Underestimating the damage: interpreting cetacean carcass recoveries in the context of the Deepwater Horizon/BP incident

Rob Williams¹, Shane Gero², Lars Bejder³, John Calambokidis⁴, Scott D. Kraus⁵, David Lusseau⁶, Andrew J. Read⁷, & Jooke Robbins⁸

¹Marine Mammal Research Unit, University of British Columbia, Vancouver, Canada
²Department of Biology, Dalhousie University, Halifax, Canada
³Centre for Fish and Fisheries Research, Cetacean Research Unit, Murdoch University, Western Australia
⁴Cascadia Research Collective, Olympia, WA, USA
⁵New England Aquarium, Boston, MA, USA
⁶School of Biology, Aberdeen University, Aberdeen, Scotland, UK
⁷Nicholas School of the Environment, Duke University, Beaufort, NC, USA
⁸Humpback Whale Studies Program, Provincetown Center for Coastal Studies, Provincetown, MA, USA

Keywords
Anthropogenic impacts; dolphin; Deepwater Horizon; Gulf of Mexico; mortality; oil; strandings.

Abstract
Evaluating impacts of human activities on marine ecosystems is difficult when effects occur out of plain sight. Oil spill severity is often measured by the number of marine birds and mammals killed, but only a small fraction of carcasses are recovered. The Deepwater Horizon/BP oil spill in the Gulf of Mexico was the largest in the U.S. history, but some reports implied modest environmental impacts, in part because of a relatively low number (101) of observed marine mammal mortalities. We estimate historical carcass-detection rates for 14 cetacean species in the northern Gulf of Mexico that have estimates of abundance, survival rates, and stranding records. This preliminary analysis suggests that carcasses are recovered, on an average, from only 2% (range: 0–6.2%) of cetacean deaths. Thus, the true death toll could be 50 times the number of carcasses recovered, given no additional information. We discuss caveats to this estimate, but present it as a counterpoint to illustrate the magnitude of misrepresentation implicit in presenting observed carcass counts without similar qualification. We urge methodological development to develop appropriate multipliers. Analytical methods are required to account explicitly for low probability of carcass recovery from cryptic mortality events (e.g., oil spills, ship strikes, bycatch in unmonitored fisheries and acoustic trauma).

Introduction
The Deepwater Horizon/BP oil spill in the Gulf of Mexico was not only the largest in the US history (Machlis & McNutt 2010) but was also the first to release oil at the sea floor (over 1.5 km below sea level) and to involve the widespread use of dispersants below the surface (Mascarelli 2010). However, many media reports have suggested that the spill caused only modest environmental impacts (Grunwald 2010; Walsh 2010), in part because of a low number of observed wildlife mortalities, especially marine mammals (Unified Area Command from the U.S. Fish and Wildlife Service and the National Oceanic and Atmospheric Association 2010). Not surprisingly, perhaps, this spill has been compared to other acute environmental disasters, such as the 1989 Exxon Valdez oil spill (EVOS). In the case of EVOS, the mortality of sea otters became emblematic of environmental impact, as well as a contentious effort to agree on compensation (Ehrenfeld 1990; Estes 1991). In contrast, the Deepwater Horizon/BP event has not left such an iconic symbol of devastation. As of November 7, 2010, “only” 101 cetacean (whale, dolphin, and porpoise) carcasses¹ had been detected across the Northern Gulf

¹ http://www.nmfs.noaa.gov/pr/health/oilspill.htm
of Mexico. The critical issue is, therefore, how to interpret this relatively low number of carcass recoveries in terms of impact to populations. The Gulf of Mexico is a semienclosed subtropical sea that forms essentially one ecosystem with many demographically independent cetacean populations (Mullin & Fulling 2004). Some of these cetacean populations, such as killer whales (Orcinus Orca), false killer whales (Pseudorca crassidentis), melon-headed whales (Peponocephala electra), and several beaked whale species, appear to be quite small, are poorly studied, or are found in the pelagic realm where they could have been exposed to oil and yet never strand. Small, genetically isolated populations of bottlenose dolphins (Tursiops truncatus) could have experienced substantial losses either inshore or offshore.

In an ideal world, one would simply compare postspill to prespill abundance estimates. But, it is rare to have good knowledge of long-term trends in wildlife abundance (Bonebrake et al. 2010). Abundance of many marine mammal populations has been monitored for decades, but the low precision of most cetacean abundance estimates would prevent us from detecting all but the most catastrophic declines using conventional null-hypothesis testing (Taylor et al. 2007b). As a result, it would not be very informative to compare pre- and postspill abundance estimates for populations of cetaceans in the northern Gulf of Mexico. An alternative approach is to count the number of carcasses recovered, acknowledging that these recoveries were subject to a number of processes (e.g., sinking, decaying, scavenging, drifting) that reduce detection probability, and then adjusting the counts upward to estimate total mortality. This is the approach that is commonly taken to estimate the effects of power lines on bird mortality, for example, in which it has been shown in one instance that carcass counts underestimate total mortality by 32% (Ponce et al. 2010). This also appears to be the approach being taken to assess impacts of the Deepwater Horizon Incident on whales and dolphins, with the important caveat that the carcass counts appear to be presented at face value, with no attempt to extrapolate to total mortality (Unified Area Command from the U.S. Fish and Wildlife Service and the National Oceanic and Atmospheric Association 2010; Grunwald 2010; Walsh 2010).

Cetacean carcasses do not necessarily strand along coastlines or remain afloat long enough to be detected at sea. The probability of detecting the death of a marine mammal depends on a wide range of physical and biological factors, including: behavioral responses prior to death, proximity of the carcass to shore (or at-sea observers), decomposition rates and processes, water temperature, wind regime, and local currents (Epperly et al. 1996). Cetaceans subject to natural predation would obviously leave no carcass at all. Shore recoveries may be very site-specific, such that the likelihood of a carcass drifting to shore varies with the geography of the coastline itself (Faerber & Baird 2010). As such, “oiled” carcasses detected subsequent to the Deepwater Horizon/BP event are expected to represent a small fraction of total mortality in the Northern Gulf of Mexico.

Given the magnitude of the spill and complexity of the response, quantifying the ecological impacts will take a long time. To contribute to this effort, we examined historical data from the Northern Gulf of Mexico to evaluate whether cetacean carcass counts in this region have previously been reliable indicators of mortality, and may therefore accurately represent deaths caused by the Deepwater Horizon/BP event.

**Methods**

We estimated historical carcass-detection rates for 14 species of cetaceans in the northern Gulf of Mexico for which species-specific estimates of abundance (Waring et al. 2009a, b; Mullin & Fulling 2004) species-level adult-survival rates (Taylor et al. 2007a), and stranding records exist (Waring et al. 2009a, b). Estimates of mortality were generated for each species by multiplying recent abundance estimates by the species-specific mortality rate. An annual carcass-recovery rate was then estimated by dividing the mean number of observed strandings each year by our estimate of annual mortality (Table 1). First, an overall pooled carcass-recovery rate was calculated across all cetacean species (n = 14) in the Gulf of Mexico for which data was available by using the expected number of deaths across all species and the total number of observed carcasses across all species. Next, species-specific carcass-recovery rates were calculated using only species-specific values and a mean (n = 14) of those was taken across species. Species-specific carcass-recovery estimates were not generated for Bryde’s whale (Balaenoptera brydei), bottlenose dolphins (Tursiops sp.), and Fraser’s dolphins (Lagenodelphis hosei) due to uncertainties in their abundance and/or population structure (Waring et al. 2009b). No attempt was made to estimate carcass-recovery rates in the two taxonomic groups that are not identified to species in the field during raw data collection: Kogia (a pooled estimate for two species, dwarf and pygmy sperm whales), or mesoplodonts (a pooled estimate for a genus of similar-looking beaked whales) (Waring et al. 2009b).

**Results**

Our analysis suggests that an average of 4,474 individual cetaceans died annually between 2003 and 2007 from all
Low probability of cetacean carcass recovery

R. Williams et al.

Table 1  Population parameters and illustrative species-specific carcass-recovery rates for 14 species from the Gulf of Mexico.

<table>
<thead>
<tr>
<th>Northern Gulf of Mexico population</th>
<th>Population estimate</th>
<th>CVa</th>
<th>Adult-estimated annual mortalityb</th>
<th>Mean observed annual strandingsc</th>
<th>Carcass-detection rate (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sperm whale</td>
<td>1665</td>
<td>0.20</td>
<td>0.986</td>
<td>23.3</td>
<td>0.8</td>
</tr>
<tr>
<td>Cuvier's beaked whale</td>
<td>65</td>
<td>0.67</td>
<td>0.95</td>
<td>3.3</td>
<td>0.2</td>
</tr>
<tr>
<td>Atlantic spotted dolphin\1</td>
<td>3761</td>
<td>0.28</td>
<td>0.95</td>
<td>1880.6</td>
<td>2.4</td>
</tr>
<tr>
<td>Pantropical spotted dolphin</td>
<td>34067</td>
<td>0.18</td>
<td>0.95</td>
<td>1703.4</td>
<td>0.8</td>
</tr>
<tr>
<td>Striped dolphin</td>
<td>323</td>
<td>0.48</td>
<td>0.95</td>
<td>99.5</td>
<td>1</td>
</tr>
<tr>
<td>Spinner dolphin</td>
<td>1989</td>
<td>0.48</td>
<td>0.95</td>
<td>132.7</td>
<td>5.8</td>
</tr>
<tr>
<td>Clymene dolphin</td>
<td>6575</td>
<td>0.36</td>
<td>0.95</td>
<td>328.8</td>
<td>0.6</td>
</tr>
<tr>
<td>Killer whale</td>
<td>49</td>
<td>0.77</td>
<td>0.99</td>
<td>0.5</td>
<td>0</td>
</tr>
<tr>
<td>False killer whale</td>
<td>777</td>
<td>0.56</td>
<td>0.99</td>
<td>7.8</td>
<td>0</td>
</tr>
<tr>
<td>Pygmy killer whale</td>
<td>323</td>
<td>0.60</td>
<td>0.95</td>
<td>16.2</td>
<td>0.2</td>
</tr>
<tr>
<td>Melon-headed whale</td>
<td>2283</td>
<td>0.76</td>
<td>0.99</td>
<td>22.8</td>
<td>1.4</td>
</tr>
<tr>
<td>Risso's dolphin</td>
<td>1589</td>
<td>0.27</td>
<td>0.95</td>
<td>79.5</td>
<td>2.8</td>
</tr>
<tr>
<td>Short-finned pilot whale</td>
<td>716</td>
<td>0.34</td>
<td>0.986</td>
<td>10.0</td>
<td>0.2</td>
</tr>
<tr>
<td>Average of all species</td>
<td>93,687</td>
<td>–</td>
<td>–</td>
<td>4,474</td>
<td>17</td>
</tr>
<tr>
<td>Pooled across all species (n = 14)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

\1Population abundance and stranding data (2002–2006) were taken from (Waring et al. 2009a).

Discussion

Our results indicate that carcass-recovery rates are historically low for cetaceans in the Gulf of Mexico. Studies of other populations show similar recovery rates. In long-term studies of killer whales off the coasts of British Columbia and Washington State, in which populations are censused completely every year, carcasses from confirmed deaths of known individuals are recovered only 6% of the time (Fisheries and Oceans Canada 2008). Similarly, low-detection rates have been estimated for carcasses of eastern gray whales (Eischricttus robustus, <5%, Heyning & Dahlheim 1990), North Atlantic right whales (Eubalaena glacialis, 17%, Kraus et al. 2005), and harbor porpoises (Phocoena phocoena, <1%, Moore & Read 2008), all of which occur in near-shore waters. Beached carcasses of other pelagic marine vertebrates have been shown to be equally poor indicators of mortality (for example, 7–13% recovery rates for four species of sea turtle, Epperly et al. 1996). As such, raw carcass counts alone are not reliable indicators of the magnitude of mortality for these species.

We do not claim to have calculated definitive multipliers for this spill. Instead, our aim is to show plausible ranges for those multipliers, in order to illustrate how much they would affect our perception of the ecological damages caused by Deepwater Horizon incident and why this topic is worthy of additional resources for methodological development. Consider, for example, one sperm whale being detected as a carcass, and a necropsy identified oozing as a contributing factor in the whale’s death. If the carcass-detection rate for sperm whales is 3.4% (Table 1), then it is plausible that 29 sperm whale deaths represents the best estimate of total mortality, given no natural and anthropogenic causes. However, during that period, an average of only 17 cetacean carcasses were detected annually along the northern Gulf of Mexico. This would suggest that the overall pooled rate of carcass recovery for cetaceans in the Gulf of Mexico is approximately 0.4% of the total estimated mortality. Table 1 breaks down the recovery rates by species. Carcasses were recovered only from a mean of 2.0% (range: 0–6.2%) of cetacean species deaths along the northern Gulf of Mexico. The disparity between this value and the overall pooled value likely results from undue influence of poorly studied and relatively rare species (e.g., Cuvier’s beaked whale and melon-headed whale; Table 1) with high estimated carcass-recovery rates that are weighted equivalently and treated as reliably in this average as estimates from species that are common and well studied. We have reason to believe that the Cuvier’s beaked whale recovery rate is positively biased. The original abundance estimate is thought to be an underestimate by a factor of 2 to 4, based on the assumption of certain track line detection (Mullin & Fulling 2004). Our carcass-recovery rate for deep-diving whales would then be biased high by a factor of 2–4.
additional information. If, for example, 101 cetacean carcasses were recovered overall, and all deaths were attributed to oiling, the average-recovery rate (2%) would translate to 5,050 carcasses, given the 101 carcasses detected (Table 1). As the necropsy results emerge, we can evaluate whether this prediction is high or low, but the sheer scope for underestimation builds a compelling case, in our view, for additional work. The vast majority of carcasses recovered appear to have been bottlenose dolphins. As necropsy results emerge and the need for recovery plans debated, we encourage such discussions to explicitly take into account the probability that the number of dolphins stranded represented something on the order of only 2% of the number of animals killed. The potential is high for the spill to have caused catastrophic impacts on small, localized populations of bottlenose dolphins in the Gulf. We note that coastal and offshore forms of bottlenose dolphins are found off California, with the coastal carcass having a 50-fold greater probability of stranding than an offshore one (Perrin et al. 2010).

Even in the case of EVOS, the large number of observed deaths was acknowledged to represent only a fraction of the total mortality (Estes 1991). Two approaches were taken to estimate total mortality in Prince William Sound: (1) a comparison of pre- and postspill population size; and (2) extrapolations from recovered carcasses to total mortality from a multiplier based on the probability of recovering a carcass (Garshelis 1997). Our estimates of carcass-recovery rates were calculated from the best available data, but we caution against using historic (i.e., pre-spill) carcass-recovery rates to generate a simple multiplier to assess total mortality in the Deepwater Horizon/BP oil spill, but also have global relevance for litigation and marine conservation policy. Given that environmental restitution in the United States can be based on a violation system (Alexander 2010), carcass-recovery rates must be explicitly considered when evaluating the impacts of such disasters. In the case of EVOS, legal damages placed the value of each sea otter killed at US$80,000, or the cost of rehabilitating each oiled otter (Estes 1991; Garshelis 1997). In terms of broader recommendations for marine policy, we note that carcass counts are used in many countries, including the United States, to monitor human impacts on cetacean populations. The tools that managers use in the United States to estimate and limit the impacts of human activities on stocks rely upon “potential biological removal” (PBR), a calculation that determines how many animals can be removed from a stock before causing harm. The PBR estimate, under the Marine Mammal Protection Act (MMPA) depends on reasonably unbiased and precise estimates of human-caused mortality (Wade 1998). In contrast, the effects of many human impacts are only witnessed opportunistically, such as a carcass being discovered on a beach. The issue arises when policymakers, legislators, or biologists treat these carcass-recovery counts as though they were complete counts or parameters estimated from some representative sample, when in fact, they are opportunistic observations. Our study suggests that these opportunistic observations should be taken to estimate only the bare minimum number of human-caused mortalities. This work suggests that carcass counts alone
are unreliable indicators of either natural or anthropogenic sources of mortality. It is vital to develop a framework that explicitly accounts for the low probability of recovering carcasses, if we are to accurately assess the sustainability of all cryptic forms of human-caused mortality.

Human impacts on marine ecosystems and marine mammals are growing both in type and scale (Kraus & Rolland 2007; Clausen & York 2008; Duce et al. 2008; Doney 2010; Hoegh-Guldberg & Bruno 2010; Tittensor et al. 2010). Establishing the proper spatial and temporal scales, at which to assess the impacts of acute events, is further complicated by potential long-term effects and a lack of basic population-specific information (Bejder et al. 2006). This highlights the need for long-term population monitoring, such as that mandated by the U.S. MMPA (Bonebrake et al. 2010). In the first year after the 1989 Exxon Valdez spill, the AT1 group of “transient” killer whales experienced a 41% loss; there has been no reproduction since the spill (Matkin et al. 2008). Although the cause of the apparent sterility is unknown, the lesson serves as an important reminder that immediate death is not the only factor that can lead to long-term loss of population viability. The recent disaster in the Gulf of Mexico provides an important opportunity to assess whether or not the intensity of monitoring conducted in the Gulf of Mexico is sufficient to detect even catastrophic effects (Taylor et al. 2007b). If line-transect survey data are found wanting, we see value in exploring new passive acoustic monitoring methods to detect trends in relative abundance (Marques et al. 2009, Rojas-Bracho et al. 2010). These could be especially useful for rare or pelagic species, or those for which g(0) estimation is particularly problematic. Accurate assessment of impacts also must consider how species are likely impacted, whether acutely on contact with oil or over the longer term through toxicity or habitat degradation (Lovett 2010; Schrope 2010b). If support for longer term assessments dwindles as the time passes and public attention moves elsewhere (Schrope 2010a), then chronic effects will remain unknown. In such cases, only immediately observable effects, such as the number of carcasses, have and will be used to determine the impact of an event, and synergistic and lagged effects will not be considered. Our findings suggest that assessments of the impact of anthropogenic events based solely on the numbers of carcasses recovered are deceptively biased. A better understanding of carcass-recovery rates and the degree to which they underestimate actual mortality, is critical to assessing the true consequences of oil spills and other human activities known to cause cryptic mortality, such as ship strikes, certain fisheries interactions, and acoustic trauma.

Acknowledgments

Lynne Barre, Dee Boersma, and Dave Thompson gave valuable comments at an early stage of the development of this manuscript. We thank Leah Gerber, Tim Gerrodette and Barb Taylor for their careful reviews.

References


Fisheries and Oceans Canada. (2008) Recovery strategy for the northern and southern resident killer whales (Orcinus Orca) in Canada. Fisheries and Oceans Canada, Ottawa, Canada.


