean Science, National Oceanic and Atmospheric Administration, 331 Fort Johnson Road, Charleston, SC 29412, USA

<sup>3</sup>National Marine Mammal Foundation, 2240 Shelter Island Drive, Suite 200, San Diego, CA 92106, USA

<sup>4</sup>Chicago Zoological Society, c/o Mote Marine Laboratory, 1600 Ken Thompson Parkway, Sarasota, FL 34236, USA

<sup>5</sup>Office of Protected Resources, National Marine Fisheries Service, National Oceanic and Atmospheric Administration, 1315 East West Highway, Silver Spring, MD 20910, USA

**ABSTRACT:** Following the *Deepwater Horizon* (DWH) oil spill, reproductive success rates in 2 northern Gulf of Mexico (GoM) bottlenose dolphin stocks exposed to oil were evaluated for 4 yr during and after the spill (2010 to 2015) in efforts to assess population-level reproductive health. Pregnancy was determined from either (1) ultrasound examinations of the reproductive tract during capture-release health assessments, or (2) endocrine evaluations of blubber tissue collected from dart biopsies of free-ranging dolphins. Follow-up photo-identification was then used to track the status of pregnant females and any associated neonatal calves for a minimum of 1 yr after the initial pregnancy detection (IPD). For all pregnant females observed following IPD, individuals seen with a calf (reproductive success) and without one (reproductive failure) were recorded. The resulting estimated reproductive success rates for both GoM stocks (19.4%; 7/36) were less than a third of those previously reported in other areas not impacted by the spill (i.e. Sarasota Bay, FL; Indian River Lagoon, FL; and Charleston Harbor, SC) using similar techniques (64.7%; 22/34). We also evaluated the relationships between reproductive success and 13 potential covariates, including stock, ordinal date, progesterone, cortisol, thyroid hormone concentrations, leukocyte count, lung health score, and total body length. Among these, the results only provide strong evidence (Bayes factor >20) of a relationship between reproductive failure and the total leukocyte count covariate. The high reproductive failure rates measured in both GoM stocks following the DWH oil spill are consistent with mammalian literature that shows a link between petroleum exposure and reproductive abnormalities and failures.

**KEY WORDS:** Oil spill · Petroleum · Pregnancy · Reproductive failure · Cetacean · Blubber progesterone · Ultrasound

## INTRODUCTION

The massive volume of oil released after the *Deepwater Horizon* (DWH) drilling rig explosion spread

throughout the northern Gulf of Mexico (GoM), including into the habitats of bay, sound, and estuary (BSE) common bottlenose dolphins *Tursiops truncatus* (Michel et al. 2013, Schwacke et al. 2014, Balmer

\*Corresponding author: nick.kellar@noaa.gov

et al. 2015). Various assessment studies have shown that the BSE dolphin stocks most exposed to DWH oiling oil exhibited substantially higher rates of mortality, disease, and reproductive failure. For the majority of time in the 4 yr after the drilling rig explosion, dead-stranding rates in areas with the heaviest of oiling exceeded the upper 95% confidence intervals of the historical baseline levels; in some years, stranding rates in certain regions were up to 3.5 to 4 times greater than the 95% confidence levels (Litz et al. 2014, Venn-Watson et al. 2015b), leading to an unprecedented duration and magnitude in rate of dead dolphin stranding. Moreover, mortality rates in heavily-oiled Barataria Bay, LA, and Mississippi Sound, MS, were approximately 2.5 to 3.5 times higher than those of non-oiled areas, as estimated from the boat-based follow-up surveys of animals examined during capture-release health assessments (CRHA) (Lane et al. 2015, Deepwater Horizon Natural Resource Damage Assessment Trustees 2016). The CRHAs of the Barataria Bay and Mississippi Bay animals also revealed a disproportionate number of grossly underweight individuals (Schwacke et al. 2014) and a high prevalence of lung disease and evidence of adrenal insufficiency that continued until at least 2014 (Schwacke et al. 2014, Smith et al. 2017, this Theme Section). The results from these live animals are consistent with histological studies of dolphins stranded after the spill in the oiled waters of Louisiana, Mississippi, and Alabama, which found significantly higher rates of bacterial pneumonia and abnormally thin adrenal cortices relative to animals outside the region or with individuals collected before the spill (Venn-Watson et al. 2015a). Barataria Bay dolphins also showed reproductive failure rates in the first years after the spill 4 times greater than those from a non-oiled stock (Lane et al. 2015). The stranding record for some areas also shows indications of impaired reproduction/early life survival, with periods of excessively high rates of perinatal strandings (Carmichael et al. 2012, Litz et al. 2014, Venn-Watson et al. 2015b).

The potential impacts of oil exposure on reproduction in wildlife and fish species are well documented within the scientific literature; the record shows marine vertebrate reproduction and early development becoming impaired in the wake of large oil spills. After the *Exxon Valdez* disaster, sea otters in Prince William Sound experienced high rates of fetal and neonatal loss (Tuomi & Williams 1995). In seabird colonies, successful annual reproduction was still 70% lower than historical rates, 5 yr after the *Prestige* oil spill (Barros et al. 2014). Reproductive and early

developmental effects have also been documented in fish species exposed to oil from the *Exxon Valdez*, *Prestige*, and *Deepwater Horizon* spills (Hawkes & Stehr 1982, Kocan et al. 1996, Bilbao et al. 2010, de Soysa et al. 2012, Kawaguchi et al. 2012, Whitehead et al. 2012, Turner et al. 2014).

One effective way to assess cetacean reproductive health at a population level is by measuring rates of reproductive success. In this case, a 'reproductive success' is specifically defined as when an identified pregnant female produces a viable calf (i.e. a calf that survives for sufficient time to be observed, recorded, and/or photographed) (Wells et al. 2005, 2014, Browning et al. 2010, Wells 2014). Conversely, a failure is when an identified pregnant female is observed without a calf for a duration of time past her expected due date<sup>1</sup>. Reproductive success rates have been estimated for free-ranging BSE bottlenose dolphin stocks in previous studies. In Sarasota Bay, FL, 83% (10/12) of females determined pregnant via ultrasound successfully gave birth to calves that survived for sufficient time to be recorded (Wells et al. 2014). In addition, reproductive successes and failures were reported for BSE bottlenose dolphins near Charleston, SC, and from the Indian River Lagoon, FL; on aggregate 50% (10/20) of the pregnancies (in this case determined via serum progesterone concentrations) met the criteria of reproductive success, i.e. resulted in viable calves (Bergfelt et al. 2013).

Pregnancy was diagnosed 2 different ways in these previous studies: using ultrasound imaging (Wells et al. 2014) and using endocrine evaluations via progesterone measurements (Bergfelt et al. 2013). These 2 methods are the most common modes of pregnancy detection, and their interpretation can affect the numerical estimates of reproductive success rate. Diagnostic rises in progesterone occur before the stage of pregnancy at which a fetus can be distinctly visualized via ultrasound (O'Brien & Robeck 2012). However, high progesterone levels can also be associated with conditions other than pregnancies, such as non-fertile ovulations and pseudo-pregnancies

<sup>1</sup>An expectant mother's specific sighting history and duration is important when assessing reproductive success. If she is seen with a neonate calf even before her estimated due date and not seen again after her due date that would be considered a success. Given the same sighting history of an expectant mother who is not seen with a calf, we would be unable to determine if the pregnancy was successful or not. In order to determine that the pregnancy is a failure the expectant mother (who has not been previously seen with a calf) must be observed and recorded as being not with a calf between 2 wk and 1 yr after her due date.

(Robeck et al. 2001, Bergfelt et al. 2011, O'Brien & Robeck 2012). As such, analysis that limits data to ultrasound detections underestimates pregnancies and consequently may underestimate failures because one must wait longer into gestation before pregnancy determination is possible, while potential mortality is ongoing. However, analysis that relies on endocrine evaluations (in serum and almost certainly in blubber), may overestimate pregnancies and consequently overestimate failures as some high progesterone concentrations may not be related to true pregnancies.

In efforts to assess the impact of the oil spill on reproduction of exposed dolphin stocks, we examined the reproductive success rates of GoM dolphin stocks in Barataria Bay, LA, and Mississippi Sound, MS, that were oiled in the years during and following the DWH disaster (2010 to 2015). These success rates were then compared to those published previously for Sarasota Bay, FL, Charleston Harbor, SC, and Indian River Lagoon, FL. In addition, covariate analyses were conducted to evaluate potential links between reproductive success and biological, demographic, and behavioral information.

## MATERIALS AND METHODS

### Overview

The overall intent in this study was to identify and monitor over time pregnant dolphins to assess whether each identified pregnancy resulted in a viable calf. Pregnancy was primarily diagnosed either from (1) sonographic images collected during CRHA operations or from (2) endocrine evaluations of blubber dart biopsies collected during boat surveys. Identified pregnant individuals were monitored via follow-up boat-based surveys and confirmed through photo ID.

### Ultrasound pregnancy evaluations and capture-release health assessments

Capture-release health assessments were conducted on 4 different occasions in 2 locations, Barataria Bay, LA (August 2011, June 2013, and June 2014) and Mississippi Sound, MS (July 2013) (Fig. 1), using previously described methods (Wells & Scott 1990, Wells et al. 2004, 2005, Schwacke et al. 2014). Briefly, individuals or small groups of dolphins were encircled in a seine net and then restrained for evaluation and sampling. Ultrasound images were collected during

these capture events to assess reproductive state as previously described (Wells et al. 2014, their supplementary information document). In addition to ultrasound, a serum sample was collected from each dolphin and progesterone concentrations were measured at the Animal Health Diagnostic Center (Cornell University, Ithaca, NY). Pregnancies were classified as (1) confirmed pregnant: a fetus was detected; (2) probable pregnancy: a corpus luteum (CL) with or without uterine fluid was present in conjunction with a serum progesterone value greater than  $5 \text{ ng ml}^{-1}$ ; or (3) confirmed not pregnant: the ovary was seen without a CL, progesterone measurements were less than  $5 \text{ ng ml}^{-1}$  and the uterus was visualized without an embryo or fetus (Lane et al. 2015). 'Confirmed pregnancies' were given estimated due dates based on fetal biparietal skull diameter (Stone et al. 1999, Lacave et al. 2004, Smith et al. 2013). One sample classified as a 'probable pregnancy' was excluded from the statistical analyses given the unknown level of statistical uncertainty associated with this pregnancy classification (i.e. unknown rate associated with pregnancies, pseudo-pregnancies, or non-fertile ovulations).

During the CRHAs, measurements and samples were collected for a wide spectrum of diagnostic analyses including sonographic images of the lungs, measurements of mass and length, and blood samples for chemical, cell composition and endocrine analysis (Schwacke et al. 2009, 2010, 2011, Hart et al. 2013). Important for this study, blubber wedge biopsies were also collected from the vast majority of these animals using previously published techniques (Wells et al. 2004, 2005). Briefly, a veterinarian excised a small ( $\sim 4 \times 3 \text{ cm}$ ) wedge of epidermis and blubber tissue from a region approximately 10 cm below and 10 cm behind the caudal insertion of the dorsal fin. From this sample, blubber progesterone and blubber cortisol were measured using previously established methods (Kellar et al. 2006, 2013b, 2015, Pérez et al. 2011, Trana et al. 2016).

### Blubber endocrine evaluations of dart biopsies

A second type of blubber biopsy was obtained from free-swimming unrestrained dolphins (i.e. individuals not captured during CRHAs) via remote dart sampling with a clean sanitized 10-mm diameter stainless steel collection tip attached to a projectile dart (Balmer et al. 2015, Sinclair et al. 2015). These dart biopsies were primarily obtained from the area from the anterior insertion of the dorsal fin caudal to the

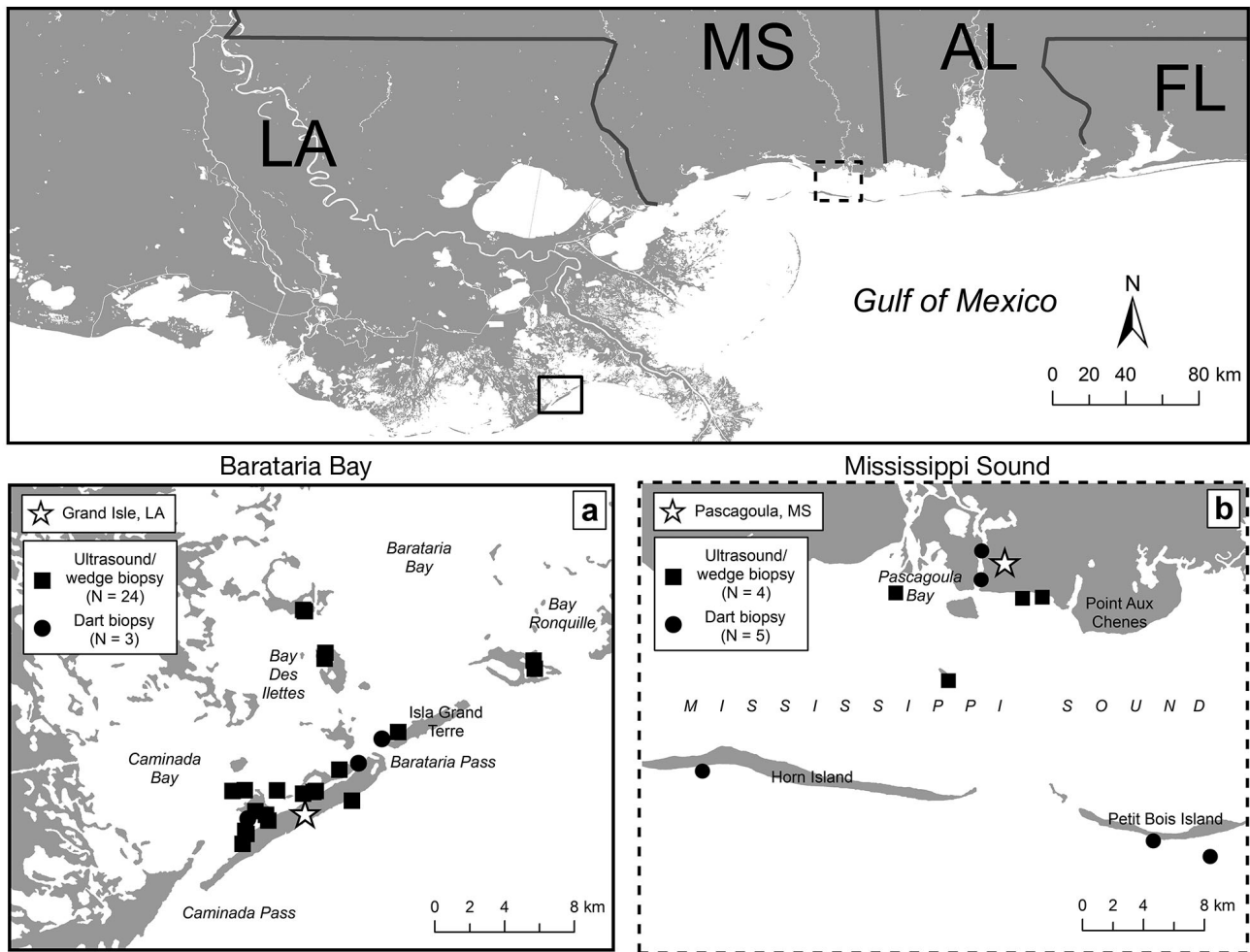


Fig. 1. Study areas (a) Barataria Bay, LA, and (b) Mississippi Sound, MS, for investigation of the effects of oiling on the reproductive success of common bottlenose dolphin *Tursiops truncatus* in the Gulf of Mexico (GoM) during and after the 2010 *Deepwater Horizon* (DWH) oil spill. The inset maps show locations of pregnant animals based on results of ultrasound examinations with wedge biopsies (■) and endocrine evaluations of blubber tissue collected from dart biopsies (●)

mid-point of the peduncle and from the dorsal ridge ventral to the mid-frontal plane. Samples were placed in fully charged liquid nitrogen dry shippers in the field, shipped in dry ice, and ultimately stored at  $-80^{\circ}\text{C}$  until processing. The sex of the sampled animal was assessed via genetic analysis of DNA obtained from the epidermis (Rosel 2003).

The blubber portion of each sample type (wedge or dart biopsy) was subsampled in pieces that were perpendicular to the skin so that the full depth of blubber was used and these subsamples contained between  $\sim 0.05$  and  $0.15$  mg of blubber tissue. Pregnancy determination was based on progesterone concentration, i.e. the mass of progesterone as a fraction of blubber mass ( $\text{ng g}^{-1}$ ), using techniques that have been validated for and implemented across numerous cetacean species including bottle-

nose dolphins (Mansour et al. 2002, Kellar et al. 2006, 2013a,b, 2014, Pérez et al. 2011, Trego et al. 2013). Blubber progesterone concentrations were determined as described previously (Trego et al. 2013). Briefly, the blubber was homogenized and then tissue debris and water were removed in a series of ethanol (100%), ethanol:acetone (4:1), and diethyl ether (100%) rinses in which the supernatant was recovered after each solvent rinse. The resulting lipid residue was mixed with acetonitrile and hexane (2 immiscible solutions) twice and, each time, the acetonitrile layer with the target hormone was collected. The final acetonitrile layer was dried and stored at  $-20^{\circ}\text{C}$  until the extract was ready to be assayed. For assaying, the extracts were suspended in  $250\ \mu\text{l}$  of 1 M phosphate buffered saline and the progesterone measurements were per-

formed using EIA kit ADI-900-011 (Enzo Life Sciences). The intra-assay coefficient of variation (CV) was between 4.9 and 7.6%, and the inter-assay CV was between 2.7 and 8.3%. All information associated with each blubber sample including reproductive state was kept blind from the laboratory measuring the blubber progesterone concentration.

To further validate the use of blubber progesterone for pregnancy determination, we present the blubber progesterone concentrations for all CRHA animals for which associated ultrasound imaging of the reproductive tract was conducted to provide pregnancy confirmation. The utility of this approach is that it can be used with dart biopsy sampling to obtain reproductive information without the need to capture individuals. For cetacean species, it has been estimated that progesterone concentrations of pregnant animals rise to levels that are diagnostically distinct from confirmed non-pregnant animals within 3 wk post conception (Kellar et al. 2006, O'Brien & Robeck 2012, Robeck et al. 2012, Steinman et al. 2016). However, unlike with the CRHA ultrasound animals, the endocrine evaluations of the blubber provide no diagnostic information about due date of the pregnant biopsied animals; therefore, the maximum expected due date was assumed to be no greater than 13 mo after the biopsy was collected, based on an estimated gestation duration of 12.5 mo, plus 2 wk (O'Brien & Robeck 2012, Smith et al. 2013, Wells et al. 2014).

Dart biopsies were collected from both DWH oil-impacted stocks, Barataria Bay and Mississippi Sound, in 2010 to 2012 along with photo identification of the biopsied individuals (Fig. 1). Boat-based follow-up surveys were then conducted in each area for at least 13 mo following the CRHAs or dart biopsy sampling events to monitor for the presence of calves. Following the approach previously described by Melancon et al. (2011), these boat-based monitoring surveys included photo-identification, VHF radio tracking, and additional remote biopsy collection and were conducted from one or two 5 to 6 m center-console outboard-powered vessels, crewed by 3 to 4 researchers. Canon EOS digital cameras equipped with 100 to 400 mm telephoto lenses were used to collect all identification images.

### Photo analysis

The analysis of the photographic images used for individual identification and calf association followed previously established methods (Melancon et al. 2011, Lane et al. 2015). Briefly, image file names were

encoded with survey and sighting numbers and then sorted in Adobe Photoshop 7.0, from which the best left and/or right-side dorsal fin image of each individual from each sighting was obtained. The resulting sorted images were graded for photographic quality based on focus, contrast, angle, dorsal fin visibility/obscurity, and proportion of the frame filled with diagnostic attributes (Urian et al. 1999, 2015). Images of individual dolphins were matched using a combination of 3 methods: freeze brand numbers on their dorsal fins, tag placement, and/or fin notches (Fig. 2). All matches were verified by 2 researchers and cataloged in FinBase (Adams et al. 2006), a customized database constructed in Microsoft Access 13, under each putative animal's unique numerical code.

Individual sighting histories were generated by compiling when each individual was observed during boat-based surveys. After the initial pregnancy detection (IPD) via ultrasound or progesterone concentration of remote biopsy samples, photographs of pregnant dolphins obtained during boat-based sightings were analyzed to determine the presence of a neonate or calf less than 1 yr old (Fig. 2). Reproductive success was defined as the sighting of an associated neonate/calf—i.e. a dolphin whose length was no greater than 75% of the presumed mother's length and swimming in echelon position (Barbara 1999)—with its expectant mother within the year after the mother's approximate due date. Reproductive failure (i.e. failed pregnancy or neonatal death)



Fig. 2. *Tursiops truncatus*. A mother bottlenose dolphin with calf. The mother was identified as dolphin 7034 (determined from visible characteristics including dorsal fin notches) from Mississippi Sound. This animal was dart biopsied in May 2010 and an endocrine evaluation of the sample indicated that she was pregnant at that time. This image was taken 10 mo later (March 2011), confirming reproductive success

was defined as the sighting of an expectant female without an associated neonate/calf either (1) between 2 wk and 1 yr following her estimated due date for ultrasound-visualized pregnancies or (2) up to 1 yr after her maximum due date for evaluated dart biopsy pregnancies. If an expectant female was not re-sighted after the 2 wk following the estimated due date (at least 13 mo after IPD for biopsy evaluated pregnancies) and was not previously observed with a neonate calf (i.e. a premature birth) the outcome was classified as 'could not be determined' and the individual was excluded from the analysis. Calf size, which was identified using photo-identification during the follow-up surveys, was also factored into each assessment of reproductive outcome to help differentiate calves conceived in different years; this can be important when there is a relatively long gap in an expectant mother sighting history<sup>2</sup>. Sightings from all survey types were used to monitor reproductive outcomes.

### Statistical analysis

#### Overall analysis design

There were 3 parts to the data analysis. First, we modeled the relationship between blubber progesterone and ultrasound-evaluated pregnancy state of female dolphins sampled during the CRHAs. Second, we compared the reproductive success rates as measured in this study in DWH-oiled dolphins from Barataria Bay and Mississippi Sound to those reported in the literature for non-oiled dolphins from Sarasota Bay, FL, Charleston Harbor, SC, and Indian River Lagoon, FL (Bergfelt et al. 2013, Wells et al. 2014). Third, we conducted 2 covariate analyses to evaluate

<sup>2</sup>In order to determine that the pregnancy is not a success an expectant mother (who is never observed with a calf) must be observed and recorded between 2 wk and 1 yr after her due date. Consequently, if an expectant mother is seen without a calf before her due date and then is not seen again until 13 mo after her due date (25.5 mo after her estimate conception date) and at that time is seen with a small neonate calf (size and characteristics of the calf are important here) the outcome of the assessed pregnancy could not be determined. In other words, although she did have a successful pregnancy it occurred from a different conception event in the year following her due date. Because that calf was not associated with the initial pregnancy detection it would not be included in the success analysis. This occurred only once for re-sighted pregnancy females; the overwhelming majority (>95%) of re-sighted animals irrespective of reproductive class were seen within 1 yr after sampling.

potential links between reproductive success and biological, demographic, and behavioral information.

#### Modeling the relationship between blubber progesterone and pregnancy state

The probability of pregnancy was modeled as a function of measured blubber progesterone (nanograms of progesterone per gram of blubber processed) using a standard logistic regression structure with pregnancy state as the Bernoulli response variable. These data were taken exclusively from CRHA females from which the reproductive tract was inspected via ultrasound, serum progesterone was measured, and a blubber wedge biopsy was also taken. The analysis was conducted within a Bayesian framework to allow for better propagation of the measured uncertainty into the subsequent analyses. All blubber progesterone measurements were log transformed prior to analysis to minimize the heteroscedasticity that is common in hormone measurements (Kellar et al. 2009, 2015). The output generated estimates of the parameters for the best logistic regression fit and the 95% Bayesian credibility envelope around the resulting function.

When the model was given the set of blubber progesterone concentrations from females of unknown pregnancy state (i.e. the concentrations from the dart biopsied females), it returned for each blubber progesterone concentration a marginal posterior probability distribution that the associated dart biopsy sample was obtained from a pregnant animal. In other words, given the model linking blubber progesterone to probability of being pregnant in the CRHA animals (i.e. females of known pregnancy status) we could then utilize each dart biopsy progesterone concentration to estimate the probability that the biopsy was collected from a pregnant animal. Estimating probability of being pregnant removed the constraint of choosing an arbitrary blubber progesterone concentration that, for diagnostic purposes, would distinguish pregnant from non-pregnant females; here the model makes that estimation with appropriate associated statistical uncertainty.

#### Comparing reproductive success rates (oiled vs. non-oiled reference stocks)

The difference in the reproductive success rate between the DWH-oiled stocks and the combined data from 3 non-oiled reference stocks was statistically evaluated using a delta mean test (Manly 1991,

Lo 1994). This test, which is similar to a frequentist paired *t*-test, was conducted in a Bayesian framework and it allowed us to evaluate the statistically relevant magnitude of this difference (i.e. by determining the probability that the difference in success rate was greater than a defined percentage). Ultimately, there were 36 data points from pregnant females representing the oiled stocks (27 from Barataria Bay and 9 from Mississippi Sound; Fig. 1) with sufficient sighting histories to be included in the reproductive success analysis.

Reproductive success data from the non-oiled reference stocks were primarily compiled from 2 peer-reviewed publications, one conducted in Sarasota Bay, FL that employed ultrasound scanning for pregnancy diagnosis (Wells et al. 2014,  $n = 12$ ), and another conducted in Charleston, SC, and Indian River Lagoon, FL that used serum progesterone concentrations with a threshold of  $6.0 \text{ ng ml}^{-1}$  to distinguish pregnancy from non-pregnancy (Bergfelt et al. 2013,  $n = 20$ ). The success data from the ultrasound and serum reference-studies were aggregated along with 2 additional reproductive outcome data points acquired for the present study. These 2 additional points were acquired via dart biopsy samples taken from females in Sarasota Bay. Out of 10 Sarasota Bay female dolphins with sufficient sighting histories screened for progesterone concentrations indicative of an active pregnancy, these 2 (identified in the Sarasota photo identification as GRWW and FRNG) had blubber progesterone concentrations consistent with >99% probability of being pregnant; both females were seen with calves (GRW2 and FRN2 respectively) within 11 mo after each dart biopsy was taken. Thus in total there were 34 total reference data points of reproductive outcome representing 3 non-oiled stocks. Also note that the primary focus of the Bergfelt et al. (2013) study was on the endocrine evaluations of pregnancy, using serum progesterone measurements as the 'gold standard' of pregnancy detection. This study was not necessarily designed to evaluate reproductive success; however, the data were acquired in such a way that reproductive success could be calculated from the results so as to provide additional reference data for future studies.

#### Covariate analysis

Supplemental data were gathered from the CRHA animals that it was not possible to obtain from the dart-biopsied dolphins. As such, the covariate analysis was separated into 2 phases. The first phase

looked at the covariates that were measured in both types of sampling effort (CRHA ultrasound and endocrine evaluations of dart biopsies), i.e. sighting frequency, season (day of the year), days after the DWH explosion, stock (Barataria Bay or Mississippi Sound), blubber progesterone concentration, and blubber cortisol concentration. The second phase examined covariate data that were solely obtained from the CRHA animals, i.e. total length, lung condition score (assessed via ultrasound), white blood cell count (WBC), and serum hormone concentrations (progesterone, cortisol,  $17\beta$  estradiol, and thyroxine).

Covariate associations with reproductive success were assessed as part of a Bayesian model fitting procedure within a generalized linear model framework, using a variable selection method for inferring which factors to include or exclude in the model (Manly 1991, Carlin & Chib 1995, Chib & Carlin 1999). In this analysis, sets of logistic generalized linear models were constructed in WinBUGS (Lunn et al. 2000, Spiegelhalter et al. 2003) with each of the 8 covariates multiplied by a covariate-specific Bernoulli selection parameter in the form:

$$\text{logit}(c_{\text{success}(i)}) = \alpha + \sum \gamma_j \times \beta_j (\mathbf{X}_{i,j}) \quad (1)$$

where  $c_{\text{success}(i)}$  is the reproductive outcome (success or failure) for individual  $i$ ,  $\alpha$  was the model intercept,  $\beta_j$  was the coefficient of covariate  $j$ ,  $\mathbf{X}$  a matrix of values for  $j$  and  $i$ , and  $\gamma_j$  was the selection parameter for each covariate with a probability of  $p(j)$ , equal to the mean of  $\gamma_j$ . A logit link function was used as reproductive success and coded as a Bernoulli response, either 0 or 1, for failure and success, respectively. Vague priors were set for all marginal slope coefficients from normal distributions, each with mean = 0 and variance = 1000 (prior distributions where variance exceeded 1000 did not converge). The prior on each selection parameter was set to be  $p = 0.5$ . The predictive weight (an indication of importance) that each covariate had on the estimate of the sampled animal's reproductive success was equal to  $p(j)$  and was directly proportional to the number of iterations where that particular covariate was selected in the Markov chain Monte Carlo (MCMC) estimates. Marginal posterior probabilities of the selection parameters with median values greater than  $p = 0.5$  were those in which the weight of evidence supported their inclusion in the model. The data for each of these covariates were normalized (mean = 0 and standard deviation = 1) prior to analysis. All 36 reproductive outcome data points from pregnant females representing the oiled stocks (27 Barataria Bay, 9 Mississippi Sound; Fig. 1) were used for this first covariate analysis.

For the second covariate analysis (i.e. only the CRHA animals), data from 4 Sarasota pregnant females (2 successes and 2 failures) were included to augment the covariate data set. These 4 females were described by Wells et al (2014), where they are catalogued as dolphins FB33, FB116, FB137, and FB225. Because data from the dart-biopsied animals ( $n = 8$ ) lacked blood, morphology, and lung information, they could not be used in this second covariate analysis. Once they were removed, too few reproductive success data points remained to inform the analysis. Consequently, the 4 additional pregnant females from Sarasota (2 successes and 2 failures), which were examined during CRHA events in 2010 ( $n = 2$ ) and 2013 ( $n = 2$ ) using the same methodology employed during the CRHAs of Barataria Bay and Mississippi Sound, were included. In total for this second covariate analysis there were 32 pregnancies (7 successes) represented (24 Barataria Bay, 4 Mississippi Sound, 4 Sarasota Bay).

#### Bayes factors

Bayesian results for hypothesis testing were reported as both Bayesian posterior probabilities and Bayes factors ( $B_{10}$ ); otherwise only posterior probability estimates were reported. We have included a table (Table 1) with description statements for the different levels of Bayes factors, adapted from Kass & Raftery (1995), to aid in interpretation.

## RESULTS

### Pregnancy determination

Ultrasound evaluations were conducted on 63 females (54 from Barataria Bay and 9 from Mississippi Sound). Of these, 33 were confirmed pregnant, 24 were confirmed not pregnant, and 6 showed evidence consistent with early pregnancy, but pregnancy status could not be confirmed (i.e. CL and uterine fluid were observed but fetus/embryo was not).

Out of the 57 females of known pregnancy state, blubber progesterone measurements from 54 females (no wedge blubber samples were taken from the other 3) were used to develop the logistic regression model to estimate the probability of being pregnant. The pregnant female blubber progesterone measurements (mean  $\pm$  SE =  $275.38 \pm 38.5$  ng g<sup>-1</sup>,  $n = 30$ ) were on average >2 magnitudes greater than non-pregnant females ( $1.16 \pm 0.56$  ng g<sup>-1</sup>,  $n = 24$ ) with no overlap in values between 20 and 40 ng g<sup>-1</sup>, providing perfect

Table 1. Bayes factor interpretation. Descriptive statements of standards of evidence in scientific investigation as proposed by Kass & Raftery (1995)

Bayes factor ( $B_{10}$ )	Evidence against $H_0$
1 to 3	Not worth a bare mention
3 to 20	Positive
20 to 150	Strong
>150	Very strong

sensitivity and specificity as a diagnostic of pregnancy within this dataset (i.e. 100% of the animals were correctly assigned to their known pregnancy state). The best fit model (Fig. 3) describing the relationship between pregnancy state and blubber progesterone (BP) concentrations with these animals was:

$$P_{\text{pregnancy}_i} = \frac{1}{1 + e^{-(\alpha_0 + \alpha_1 \times \text{BP}_{4_i})}} \quad (2)$$

where the median coefficients and their 95% credibility intervals were:  $\alpha_0 = -11.8$  (CI =  $-32.9$  to  $-0.010$ ) and  $\alpha_1 = 71.9$  (17.9 to 98.8).

For dart biopsies, ultrasound data were not available to inform pregnancy state. Instead, the above model, which accounts for the statistical uncertainty of the coefficient estimates, was used to generate a posterior probability distribution of the probability of pregnancy (see Table S1 in the Supplement at [www.int-res.com/articles/suppl/n033p143\\_supp.xlsx](http://www.int-res.com/articles/suppl/n033p143_supp.xlsx)) based

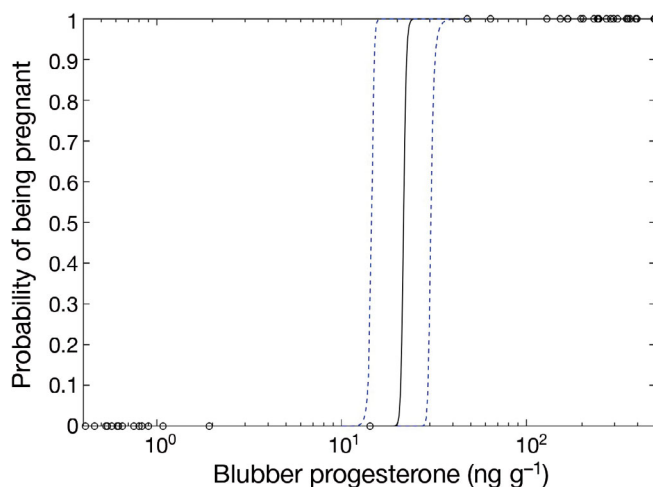


Fig. 3. *Tursiops truncatus*. Logistic model for the probability of pregnancy in bottlenose dolphins relative to blubber progesterone concentration. Open circles represent the observed measurement data. Dashed lines represent 95% credibility interval from the 10 000 model iterations. x-axis values are log<sub>10</sub> scaled. Model priors and summaries of the posterior probability distributions for the parameters are given in 'Results: Pregnancy determination'



Table 2. *Tursiops truncatus*. Reproductive success rates for oiled and non-oiled reference stocks of common bottlenose dolphins in the Gulf of Mexico (GoM) following the *Deepwater Horizon* (DWH) oil spill

Location	Years	Pregnancy determination		Aggregated total (%)
		Ultrasound imaging (successes/total)	Blubber endocrine evaluation (successes/total)	
Reference stocks (Sarasota Bay, FL; Charleston, SC; Indian River Lagoon, FL)	1998–2014	20/32	2/2	64.7% = 22/34
Oiled stocks				
Barataria Bay, LA	2010–2014	4/24	1/3	19.4%
Mississippi Sound, MS	2010–2014	1/4	1/5	= 7/36

on each sample's measured blubber progesterone concentration. Of the 85 females that were biopsied, 17 individuals were estimated to have a higher than 99.9% median probability of being pregnant (blubber progesterone: mean  $\pm$  SE = 161.0  $\pm$  21.4 ng g<sup>-1</sup>, n = 17). The model indicated 1 individual having between 0.1 and 99.9% pregnancy probability (blubber progesterone: 25.24 ng g<sup>-1</sup>); however, this individual was not included in the success analysis because there were insufficient sightings after pregnancy to assess success status. The rest of the biopsies came from females with a median probability of being pregnant of <0.1% (blubber progesterone: mean  $\pm$  SE = 1.55  $\pm$  0.42 ng g<sup>-1</sup>, n = 67). These, too, were not included in the reproductive success analysis because of the very low probability that they were pregnant at the time of sampling; none were resighted in the following year with a calf.

### Reproductive success comparison

In total, 50 females from the DWH oiled areas (Barataria Bay and Mississippi Sound) were determined to be pregnant (33 ultrasound diagnosed and 17 blubber-endocrine evaluated) out of 148 that were evaluated. Of these 50 pregnant females, 36 (27 Barataria Bay, 9 Mississippi Sound; Fig. 1) had sufficient sighting histories to be included in the reproductive success analysis, having been resighted either (1) 2 wk or more following the estimated due date for ultrasound-visualized pregnancies (n = 28), or (2) within 13 mo after IPD for dart biopsy evaluated pregnancies (n = 8; see Table S2 in the Supplement). Of these 36 pregnancies, all but 7 (5 Barataria Bay, 2 Mississippi Sound) resulted in failure (i.e. the females were resighted without a calf). The resulting estimated aggregated reproductive success rate for Barataria Bay (0.185, n = 27) and Mississippi Sound (0.222, n = 9) was 0.194; i.e. less than 1 in 5 detected pregnancies

resulted in a viable calf (Table 2). In comparison, the expected success rate based on the aggregate of previous observations in reference areas is over 3-fold higher (0.647, n = 34) (Table 2). Note that this result includes the 2 additional reproductive outcomes from Sarasota Bay, FL (as described above). Also, one of the observations from Charleston Harbor, SC, (individual '82503'; Bergfelt et al. 2013) was updated with additional sighting data leading to 1 correction of status from presumed failure to known success.

The posterior probability distribution on the difference in success rates between the reference areas and the oiled areas (Fig. 4a) shows the weight of evidence was 12 499 to 1 (i.e. a Bayes factor of 12 499), equating to a 99.999% probability that the reference areas had a higher reproductive success rate than the rate observed in the 2 oiled areas. To illustrate the magnitude of this difference, we present the weight evidence at 2 (70.4 to 1; Fig. 4b) and 3 times (2.7 to 1; Fig. 4c) the observed success rate of the oiled areas. That translates into 98.5% probability that the reference success rate is greater than twice, and a 73.2% probability that the difference is greater than 3 times, that of the oiled areas.

### Covariate analysis

The first covariate analysis contained data from pregnancy determinations using both the dart biopsies and the CRHA animals (see Table S3 in the Supplement). It indicated that none of the 7 covariates analyzed were included more than 10% of the time. The low inclusion rate indicates that there was no discernible evidence that success rate varied by sighting frequency, day of the year, days after DWH explosion, blubber progesterone concentration, or blubber cortisol concentration (Table 3). For reference, the mean inclusion rate for random covariates (as simulated by permuting observed values relative

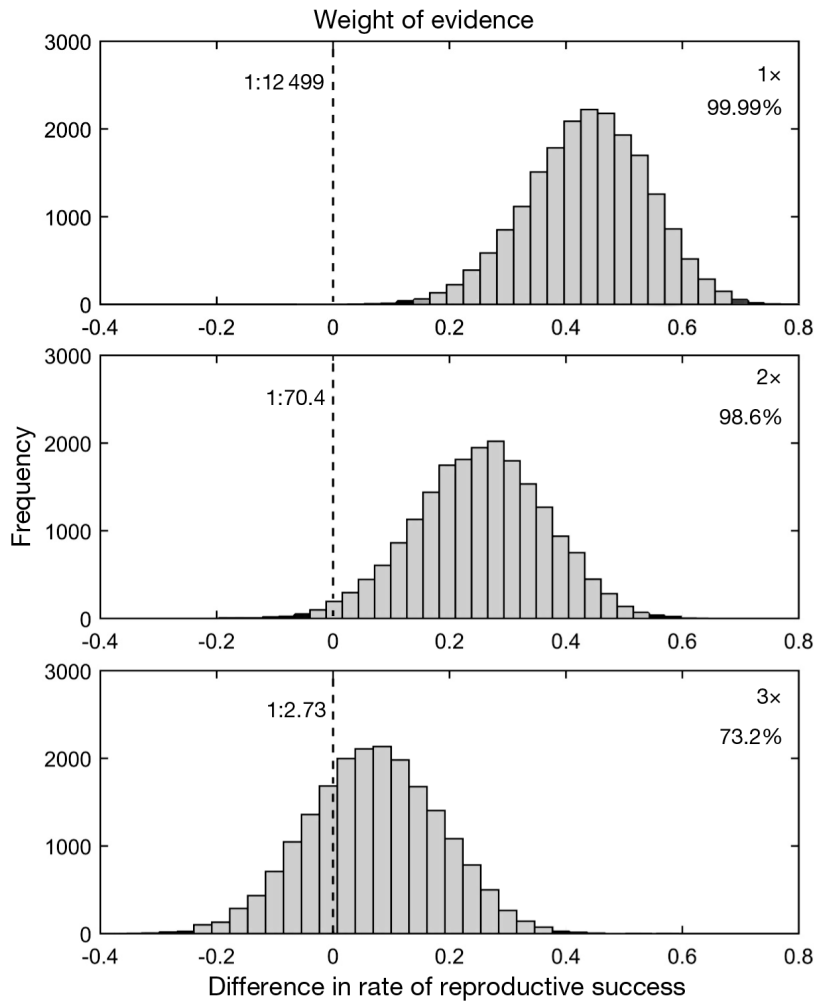


Fig. 4. *Tursiops truncatus*. Posterior probability distributions for the difference in reproductive success between the DWH-oiled GoM stocks of bottlenose dolphins and non-oiled reference stocks at (top to bottom) 1, 2 and 3 times the observed reproductive success rate of the oiled GoM stocks. The point estimates for reproductive success were 0.222 and 0.647 for the oiled and reference stocks, respectively

to success status) was 1.95%.

The second covariate analysis included only the CRHA animal data: morphometrics, ultrasound imagery, and blood data. This analysis revealed that there was 1 covariate, the leukocyte or white blood cell count (WBC), with a strong inverse relationship with reproductive success (Table 4). All 7 successes (success rate = 7/14) were observed in animals with WBC counts  $<11$  cells  $\text{nl}^{-1}$  (Table 4). None of the 18 pregnancies (success rate = 0/18) associated with levels  $>11$  cells  $\text{nl}^{-1}$  were successful. The differences were primarily driven by differences in neutrophils and eosinophils (bivariate analysis:  $p = 0.0057$  and  $0.0039$  respectively). The dramatic relationship relative to WBC count is in sharp contrast to those of the

other covariates, which showed no appreciable evidence of a relationship with reproductive success; all had inclusion rates of  $<11\%$  and, for context, the mean inclusion rate for random covariates for this second covariate analysis was 1.88%.

## DISCUSSION

The data presented in this study indicate that BSE bottlenose dolphin reproductive success was aberrantly low ( $<20\%$ ) in oiled areas following the DWH oil rig explosion. When compared to reference non-oiled areas (success  $>60\%$  in aggregate), the weight of evidence was strong ( $>98\%$  probability) that there was at minimum a 2-fold difference in success rate between the study and reference stocks. In fact, given the observed data, the point estimate indicates that the difference was more likely greater than 3-fold ( $>70\%$  probability).

As previously reported, during the 2011 capture-release health assessments and follow-up boat surveys, direct observations of offspring mortalities were recorded (Lane et al. 2015). A dead fetus *in utero*, i.e. one without a heartbeat, belonging to dolphin Y31, was observed via ultrasound examination. During the time period of this study, a number of presumed mothers were observed pushing perinate carcasses. These included dolphin Y01,

who was seen pushing a carcass of a perinate 10 mo after her initial due date, an indication of 2 consecutive reproductive failures during a 2-yr period (Lane et al. 2015).

These findings and direct observations are consistent with numerous studies directly linking polycyclic aromatic hydrocarbon (PAH) and general petroleum exposure to reproductive abnormalities and early developmental impairments. Controlled experimental studies have demonstrated a causal link in model species including mice, rats, rabbit, and mink. Reproductive failure or impairment has been documented to be caused by pyrogenic PAH exposure in mice, which exhibit a 2-fold increase in offspring loss with evidence of increased embryonic

Table 3. *Tursiops truncatus*. Model averaged coefficients for factors associated with reproductive success in oiled stocks of northern GoM bottlenose dolphin following the DWH oil spill, based on pregnancies detected by both ultrasound and blubber endocrine evaluation (n = 36; 27 from Barataria Bay, 9 from Mississippi Sound). Median values are given with 95 % probability interval values. '% selected' is the percent of iterations in which the corresponding factor was selected for inclusion in the final model. Positive and negative median coefficient values indicate direct and inverse relationships, respectively. No covariate was included in more than 50 % of the iterations; the weight of evidence is against their inclusion in the final model. For reference, the mean inclusion rate of a random covariate (modeled here as the observed covariate permuted relative to the observed outcome) is shown (gray shaded row)

Covariate	— Model coefficients —			% selected
	2.50 %	Median	97.50 %	
Sighting frequency	-2.045	-0.779	0.192	4.63
Blubber cortisol	-1.297	-0.329	0.689	1.82
Permuted covariates	Benchmark: incorporation of random no.			1.74
Days after DWH explosion	-1.053	-0.078	1.189	1.57
Blubber progesterone	-0.712	0.204	1.158	1.57
Stock (Barataria/Mississippi)	-1.127	-0.073	0.810	1.51
Day of the year	-0.931	-0.042	1.005	1.41

Table 4. *Tursiops truncatus*. Model averaged coefficients for factors associated with reproductive success based on pregnancies detected solely by ultrasound imaging (n = 32; 24 from Barataria Bay, 4 from Mississippi Sound and 4 from Sarasota Bay). Median values are given with 95 % probability interval values. '% selected' is the percent of iterations in which the corresponding factor was selected for inclusion in the final model. Positive and negative median coefficient values indicate direct and inverse relationships, respectively. Only the covariate white blood cell count was included in more than 50 % of the iterations; the weight of evidence supports its inclusion in the final model. For reference the mean inclusion rate of a random covariate (modeled here as the observed covariate permuted relative to the observed outcome) is shown (gray shaded row)

Covariate	— Model coefficients —			% selected
	2.50 %	Median	97.50 %	
White blood cell count	-5.605	-2.661	-0.817	77.77
Serum: total thyroxine	-0.119	1.498	3.973	10.22
Total length	-2.236	-0.701	0.535	3.53
Serum: progesterone	-2.095	-0.650	0.591	3.33
Lung condition score (healthy/unhealthy)	-0.965	0.590	2.285	3.07
Permuted covariates	Benchmark: incorporation of random no.			2.06
Day of the year	-0.946	0.250	1.551	2.00
Serum: 17 $\beta$ estradiol	-1.347	-0.042	1.255	1.91
Serum: cortisol	-0.867	0.175	1.319	1.63

resorption (Detmar & Jurisicova 2010). Crude-oil exposed rats exhibit high perinatal mortality with up to 72 % of parturitions of exposed pregnancies concluding in still births (Nwaigwe et al. 2012), and fetal survival rates decreasing by as much as 3-fold in offspring whose mothers were exposed to the pyrogenic PAH benzo(a)pyrene (Bui et al. 1986, Archibong et al. 2002). Female rabbits show substantial ovarian

malformations and reproductive endocrine impairments suggestive of polycystic ovarian syndrome when orally administered Escravos crude oil (Ogechukwu et al. 2014). Finally, ranch mink sows fed crude oil and bunker C oil also show a 2 to 5-fold reduction in reproductive output and much lower survival rates of their offspring before weaning (Mazet et al. 2001).

Field studies of species exposed to petroleum and petroleum products find evidence of similar reproductive and developmental effects. Numerous studies have shown that humans exposed to higher levels of PAHs across many different conditions are disproportionately likely to experience pregnancy failures or impaired prenatal and natal development (San Sebastián et al. 2002, McCoy & Salerno 2010, Merhi 2010, Wu et al. 2010). Long-term petroleum exposure in cattle can lead to poisoning that effects reproduction and early calf development (Osweiler 2005). In marine mammals studied during and after the *Exxon Valdez* oil spill, sea otters in Prince William Sound experienced high rates of fetal and neonatal loss (Tuomi & Williams 1995, Mazet et al. 2001) and killer whale pods exhibited impaired recruitment that lasted for years after the spill (Matkin et al. 2008). Similarly, long-term reproductive failures were observed in oiled seabird colonies after the *Prestige* oil spill (Velando et al. 2005, Barros et al. 2014).

Beyond the direct effects of oiling on reproduction, there are many potential maternal health effects that have been documented in other vertebrate species to significantly reduce reproductive success. The primary health effects documented in oiled dolphin stocks during and after the DWH oil spill include high rates of atypical pulmonary and adrenal diseases and disproportionate rates of poor body condition (Schwacke et al. 2014, Smith et al. 2017). Untreated pulmonary disease can cause complications during human pregnancy and early development, leading to greater likelihood of fetal and

neonatal mortality (Ramsey & Ramin 2001, Ie et al. 2002, Hartert et al. 2003, Goodnight & Soper 2005). During normal pregnancies, maternal oxygen consumption increases between 15 and 20% to support increased metabolic activity and to enrich the partial pressure of blood oxygen, which must increase to become more alkaline to create conditions that allow oxygen exchange with the fetus (Goodnight & Soper 2005). Poor maternal lung function limits the ability to provide the additional oxygen needed leading to greater likelihood of early terminations, poor fetal development, or spontaneous abortions (Goodnight & Soper 2005).

Another major health effect observed in the DWH oiled areas was impaired adrenal corticosteroid hormone production or hypoadrenocorticism (Schwacke et al. 2014, Smith et al. 2017). Humans with impaired adrenal cortex function have a higher risk of adrenal crisis, which is linked to higher fetal and maternal mortality when left untreated (Brent 1950, Keller-Wood & Wood 2001, Ambrosi et al. 2003, Beehner et al. 2006). At lower severity, adrenal cortical impairment can cause metabolic abnormalities and imbalances in blood chemistry that can lead to higher risk to fetal health and development (Mazet et al. 2001, Hobel & Culhane 2003, Mohr et al. 2008, 2010).

A disproportionate number of individuals from DWH oiled areas were also underweight relative to body length (Hart et al. 2013, Schwacke et al. 2014). Extensive research across the mammalian literature documents the relationship between low maternal body mass or nutritional deficits and reproductive failure (Verme 1969, Felig & Lynch 1970, Robinette et al. 1973, Keech et al. 2000, Bishop et al. 2009). Mechanisms by which this can occur range from direct insufficient fetal nourishment to maternal blood acidosis limiting fetal gas exchange (Felig & Lynch 1970). Moreover, there are many associated indirect health effects that increase the risk of fetal mortality as well as poor early life growth and development (Felig & Lynch 1970, Keech et al. 2000, Hobel & Culhane 2003). However, there is some evidence that poor maternal body condition may not be the primary driver of reproductive failure: although high reproductive failure continued into 2013 and 2014, the number of dolphins that exhibited abnormally low body mass dropped to near normal by 2013 (Smith et al. 2017).

Impairment of vertebrate immune systems is a commonly observed effect of PAH and general petroleum exposure, and increased susceptibility to infection has been documented in association with the DWH oil spill (Detmar & Jurisicova 2010, Whitehead et al. 2012, Whitehead 2013, Ali et al. 2014, Venn-

Watson et al. 2015a, S. De Guise et al. 2017). Of course, numerous infectious agents and abnormal inflammatory responses can in turn cause reproductive failure and impair early development (Wilson et al. 2015). From this study, there is evidence that female dolphins with total WBC counts  $>11$  cells  $\text{nl}^{-1}$  have lower probability of reproductive success. Interestingly, the range of WBC values associated with all reproductive successes observed in this study ( $<11$  cells  $\text{nl}^{-1}$ ) is nearly identical to the WBC reference ranges reported for managed-care animals (Venn-Watson et al. 2007). However, the leukocyte counts associated with all animals found in this study (i.e. both those with reproductive success and failures) were within the range (95<sup>th</sup> percentiles) previously reported for wild dolphins found in the reference non-oiled areas, where reproductive success rates were much higher (Schwacke et al. 2009). This suggests that factor(s) other than the high-normal leukocyte counts, or in addition to the leukocyte counts, may be responsible for the abnormally high reproductive failures seen in the oiled dolphin stocks. These observations are consistent with a potential synergistic adverse effect of oil exposure on the immune system with pathogen presence (S. De Guise et al. 2017), which has been noted as a potential mode of causation for the high morbidity levels observed in the oiled stocks, especially as it relates to the abnormally high prevalence of lung disease (Venn-Watson et al. 2015a) and clusters of perinatal mortality (Colegrove et al. 2016). However, these observations are also consistent with inflammatory responses associated with injuries like those found in these dolphins' lung tissue and impaired adrenal health leading to poor reproductive outcomes; consequently, additional work is needed to determine the relationship between high WBC counts and reproductive failure in these animals.

Agents other than those directly associated with petroleum have the potential to impair reproduction. However, there is strong evidence that 3 of the most likely agents, *Brucella* spp., persistent organic pollutants, and biotoxins were not disproportionately present across demographic groups in the DWH-impacted stocks (Litz et al. 2014, Balmer et al. 2015, Venn-Watson et al. 2015a, Smith et al. 2017) compared with the non-impacted reference areas. However, it has been noted that there were higher than expected *Brucella* spp. detections (representing various genetic clades) specifically in perinatal dolphin strandings (Colegrove et al. 2016) in Mississippi and Alabama (2011 to 2013), whose coastal areas were oiled during the 2010 spill. The prevalence in

Louisiana, however, was no different than non-oiled reference areas. Though the authors do not conclude that the *Brucella* spp. detection prevalence in perinates was definitely related to the oil spill, they indicate that there is evidence that oil exposure can lead to increase susceptibility to or persistence of pathogens like *Brucella* spp. through immune perturbations (Colegrove et al. 2016). We therefore conclude that non-petroleum agents are unlikely to be the primary drivers of the high reproductive failure rates observed in the oiled stocks, though multiple *Brucella* spp. clades may be part of a multifactorial insult as the physiology of these animals was strained during and after the spill. Given the anomalous massive volume of oil that entered into the habitat of these animals (National Commission on the BP Deepwater Horizon Oil Spill and Offshore Drilling 2011), the long duration of exposure (spilled oil persisted in marsh sediment at least until 2014; Turner et al. 2014), and the well-documented effects of PAHs, petroleum, and petroleum products on mammalian and vertebrate reproduction, early development, and early survival, other plausible explanations for the observed 3-fold decrease in dolphin reproductive success are difficult to postulate.

**Acknowledgements.** The authors thank Brian Quigley and Nicole Beaulieu for help with data preparation. Additionally, we express our gratitude to Wayne Perryman, Kathryn Sherman, James Carretta, Vicki Pease, Gaby Serra-Valente, Kelly Robertson, Susan Chivers, Kerri Danil, and John Durban for facilitating our research in the laboratory and offering valuable feedback during the writing process. This research was part of the DWH NRDA performed cooperatively among NOAA, other federal and state Trustees, and BP plc. The findings and conclusions in this paper are those of the authors and do not necessarily represent the view of NOAA or of any other natural resource Trustee for the BP/Deepwater Horizon NRDA. The dolphin health assessments protocols were approved by the NOAA Institutional Animal Care and Use Committee. The field work was conducted under the authority of Scientific Research Permit nos. 779-1633 and 932-1905/MA-009526 issued by NOAA's National Marine Fisheries Service pursuant to the US Marine Mammal Protection Act. Logistical support was provided in part by the Louisiana Department of Fisheries and Wildlife.

#### LITERATURE CITED

- Adams JD, Speakman T, Zolman E, Schwacke LH (2006) Automating image matching, cataloging, and analysis for photo-identification research. *Aquat Mamm* 32:374–384
- Ali AO, Hohn C, Allen PJ, Ford L, Dail MB, Pruett S, Petrie-Hanson L (2014) The effects of oil exposure on peripheral blood leukocytes and splenic melano-macrophage centers of Gulf of Mexico fishes. *Mar Pollut Bull* 79:87–93
- Ambrosi B, Barbetta L, Morricone L (2003) Diagnosis and management of Addison's disease during pregnancy. *J Endocrinol Invest* 26:698–702
- Archibong AE, Inyang F, Ramesh A, Greenwood M and others (2002) Alteration of pregnancy related hormones and fetal survival in F-344 rats exposed by inhalation to benzo(a)pyrene. *Reprod Toxicol* 16:801–808
- Balmer BC, Ylitalo GM, McGeorge LE, Baugh KA and others (2015) Persistent organic pollutants (POPs) in blubber of common bottlenose dolphins (*Tursiops truncatus*) along the northern Gulf of Mexico coast, USA. *Sci Total Environ* 527–528:306–312
- Barbara JM (1999) Behavioral development in wild bottlenose dolphin newborns (*Tursiops* sp.). *Behaviour* 136: 529–566
- Barros Á, Álvarez D, Velando A (2014) Long-term reproductive impairment in a seabird after the *Prestige* oil spill. *Biol Lett* 10:20131041
- Beehner JC, Nguyen N, Wango EO, Alberts SC, Altmann J (2006) The endocrinology of pregnancy and fetal loss in wild baboons. *Horm Behav* 49:688–699
- Bergfelt DR, Steinetz BG, Lasano S, West KL, Campbell M, Adams GP (2011) Relaxin and progesterone during pregnancy and the post-partum period in association with live and stillborn calves in bottlenose dolphins (*Tursiops truncatus*). *Gen Comp Endocrinol* 170:650–656
- Bergfelt DR, Steinetz BG, Reif JS, Schaefer AM and others (2013) Evaluation of single-sample analysis of progesterone in combination with relaxin for diagnosis of pregnancy in wild bottlenose dolphins (*Tursiops truncatus*). *Aquat Mamm* 39:187
- Bilbao E, Raingard D, de Cerio OD, Ortiz-Zarragoitia M and others (2010) Effects of exposure to Prestige-like heavy fuel oil and to perfluorooctane sulfonate on conventional biomarkers and target gene transcription in the thicklip grey mullet *Chelon labrosus*. *Aquat Toxicol* 98: 282–296
- Bishop CJ, White GC, Freddy DJ, Watkins BE, Stephenson TR (2009) Effect of enhanced nutrition on mule deer population rate of change. *Wildl Monogr* 172:1–28
- Brent F (1950) Addison's disease and pregnancy. *Am J Surg* 79:645–652
- Browning CL, Rolland RM, Kraus SD (2010) Estimated calf and perinatal mortality in western North Atlantic right whales (*Eubalaena glacialis*). *Mar Mamm Sci* 26:648–662
- Bui QQ, Tran MB, West WL (1986) A comparative study of the reproductive effects of methadone and benzo [a] pyrene in the pregnant and pseudopregnant rat. *Toxicology* 42:195–204
- Carlin BP, Chib S (1995) Bayesian model choice via Markov chain Monte Carlo methods. *J R Stat Soc B* 57:473–484
- Carmichael RH, Graham WM, Aven A, Worthy G, Howden S (2012) Were multiple stressors a 'perfect storm' for northern Gulf of Mexico bottlenose dolphins (*Tursiops truncatus*) in 2011. *PLOS ONE* 7:e41155
- Chib S, Carlin BP (1999) On MCMC sampling in hierarchical longitudinal models. *Stat Comput* 9:17–26
- Colegrove KM, Venn-Watson S, Litz J, Kinsel MJ and others (2016) Fetal distress and *in utero* pneumonia in perinatal dolphins during the Northern Gulf of Mexico unusual mortality event. *Dis Aquat Org* 119:1–16
- De Guise S, Levin M, Gebhard E, Jasperse L and others (2017) Changes in immune functions in bottlenose dolphins in the northern Gulf of Mexico associated with the *Deepwater Horizon* oil spill. *Endang Species Res* (in press) doi:10.3354/esr00814

- de Soysa TY, Ulrich A, Friedrich T, Pite D and others (2012) Macondo crude oil from the Deepwater Horizon oil spill disrupts specific developmental processes during zebrafish embryogenesis. *BMC Biol* 10:40
- Deepwater Horizon Natural Resource Damage Assessment Trustees (2016) *Deepwater Horizon* oil spill: final programmatic damage assessment and restoration plan (PDARP) and final programmatic environmental impact statement (PEIS). [www.gulfspillrestoration.noaa.gov/restoration-planning/gulf-plan](http://www.gulfspillrestoration.noaa.gov/restoration-planning/gulf-plan)
- Detmar J, Jurisicova A (2010) Embryonic resorption and polycyclic aromatic hydrocarbons: putative immune-mediated mechanisms. *Syst Biol Reprod Med* 56:3–17
- Felig P, Lynch V (1970) Starvation in human pregnancy: hypoglycemia, hypoinsulinemia, and hyperketonemia. *Science* 170:990–992
- Goodnight WH, Soper DE (2005) Pneumonia in pregnancy. *Crit Care Med* 33:S390–S397
- Hart LB, Wells RS, Schwacke LH (2013) Reference ranges for body condition in wild bottlenose dolphins *Tursiops truncatus*. *Aquat Biol* 18:63–68
- Hartert TV, Neuzil KM, Shintani AK, Mitchel EE and others (2003) Maternal morbidity and perinatal outcomes among pregnant women with respiratory hospitalizations during influenza season. *Am J Obstet Gynecol* 189:1705–1712
- Hawkes JW, Stehr CM (1982) Cytopathology of the brain and retina of embryonic surf smelt (*Hypomesus pretiosus*) exposed to crude oil. *Environ Res* 27:164–178
- Hobel C, Culhane J (2003) Role of psychosocial and nutritional stress on poor pregnancy outcome. *J Nutr* 133:1709S–1717S
- Ie S, Rubio ER, Alper B, Szerlip HM (2002) Respiratory complications of pregnancy. *Obstet Gynecol Surv* 57:39–46
- Kass RE, Raftery AE (1995) Bayes factors. *J Am Stat Assoc* 90:773–795
- Kawaguchi M, Sugahara Y, Watanabe T, Irie K and others (2012) Nervous system disruption and concomitant behavioral abnormality in early hatched pufferfish larvae exposed to heavy oil. *Environ Sci Pollut Res Int* 19:2488–2497
- Keech MA, Bowyer RT, Jay M, Hoef V, Boertje RD, Dale BW, Stephenson TR (2000) Life-history consequences of maternal condition in Alaskan moose. *J Wildl Manage* 64:450–462
- Kellar NM, Trego ML, Marks CI, Dizon AE (2006) Determining pregnancy from blubber in three species of delphinids. *Mar Mamm Sci* 22:1–16
- Kellar NM, Trego ML, Marks CI, Chivers SJ, Danil K, Archer FI (2009) Blubber testosterone: a potential marker of male reproductive status in short-beaked common dolphins. *Mar Mamm Sci* 25:507–522
- Kellar NM, Keliher J, Trego ML, Catelani KN, Hanns C, George J, Rosa C (2013a) Variation of bowhead whale progesterone concentrations across demographic groups and sample matrices. *Endang Species Res* 22:61–72
- Kellar NM, Trego ML, Chivers SJ, Archer FI (2013b) Pregnancy patterns of pantropical spotted dolphins (*Stenella attenuata*) in the eastern tropical Pacific determined from hormonal analysis of blubber biopsies and correlations with the purse-seine tuna fishery. *Mar Biol* 160:3113–3124
- Kellar NM, Trego ML, Chivers SJ, Archer FI, Perryman WL (2014) From progesterone in biopsies to estimates of pregnancy rates: large scale reproductive patterns of two sympatric species of common dolphin, *Delphinus* spp. off California, USA and Baja, Mexico. *Bull South Calif Acad Sci* 113:58–80
- Kellar NM, Catelani KN, Robbins MN, Trego ML, Allen CD, Danil K, Chivers SJ (2015) Blubber cortisol: a potential tool for assessing stress response in free-ranging dolphins without effects due to sampling. *PLOS ONE* 10:e0115257
- Keller-Wood M, Wood CE (2001) Pituitary-adrenal physiology during pregnancy. *Endocrinologist* 11:159–170
- Kocan R, Marty G, Okihiro M, Brown E, Baker T (1996) Reproductive success and histopathology of individual Prince William Sound Pacific herring 3 years after the (*Exxon Valdez*) oil spill. *Can J Fish Aquat Sci* 53:2388–2393
- Lacave G, Eggermont M, Verslycke T, Brook F, Salbany A, Roque L, Kinoshita R (2004) Prediction from ultrasonographic measurements of the expected delivery date in two species of bottlenosed dolphin (*Tursiops truncatus* and *Tursiops aduncus*). *Vet Rec* 154:228–233
- Lane SM, Smith CR, Mitchell J, Balmer BC and others (2015) Reproductive outcome and survival of common bottlenose dolphins sampled in Barataria Bay, Louisiana, USA, following the *Deepwater Horizon* oil spill. *Proc R Soc B* 2015:282 20151944
- Litz JA, Baran MA, Bowen-Stevens SR, Carmichael RH and others (2014) Review of historical unusual mortality events (UMEs) in the Gulf of Mexico (1990–2009): providing context for the multi-year northern Gulf of Mexico cetacean UME declared in 2010. *Dis Aquat Org* 112:161–175
- Lo NC (1994) Level of significance and power of two commonly used procedures for comparing mean values based on confidence intervals. *CCOFI Rep* 35:246–253
- Lunn DJ, Thomas A, Best N, Spiegelhalter D (2000) WinBUGS—a Bayesian modelling framework: concepts, structure, and extensibility. *Stat Comput* 10:325–337
- Manly BF (1991) Randomization, bootstrap and Monte Carlo methods in biology. CRC Press, Boca Raton, FL
- Mansour AA, Mokay DW, Lien J, Orr JC, Banoub JH, Ien NØ, Stenson G (2002) Determination of pregnancy status from blubber samples in minke whales (*Balaenoptera acutorostrata*). *Mar Mamm Sci* 18:112–120
- Matkin CO, Saulitis EL, Ellis GM, Olesiuk P, Rice SD (2008) Ongoing population-level impacts on killer whales *Orcinus orca* following the 'Exxon Valdez' oil spill in Prince William Sound, Alaska. *Mar Ecol Prog Ser* 356:269–281
- Mazet JA, Gardner IA, Jessup DA, Lowenstine LJ (2001) Effects of petroleum on mink applied as a model for reproductive success in sea otters. *J Wildl Dis* 37:686–692
- McCoy MA, Salerno JA (2010) Assessing the effects of the Gulf of Mexico oil spill on human health. The National Academy of Sciences, Washington, DC
- Melancon R, Lane S, Speakman T, Hart LB and others (2011) Photo-identification field and laboratory protocols utilizing FinBase version 2. NOAA Tech Memo NMFS-SEFSC 627
- Merhi ZO (2010) Gulf Coast oil disaster: impact on human reproduction. *Fertil Steril* 94:1575–1577
- Michel J, Owens EH, Zengel S, Graham A and others (2013) Extent and degree of shoreline oiling: *Deepwater Horizon* oil spill, Gulf of Mexico, USA. *PLOS ONE* 8:e65087
- Mohr FC, Lasley B, Bursian S (2008) Chronic oral exposure to bunker C fuel oil causes adrenal insufficiency in ranch mink (*Mustela vison*). *Arch Environ Contam Toxicol* 54:337–347
- Mohr FC, Lasley B, Bursian S (2010) Fuel oil-induced adrenal hypertrophy in ranch mink (*Mustela vison*): effects of

- sex, fuel oil weathering, and response to adrenocorticotropic hormone. *J Wildl Dis* 46:103–110
- National Commission on the BP Deepwater Horizon Oil Spill and Offshore Drilling (2011) Deep water: the Gulf oil disaster and the future of offshore drilling. Report to the President. US Government Printing Office, Washington, DC. <https://www.gpo.gov/fdsys/pkg/GPO-OILCOMMISSION/pdf/GPO-OILCOMMISSION.pdf>
- Nwaigwe A, Anya K, Nwaigwe C, Nwaigwe C (2012) The effect of Nigerian Qua-iboe Brent crude oil on the reproductive performance of female Wistar albino rats. *J Environ Sci Technol* 5:84–90
- O'Brien JK, Robeck T (2012) The relationship of maternal characteristics and circulating progesterone concentrations with reproductive outcome in the bottlenose dolphin (*Tursiops truncatus*) after artificial insemination, with and without ovulation induction, and natural breeding. *Theriogenology* 78:469–482
- Ogechukwu OJ, Ajuluchukwu NA, Ifedioranma OS, Obianuju OA (2014) Histopathological and hormonal disrupting effects of Escravos crude oil on the ovary of Chinchilla rabbits. *J Toxicol Environ Health Sci* 6:31–37
- Oswieiler GD (2005) Petroleum product poisoning. In: Kahn CM, Line S, Aiello S (eds) *The Merck veterinary manual*. Merck & Co, Whitehouse Station, NJ, p 2015–2160
- Pérez S, García-López Á, De Stephanis R, Giménez J and others (2011) Use of blubber levels of progesterone to determine pregnancy in free-ranging live cetaceans. *Mar Biol* 158:1677–1680
- Ramsey PS, Ramin KD (2001) Pneumonia in pregnancy. *Obstet Gynecol Clin North Am* 28:553–569
- Robeck T, Atkinson S, Brook F (2001) Reproduction. In: Dierauf L, Gulland F (eds) *CRC handbook in marine mammal medicine*, 2nd edn. CRC Press, Boca Raton, FL, p 193–236
- Robeck TR, Gili C, Doescher BM, Sweeney J, De Laender P, Van Elk CE, O'Brien JK (2012) Altrenogest and progesterone therapy during pregnancy in bottlenose dolphins (*Tursiops truncatus*) with progesterone insufficiency. *J Zoo Wildl Med* 43:296–308
- Robinette WL, Baer CH, Pillmore RE, Knittle CE (1973) Effects of nutritional change on captive mule deer. *J Wildl Manage* 37:312–326
- Rosel PE (2003) PCR-based sex determination in Odontocete cetaceans. *Conserv Genet* 4:647–649
- San Sebastián MS, Armstrong B, Stephens C (2002) Outcomes of pregnancy among women living in the proximity of oil fields in the Amazon basin of Ecuador. *Int J Occup Environ Health* 8:312–319
- Schwacke LH, Hall AJ, Townsend FI, Wells RS and others (2009) Hematologic and serum biochemical reference intervals for free-ranging common bottlenose dolphins (*Tursiops truncatus*) and variation in the distributions of clinicopathologic values related to geographic sampling site. *Am J Vet Res* 70:973–985
- Schwacke LH, Twiner MJ, De Guise S, Balmer BC and others (2010) Eosinophilia and biotoxin exposure in bottlenose dolphins (*Tursiops truncatus*) from a coastal area impacted by repeated mortality events. *Environ Res* 110:548–555
- Schwacke LH, Zolman ES, Balmer BC, De Guise S and others (2011) Anaemia, hypothyroidism and immune suppression associated with polychlorinated biphenyl exposure in bottlenose dolphins (*Tursiops truncatus*). *Proc R Soc B* 279:48–57.
- Schwacke LH, Smith CR, Townsend FI, Wells RS and others (2014) Health of common bottlenose dolphins (*Tursiops truncatus*) in Barataria Bay, Louisiana, following the Deepwater Horizon oil spill. *Environ Sci Technol* 48:93–103
- Sinclair C, Sinclair J, Zolman E, Martinez A, Balmer B, Barry K (2015) Remote biopsy sampling field procedures for cetaceans used during the natural resource damage assessment of the MSC252 Deepwater Horizon oil spill. NOAA Tech Memo NMFS-SEFSC-670
- Smith CR, Jensen ED, Blankenship BA, Greenberg M and others (2013) Fetal omphalocele in a common bottlenose dolphin (*Tursiops truncatus*). *J Zoo Wildl Med* 44:87–92
- Smith CR, Rowles TK, Hart LB, Townsend FI and others (2017) Slow recovery of Barataria Bay dolphin health following Deepwater Horizon oil spill (2013–2014), with evidence of persistent lung disease and impaired stress response. *Endang Species Res* 33:127–142
- Spiegelhalter D, Thomas A, Best N, Lunn D 2003. WinBUGS user manual, Version 1.4. <http://www.mrc-bsu.ac.uk/wp-content/uploads/manual14.pdf>
- Steinman KJ, Robeck TR, O'Brien JK (2016) Characterization of estrogens, testosterone, and cortisol in normal bottlenose dolphin (*Tursiops truncatus*) pregnancy. *Gen Comp Endocrinol* 226:102–112
- Stone L, Johnson R, Sweeney J, Lewis M, Fowler M, Miller R (1999) Fetal ultrasonography in dolphins with emphasis on gestational aging. In: Fowler M, Miller R (eds) *Zoo & wild animal medicine*, 4th edn. WB Saunders, Philadelphia, PA, p 501–506.
- Trana MR, Roth JD, Tomy GT, Anderson WG, Ferguson SH (2016) Increased blubber cortisol in ice-entrapped beluga whales (*Delphinapterus leucas*). *Polar Biol* 9:1563–1569
- Trego ML, Kellar NM, Danil K (2013) Validation of blubber progesterone concentrations for pregnancy determination in three dolphin species and a porpoise. *PLOS ONE* 8:e69709
- Tuomi P, Williams T (1995) Rehabilitation of pregnant sea otters and females with newborn pups. In: Williams TM, Davis DW (eds) *Emergency care and rehabilitation of oiled sea otters: a guide for oil spills involving fur-bearing marine mammals*. University of Alaska Press, Fairbanks, AK, p 121–132.
- Turner RE, Overton E, Meyer B, Miles M, Hooper-Bui L (2014) Changes in the concentration and relative abundance of alkanes and PAHs from the Deepwater Horizon oiling of coastal marshes. *Mar Pollut Bull* 86:291–297
- Urian KW, Hohn AA, Hansen LJ (1999) Status of the photo-identification catalog of coastal bottlenose dolphins of the western North Atlantic: report of a workshop of catalog contributors. NOAA Administrative Report NMFS-SEFSC-425, NOAA-NMFS, US Dept of Commerce, Beaufort, NC
- Urian K, Gorgone A, Read A, Balmer B and others (2015) Recommendations for photo-identification methods used in capture-recapture models with cetaceans. *Mar Mamm Sci* 31:298–321
- Velando A, Alvarez D, Mourino J, Arcos F, Barros A (2005) Population trends and reproductive success of the European shag, *Phalacrocorax aristotelis*, on the Iberian Peninsula following the Prestige oil spill. *J Ornithol* 146:116–120
- Venn-Watson S, Jensen ED, Ridgway SH (2007) Effects of age and sex on clinicopathologic reference ranges in a healthy managed Atlantic bottlenose dolphin population. *J Am Vet Med Assoc* 231:596–601

- Venn-Watson S, Colegrove KM, Litz J, Kinsel M and others (2015a) Adrenal gland and lung lesions in Gulf of Mexico common bottlenose dolphins (*Tursiops truncatus*) found dead following the *Deepwater Horizon* oil spill. PLOS ONE 10: e0126538,
- ✦ Venn-Watson S, Garrison L, Litz J, Fougères E and others (2015b) Demographic clusters identified within the Northern Gulf of Mexico common bottlenose dolphin (*Tursiops truncatus*) unusual mortality event: January 2010–June 2013. PLOS ONE 10:e0117248
- ✦ Verme LJ (1969) Reproductive patterns of white-tailed deer related to nutritional plane. J Wildl Manag 33:881–887
- Wells RS (2014) Social structure and life history of bottlenose dolphins near Sarasota Bay, Florida: insights from four decades and five generations. In: Yamagiwa J, Karczmarski L (eds) Primates and cetaceans: field research and conservation of complex mammalian societies. Springer, Tokyo, p 149–172
- Wells RS, Scott MD (1990) Estimating bottlenose dolphin population parameters from individual identification and capture-release techniques. Rep Int Whaling Comm 12: 407–415
- ✦ Wells RS, Rhinehart HL, Hansen LJ, Sweeney JC and others (2004) Bottlenose dolphins as marine ecosystem sentinels: developing a health monitoring system. Eco-Health 1:246–254
- ✦ Wells RS, Tornero V, Borrell A, Aguilar A and others (2005) Integrating life-history and reproductive success data to examine potential relationships with organochlorine compounds for bottlenose dolphins (*Tursiops truncatus*) in Sarasota Bay, Florida. Sci Total Environ 349:106–119
- ✦ Wells RS, Smith CR, Sweeney JC, Townsend FI and others (2014) Fetal survival of common bottlenose dolphins (*Tursiops truncatus*) in Sarasota Bay, Florida. Aquat Mamm 40:252
- ✦ Whitehead A (2013) Interactions between oil-spill pollutants and natural stressors can compound ecotoxicological effects. Integr Comp Biol 53:635–647
- ✦ Whitehead A, Dubansky B, Bodinier C, Garcia TI and others (2012) Genomic and physiological footprint of the *Deepwater Horizon* oil spill on resident marsh fishes. Proc Natl Acad Sci USA 109:20298–20302
- Wilson CB, Nizet V, Maldonado Y, Klein JO, Remington JS (eds) (2015) Remington and Klein's infectious disease of the fetus and newborn infant (8th edn). Elsevier Health Sciences, Philadelphia, PA
- ✦ Wu J, Hou H, Ritz B, Chen Y (2010) Exposure to polycyclic aromatic hydrocarbons and missed abortion in early pregnancy in a Chinese population. Sci Total Environ 408:2312–2318

Editorial responsibility: Brendan Godley,  
University of Exeter, Cornwall Campus, UK

Submitted: May 26, 2016; Accepted: October 3, 2016  
Proofs received from author(s): December 12, 2016