Draft report of the Biologically Relevant Sound Levels Technical Working Group

Part of the 2015–2016 Seismic Code of Conduct Review process

New Zealand Government

Department of Conservation
Te Papa Atawhai

Technical Working Group Members, in alphabetical order:

Natacha Aguilar de Soto (St. Andrews University & La Laguna University); Manuel Castellote (NOAA); Bob Gisiner (IAGC); Russell Leaper (Aberdeen University); Rob McCauley (Curtin University); Jakob Tougaard (Aarhus University); Sander von Benda-Beckmann (TNO); Rob Williams (St. Andrews University).

Additional contributions from: Dorian Houser (National Marine Mammal Foundation)
Support from: Andrew Wright, Dave Lundquist and William Arlidge (Department of Conservation)

Cite as: Department of Conservation (Ed.) 2016: Draft Report of the Biologically Relevant Sound Levels Technical Working Group. Marine Species and Threats, Department of Conservation, Wellington, New Zealand. 57 p.

Disclaimer:
This report of the Biologically Relevant Sound Levels Technical Working Group remains in draft form, as the final language could not be agreed by all TWG members in the time provided. The report as a whole should not be considered the opinion of any individual TWG member, nor should it be considered as a consensus position. Instead, the material contained within the following pages should be considered an attempt by the Department of Conservation to accurately and fairly represent the various viewpoints and differences of opinion expressed within the TWG. As a draft, this report has also not received a final copy-edit from the DOC Publishing Team.

Publishing information:

Author: Department of Conservation (Ed)
Published by: Marine Species and Threats
Department of Conservation, National Office
PO Box 10420, Wellington 6143
Marine@DOC.govt.nz

Contents

Preface: Background to the Technical Working Group

The review of the Code

In 2012, the Department of Conservation (DOC) developed a voluntary Code of Conduct for Minimising Acoustic Disturbance to Marine Mammals from Seismic Survey Operations ('the Code'), in consultation with international and domestic stakeholders representing industry, operators, observers and marine scientists. The Code (and its supporting reference document) aims to provide effective, practical measures to minimise the acoustic disturbance of marine mammals during seismic surveys. It was updated in 2013 after being incorporated by reference into the Exclusive Economic Zone and Continental Shelf (Environment Effects – Permitted Activities) Regulations 2013 ('the EEZ Regulations'; see SR2013/283).

At the time the 2012 Code was implemented, DOC committed to the Code being reviewed after three years. Accordingly, the review of the 2013 Code began in July 2015, with a request for feedback from numerous stakeholders (the Seismic Code Review Group; SCRG). In August 2015, this feedback was combined with that obtained during the three years since implementation.

Role of the Technical Working Groups

In August 2015, DOC established nine technical working groups (TWGs) to address the technical issues raised in the feedback and to provide expert advice on the most suitable methods for addressing them. It was intended that DOC would then draw on this advice when redrafting the Code. The TWGs were:

1. Marine Mammal Observer/Passive Acoustic Monitoring Requirements
2. Marine Mammal Observer/Passive Acoustic Monitoring Observer Data
3. Marine Mammal Impact Assessments/Marine Mammal Mitigation Plans
4. Consultation Requirements for Operators
5. Sound Propagation and Cumulative Exposure Models
6. Acoustic Ground-truthing
7. Non-Standard Surveys
8. Non-Commercial Surveys
9. Biologically Relevant Sound Levels

The work of these TWGs was supplemented by two workshops that were co-hosted by DOC in association with scientific conferences in 2015, to discuss the appropriate mechanisms to facilitate the integration of methodological and technological advances into the revised Code.

The nine TWGs worked until January 2016 to provide feedback on the issues assigned to them. This is the report of the ninth TWG: Biologically Relevant Sound Levels.

Scope of work for the Biologically Relevant Sound Levels TWG

New scientific information on the impacts of noise on marine mammals has become available since the 2012 version of the Code was produced. Any effort to minimise disturbance to marine life from sound should consider the biological relevance of different types of exposure. Accordingly, this TWG will consider the current state of scientific information, and provide guidance on the extent to which sound characteristics, levels, frequencies and exposures (single, cumulative, etc) may influence the various biological consequences of exposure for the animals concerned. This TWG will consider all levels of impact, from disturbance to direct acoustic trauma, and will provide scientifically based expert opinion on the importance of the outcomes for individuals and populations.

The output of this TWG will help guide DOC and the Steering Group to make wider policy decisions in the Code pertaining to thresholds – for example, in relation to establishing mitigation zone sizes. It is noted that DOC and the Steering Group must also consider other factors in making ultimate decisions, including socio-economic and cultural drivers. Advice will be delivered in the form of a report containing two or more options (where appropriate) for addressing the specific issues raised in this subject area in the revised Code.

Accordingly, this report discusses:

1. The implications of the various relatively recent updates to international guidelines and related science including, but not limited to:
   - The USA's acoustic injury criteria (NOAA 2015a[1]; but noting the reservations of Tougaard et al. 2015)
   - Ireland's seismic survey guidelines (Department of Art, Heritage and the Gaeltacht 2014)
   - Greenland's seismic guidelines (Kyhn et al. 2011)
   - Germany's pile-driving regulations (Werner 2012)
   - Denmark's pile-driving guidelines (Energinet.dk 2015)
   - The potential for multiple behavioural response thresholds (eg Blackwell et al. 2015)

   with particular consideration for:

   1. The merits of an acoustic footprint-based approaches (eg Germany's threshold of 160 dB re 1 $\mu Pa^2 s$ / 190 dB re 1 μPa at 750 m for pile-driving) v. exposure-level

[1] Since this report was compiled, the NOAA criteria have been finalised. They, and the associated Federal Register notice addressing raised issues and limitations, are available here: www.nmfs.noaa.gov/pr/acoustics/guidelines.htm.

threshold-based approaches (eg shutdowns when marine mammals come within 500 m):

a) The definition and application of mitigation zones

b) The use of soft starts

c) The integrity of sanctuary/reserve-based protection given the propagation of sound

2. Given that the Code is entitled 'The Code of Conduct for Minimising Acoustic Disturbance to Marine Mammals from Seismic Survey Operations':

   a) Provide an indication of how well 'acoustic disturbance' is minimised by the current provisions and, if necessary, suggest options for ways this could be improved

   b) Provide options for a working definition for what constitutes a 'disturbed' animal from a biological perspective

3. With due consideration to the International Whaling Commission (IWC) report on the potential impacts of multibeam sonar (http://iwc.int/2008-mass-stranding-in-madagascar), discuss the need and value of including non-seismic sources in the Code.

4. Provide opinion on the utility and practicality of frequency weightings appropriate to the species in New Zealand waters.

5. Provide opinion on the utility of soft starts and the most biologically appropriate way of implementing them.

6. Provide opinion on what requirements should be put in place for equipment that is designed to detect marine mammals acoustically.

Part 1: Introduction

1. Scientific data, interpretation and scope

The aim of this Technical Working Group (TWG) was to address concerns raised during the review of the Code of Conduct for Minimising Acoustic Disturbance to Marine Mammals from Seismic Survey Operations ('the Code') that related to its basis in biology. Although the nature of the underlying document crosses science and policy, this TWG was tasked to consider only the available science. This was not an easy task, as in several important areas the available scientific information is very sparse. Accordingly, disagreements arose over appropriate extrapolations, inferences from the available data and courses of action in the absence of reliable information. The following sections outline these discussions and disagreements. However, an attempt was made to limit the conclusions to statements based on underlying scientific information and, if necessary, acknowledge that science is not yet able to adequately advise the wider policy process of the review of the Code.

Some members of the TWG noted that seismic survey noise is known to have impacts on species other than marine mammals (including, but not limited to, fish and sea turtles), and expressed concern that the Code did not consider or address such effects. However, the following text is limited primarily to a discussion of the impact of seismic surveys on marine mammals due to the focus of the Code and the associated scope of the TWG deliberations.

2. Report structure

This report provides expert opinion on the issues put to the TWG as part of the scope of work. It is not intended to be a comprehensive work on the subject area. As such, it covers a wide range of topics, often without substantial introductory text. It is thus not expected to be suitable for readers that do not already have some familiarity with the concepts involved.

Part 2: Bringing science into management

3. Management approach

3.1. Source/threshold-based v. receiver/exposure-based approaches

One fundamental question relates to the approach that DOC should be adopting. One option is an approach based on a measure of the acoustic footprint (such as the German threshold of 160 dB re 1 $\mu Pa^2 s$ / 190 dB re 1 µPa at 750 m for pile-driving; see Werner 2012). However, any thresholds that are defined for source levels may need to be revised dynamically as new mitigation technologies are developed that allow reduced source levels while maintaining effectiveness. The other alternative is a receiver-based approach (eg shutdowns when marine mammals are exposed to seismic sounds within 500 m of the source). The answer to this question seems to be tending towards a combination of actions that both manage source characteristics, such as the source level or frequency structure, and mitigate effects at the receiver, such as through soft starts, shutdown or power-down thresholds, and monitoring activities (eg observers and passive acoustic monitoring – PAM).

Several approaches can be taken to manage the sound input into the water from a seismic source, which can relate to either the characteristics of the source or its operation. It was agreed that whatever management options are implemented, these should have clearly stated objectives. These include not only biological or environmental objectives, but also operational objectives – such as the quality of geological data, availability, cost, safety and reliability of operation. The development of a set of recommendations for source modifications will require a multi-disciplinary team with expertise in hearing, behaviour and general biological issues, as well as acoustics, geophysics, engineering and geophysical survey operations.

From a biological perspective there are many factors to consider, such as the occurrence/concentration of protected species (eg marine mammals) in the target seismic area at the time of the survey. One common and widely agreed objective is to reduce the risk of direct injury to the ears or other anatomical structures. Exposure level or 'safe range for hearing injury' criteria are based on an estimated distance at which a source at level X is expected to have declined to a 'safe' range, which is usually interpreted as being not directly injurious on hearing. A safe range has the advantage of being relatively easier to implement than received level values, which involve a complex dynamic interaction between source movement, animal movement and properties of the transmitting medium.

One proposed option for 'safe' levels (upon which 'safe' ranges might be based) was the newly revised draft US National Oceanic and Atmospheric Administration (NOAA) regulatory criteria (NOAA 2015a; 'draft NOAA (2015a) criteria'[2]) for direct hearing injury (other types of injuries are not assessed). The safe range values for these new hearing risk thresholds are quite short, sometimes in the range of only tens of metres, even for very large sources – although NOAA does acknowledge that these have been extrapolated from experiments based on a limited number of species/individuals/sources due to data limitations and the related uncertainties. The short spatial range derived from these thresholds means that use of a practicable criterion (eg 500 m, 1 km) may be acceptable to the operator, and will have the advantage of familiarity and simplicity while still providing some buffer before an animal would actually be within a range that is expected to pose any risk of hearing injury. However, it was noted that managing seismic operations through shutdowns is only intended to reduce the risk of auditory injury and does nothing to reduce (and may even slightly increase) other potential impacts, including disturbance and masking. This is because shutdowns simply displace seismic activity in time, rather than reducing it overall. In fact, shutdowns increase the total duration of source introduction, due to the need for additional start-up periods and overlap with already collected line data during a back-fill shoot.

The other approach involves managing source characteristics (level, frequency structure, duty cycle, etc) to the parameters necessary for the job, being a desirable goal to reduce overall acoustic energy input in the ocean by the seismic survey as much as it is practical for seismic data acquisition. The main advantage of this approach is that altering source parameters, including at frequencies and in directions that are not used for geophysical purposes, will consequently reduce the overall potential for adverse effects, although it is understood that this is not necessarily a simple linear relationship.

3.2. Effectiveness of current mitigation measures

Neither visual observers nor PAM are perfect detectors of animals,[3] so their value for mitigation really depends on how much risk reduction is obtained from their use. Using simulations of the mitigation process, Leaper et al. (2015) made a case for there being few instances where mitigation actions based on visual sightings could achieve a greater risk reduction than a 3 dB reduction in source level. This is due to the assumption that there would be a low probability that a species would be detected visually at a range that was sufficient to prevent high exposures, where 'high exposures' translates into some level of biological consequence. However, it was noted that the actual demonstrated effectiveness of observers varies widely from near-0% to near-100% across species and conditions.

---

[2] Since this report was compiled, the NOAA criteria have been finalised. They and the associated Federal Register notice addressing raised issues and limitations are available here www.nmfs.noaa.gov/pr/acoustics/guidelines.htm.

[3] It was noted that the effectiveness of PAM depends on the equipment and methods used, such as the number of channels, capacity to localise, deployment method (towed or over the side), monitoring technique (cabled or wireless remote), sensitivities, sampling rates, frequency responses, processing software and auto-detection methods, as well as training level of the operator to recognise the vocalisations of different tax. Some minimum technical specifications are required for effective implementation, such as the American National Standards Institute (ANSI) standards for operating towed PAM operators during seismic surveys that are currently under development. However, it also needs to be noted that the effectiveness of PAM also relies on factors beyond human control, such as whether an animal is vocalising or not.

These results change with increasing detection ranges and can also differ with other sensor, such as passive acoustic monitoring (PAM) for species such as sperm whales (*Physeter macrocephalus*) or some species of baleen whales whose signals propagate well for long distances, in some areas and times when they are vocally active.[4]

Despite their imperfections, the relative cost and ease of implementation of marine mammal observers (MMOs) and PAM warrant their continuation as mitigation measures, to gain whatever incremental reduction in risk they can be demonstrated to provide.[5] Similar arguments also apply for soft starts (ramp-ups), which on first consideration seemed like a relatively simple and easy mitigation method, but may not be having the desired effect of moving animals away from the source before it goes to full power (see Section 7 for further detail). However, some TWG members noted that soft starts may be more useful for species displaying greater sensitivity than others to anthropogenic sound (more broadly), such as beaked whales or porpoises, which have been reported to respond at distances of several kilometres from different noise sources (eg Tougaard et al. 2009; Moretti et al. 2010; Tyack et al. 2011; DeRuiter et al. 2013; Miller et al. 2015). (A more detailed discussion of soft starts can be found in Section 6.)

3.3. Protecting marine mammals outside mitigation zones

Part of the challenge with addressing impacts outside mitigation zones is finding common ground on what is an acceptable behavioural effect. It is conceivable that any audible sound could elicit a behavioural response, so although reports of behavioural responses to received levels of sound near ambient baseline at 100 or 120 dB of sound pressure level (SPL) (eg Miller et al. 2014, 2015) are plausible, it is not known if they are biologically significant. Such significance will be influenced by a number of factors including the level of repetition of disturbance and on the species'/individual's sensitivity to noise (Aguilar de Soto & Kight 2016). Behavioural avoidance of a critical feeding area for days or weeks obviously has high potential for leading to biologically meaningful fitness consequences for that individual and, in turn, for the population to which it belongs (eg see Claridge 2013; King et al. 2015). A proposed alternative, which is widely used to measure the physiological effects of chemical pollutants, is to monitor thresholds for a specific response (eg a behavioural response such as cessation of diving for marine mammals) and then plot the probability of that response against the measured or estimated received level to create something like a clinical dose-response function to a pollutant (eg Miller et al. 2014). A regulatory threshold is then based on some arguably arbitrary point along the dose function, eg a 90%, 50% or 10% response, which is the proportion of responding individuals that is deemed to result in a significant impact on the population. This has been applied to marine mammals only recently, and requires

[4] Hence, there is an argument for focusing more attention on source-level reductions rather than operation procedures using exclusively visual detection methods, as has most frequently occurred in the past.

[5] Some in the TWG felt that resources directed at poorly performing detection efforts might be better channelled into developing quieter source technologies or other more effective mitigation tools. Others noted this implied some sort of levy that represented an entirely different approach, and that the two elements (detection, and research and development) should be kept separate.

substantial research in order to identify suitable behavioural response criteria for different species (eg beaked whales may dive deeper instead of ceasing to dive).

This approach does not explicitly consider the context in which the exposure occurs, which may have a large impact on exposure tolerance, sensitisation and habituation (eg Ellison et al. 2012; Goldbogen et al. 2013). However, if the sample size in terms of numbers of observed exposures is large enough, the entire range of activity states may have been sampled. 'Context' in this instance will simply increase the observed variance in the dose-response curve – which may explain why some studies with larger sample sizes have larger confidence intervals (CIs) than those based on smaller datasets (eg Williams et al. 2014). On the other hand, the sample size for studies requiring, for example, playback of seismic sounds to tagged animals will be rather small. Some TWG members noted that the context of the exposures in one location and study may also be rather similar, while others countered that behavioural states may still vary substantially. Context may be incorporated into the assessment by including two (or more) probabilistic thresholds (eg 80% probability of 50% or more of the animals responding), if such data are available.

Combining dose-response functions with a consideration of foraging or breeding areas/periods may produce better tailored thresholds for behavioural responses. However, some TWG members noted that this presumes we have sufficient information about sensitive activity states. Additionally, the presumption that breeding and feeding grounds can be clearly separated may be more accurate for some species (eg many baleen whales) than others (eg many coastal dolphins). These TWG members accepted that we can more generally develop mitigation tools that are most precautionary in the places where animals engage in their least-tolerant activities if we have: (a) spatially explicit data on how animals use their habitat; and (b) information on the activity state in which animals are most vulnerable to disturbance (eg Ashe et al. 2010). However, they also noted that these approaches require substantial data. These TWG members noted further that caution should also be taken when using the apparent tolerance to noise to indicate impact, as some baleen whales are known to deflect easily during migration, while others tolerate much greater levels of noise when feeding conditions are good. Thus, effects that have no outwardly observable behavioural indicator, such as masking or physiological stress responses, may also need to be factored into estimates of consequences beyond direct hearing injury (see Section 6).

Finally, the duration of the disturbance is also likely to be an important parameter. One useful consideration is the number of days over which sound in a given area exceeds certain levels, which can help in assessing the cumulative effects from different sources and for setting target exposure levels over time. Attention needs to be given to the appropriate unit of time under discussion, as while surveys may last for months, they may pass a specific location only once during that time. They might alternatively only approach locations again within a few km on successive legs of the survey at intervals of half a day to several days. Some TWG members believe that there is no substitute for a careful time/area modelling of the actual tracks of source vessel or vessels to best express the likely pattern of ensonification for a specified location of interest. Other TWG members noted that the European Marine Strategy Framework Directive (MSFD) provides and alternative approach. The MSFD proposes the number of days that a given area experiences low frequency intense pulses (such as seismic pulses or explosions) as one indicator of Good Environmental Status (Descriptor 11).

Arriving at meaningful and widely acceptable thresholds for behavioural effects will be a real challenge for New Zealand and the expert community. It should be noted, however, that there is an ongoing effort to produce behavioural response criteria in the US.

3.4. Assessing the proportion of the population affected

Assessment of the anticipated extent of impact at the population level is a controversial topic. Some TWG members emphasised methods that are currently in widespread use, including individual-based models of sound exposure coupled with models of the subsequent population effects, such as the Population Consequences of Disturbance (PCOD) Framework (eg New et al. 2015). However, others noted that there are issues with these approaches, including difficulties in accounting for cumulative effects and incorporating population-level consequences when these are extremely difficult to detect empirically in the field. Such effects may accumulate over multiple surveys or sound sources, such as pile-driving, military activities, shipping, etc and multiple stressors can also interact with each other (eg the effects of seismic survey sound during an El Niño/El Niño Southern Oscillation (ENSO) event may be greater than in another year).[6]

Instead, it was argued that if seismic surveying is a persistent, regular activity in an area, observer data, more frequent dedicated biological surveys and other types of monitoring may provide a more sensitive indication of population health than has historically occurred due to infrequent and irregular resources for dedicated population surveys. However, others in the TWG pointed out that the detection of trends in a population is unlikely to allow the consequences of a threat to be detected in time for management action to be taken that will prevent its serious depletion (eg see Taylor et al. 2007).

Some in the TWG noted that, from a health and welfare perspective, the total number of individuals that are likely to be affected by injury or displacement is also an important consideration, in addition to any population-level effects. Thus, estimating the number of individuals that are likely to be exposed to certain sound doses that can be related to responses (eg injury or displacement) is a useful first step when considering a proposed seismic survey. The consequences of different operational strategies (eg survey timing, source characteristics, operational procedures) could then be investigated, to provide estimates of the expected numbers of individuals that are affected at each dose level for each option. There is still a need to be able to relate the relative consequences of, for example, a certain behavioural response compared with an auditory injury for an animal, which might be addressed within the framework of modelling exercises such as PCOD.

Some TWG members noted that individual-based models such as the Acoustic Integration Model (AIM; eg DeAngelis & Grimm 2014) are capable of accounting for the accumulated exposure to multiple sound sources and over long periods of time, although others argued that these may be often based on unrealistic assumptions of animal distribution and movements. With regard to the difficulties of detecting the population-level consequences of sound in the field, it was acknowledged that they may be hidden by

[6] As an aside, some in the TWG noted that this implies that the regulators should play a role in managing sound or human disturbance generally, and distributing any adverse socio-economic impacts equitably and fairly across all stakeholders.

other factors such as climatic fluctuations (eg ENSO), disease, predation, fishery bycatch, etc. In addition, marine mammals are long-lived and so small effects at the population level may not be detectable for many years or even decades. Some argued that this is a persistence of monitoring issue, rather than a problem with the models per se (ie that a model is only as good as the underlying data).

It was, however, noted that the limited data available in New Zealand (especially in comparison to other wealthy countries; see Kaschner et al. 2012) would reduce the confidence in any model results in comparison to what could be achieved with a good understanding of the distribution and habitat use of marine mammals. Estimating the proportion of a given population that is impacted also requires spatially explicit abundance and density estimates, which are lacking for most of the marine mammal populations in New Zealand waters. Thus, there is a need for solid pre and post-seismic survey monitoring as part of the protocol of any seismic activity in New Zealand waters. A solid data collection procedure is also needed to identify whether an activity triggered displacement or other changes in spatial distribution at a scale that is much larger than the seismic survey data acquisition zone.

3.5. The Code should not address multibeam sonar

Considering the International Whaling Commission (IWC) report on the potential impacts of multibeam sonar (eg Southall et al. 2013), the TWG discussed the need and value of including non-seismic sources in the Code. It was agreed that there is a need to evaluate the risk from other sound sources and, depending on the outcome, to develop regulatory structures for managing those risks. This task would best be undertaken separately from the seismic Code of Conduct, however, as there would be concerns around delaying or making overly complicated the procedures for regulating seismic surveys by including other sources that differ in their mode of operation, reason for use and environmental impacts.

4. Hearing and frequency weighting

4.1. When and how to apply weighting

There are three underlying questions to the issue of weighting:[7]

- Should we use weighting?
- Which curve(s) should we use?
- How do we define the curve(s)?

It was agreed that weighting functions provide a useful way of broadly accounting for the hearing sensitivities of different species across different frequency ranges, especially for the assessment of cumulative exposure using some sort of sound exposure level (SEL)

[7] Weightings are essentially filters that represent the relative sensitivity of the ear to different frequencies included in a sound to better assess the way it would be perceived by a given listener.

metric (it is less important for considering peak pressure impacts). Assuming that the ears are the most sensitive organs for noise, both with respect to direct injury and behavioural responses, there is a clear argument for frequency weighting as a principle,[8] as the noise can only inflict damage in the acoustic system and/or elicit behaviours if it reaches the inner ear. Moreover, in the absence of weighting, there may be little or no incentive to develop sound sources that are intended specifically to reduce sound in frequencies outside those of operational use to seismic surveys, such as the Teledyne Bolt eSource.[9] This is because the main energy of seismic pulses is typically at low frequencies, and so the selective removal of higher frequencies will have almost no effect on the overall broadband source level; however, the effect of modifying the source should be clearly reflected in the weighted source levels, at least for odontocetes, which have greater sensitivity to sound at frequencies higher than those used in geophysical imaging, but which are a by-product of the bubbles produced by compressed air sources. Despite this, the TWG agreed that a weighted received level is more closely linked to hearing damage (temporary or permanent threshold shifts: TTS/PTS) than to the level of behavioural responses.

In summary, weighting is useful for predicting hearing damage and some behavioural responses, but animals may also be physiologically sensitive to frequencies outside their best hearing range. Thus, reducing high frequencies from seismic sources will likely offer some general benefit, and not only for those marine fauna with peak hearing sensitivities at these high frequencies.

4.2. Hearing damage

The weighted levels that induce TTS have been reasonably well studied, at least in certain species. This means that they are suited for most considerations of auditory injury.[10]

Since the TWG agreed that weighted levels will generally best reflect the likelihood of hearing damage, the question becomes largely practical: which weighting curve(s) should be used? M-weighting curves were developed based on a hypothetical link between loudness weighted levels and injury/behavioural reactions. However, their use has been criticised because it has not been demonstrated experimentally that their levels correlate with the degree of injury and/or severity of behavioural reactions. Thus, work to modify the weighting functions is currently underway. The latest draft NOAA (2015a) criteria weighting functions are another step forward, but are still extrapolating beyond our current knowledge on very few species to all marine mammals. Some in the TWG noted the inherent risk of this extrapolation, which urgently needs to be addressed with newer empirical data.

---

[8] It was, however, noted that stress responses can have strong physiological consequences on wellbeing and reproductive success, or elicit behaviours resulting in homeostasis alteration (eg beaked whales and sonar). These responses may happen at levels well below hearing injury.

[9] http://www.teledynemarine.com/eSource.

[10] Some TWG members noted that hearing directivity remains a complicating issue, and that there is increasing scientific evidence that infrasonic noise below hearing thresholds can also affect the inner ear of humans and experimental terrestrial mammals in experimental situations (eg Salt & Hullar 2010; Kugler et al. 2014) – suggesting that this may also be possible in marine mammals.

Some TWG members noted that, given the limited available data, the choice of weighting curve would need to be carefully considered for each species and also made open to change. For example, it may be most appropriate to use the porpoise weighting function for Māui dolphins (*Cephalorhynchus hectori maui*) at present due to their comparable acoustic capabilities, but the as-yet undefined function of a Commerson's dolphin (*Cephalorhynchus commersonii*) might be more appropriate in the future, as they are more closely related and have similar acoustic capabilities. Nonetheless, cross-species extrapolations need to be carefully considered in all instances. For example, we know that high-frequency-group harbour porpoises (*Phocoena phocoena*) are more sensitive than mid-frequency-group bottlenose dolphins (*Tursiops* spp), but can that relative sensitivity be extended to all other high and mid-frequency species? However, others in the TWG noted that hearing abilities tend to be taxonomically conservative, with all or almost all canids (dogs), felids (cats), otariids (sea lions and fur seals), phocids (true seals) and delphinids (medium to small dolphin species) exhibiting very similar hearing abilities. While the only high-frequency specialist tested to date is the harbour porpoise, its anatomical and sound production similarities to other species such as Māui and Commerson's dolphins suggest that its weighting functions are likely to be the 'best available' fit to New Zealand's as-yet untested species. This designation is also considered a precautionary assignment, as high-frequency cetacean hearing risk thresholds and weighting functions are considerably more conservative than those of other marine mammal hearing groups.

The TWG then discussed the use of auditory brainstem response (ABR) v. behaviourally acquired data in the construction of weighting curves. While these techniques are actually measuring different things, they both assess just different aspects of the auditory processing of sound. However, some TWG members noted that ABR tends to underestimate sensitivity at the ends of the ranges. Others noted that there is an extensive literature on the relationship between evoked potential audiometry and behavioural measurements (traditional hearing testing) both for animals in general and for marine mammals in particular. While such a discussion may be beyond the scope of the Code of Conduct review, they believed it would be misleading to suggest that the actual specific relationships between evoked potential (EP) audiometry have not been well-explored, with over-simplifications such as 'ABR underestimates sensitivity at the ends of the ranges' not adequately characterising the scientific knowledge on this topic. This issue remained unresolved within the TWG.

4.3. NOAA 2015 draft criteria[11]

Endorsement of the weighting functions in the draft NOAA (2015a) criteria was proposed, but there was no agreement in the TWG. It was noted that this remains in draft and, while it was not anticipated that the recommendations would change greatly, some flaws or potential improvements had been identified. Some in the TWG also noted that more comprehensive approaches have been advised (eg see Weir & Dolman 2007; Wright et al.

---

[11] Since this report was compiled, the NOAA criteria have been finalised. They, and the associated Federal Register notice addressing raised issues and limitations, are available here www.nmfs.noaa.gov/pr/acoustics/guidelines.htm.

2013; CBD 2014; Nowacek et al. 2015; Prideaux & Prideaux 2015; Erbe et al. 2016). Other TWG members were not convinced by the proposed evidence of possible 'flaws' and strongly believed that the draft NOAA (2015a) criteria represent the 'best available' consensus among the expert scientific community, as reflected by the two very thorough external peer reviews. It was also noted that the current draft NOAA (2015a) criteria address only auditory injury, which is defined as a PTS by NOAA. Some in the TWG believe that since the ear is the most sensitive organ in the body to trauma from acoustic exposure, mitigating PTS (even imperfectly) is a reasonable approach to reducing all injury from acoustic exposures. Others noted that PTS is used as it is easier to assess than injuries to other organs. The TWG did not agree that using PTS onset as an indicator of onset of injury was precautionary, but did agreed that the approach NOAA took to define PTS onset (as TTS onset plus 34 dB) was precautionary. However, some in the TWG noted that there was still ongoing scientific discussion about using PTS for defining criteria and that more conservative thresholds have been proposed (eg Tougaard et al. 2015).

While the TWG agreed that the review contained within the pages (and references) of the draft NOAA (2015a) criteria outlined much of the best available science about hearing damage, there was less agreement on the application of that scientific information to the formation of the criteria. For example, there was disagreement over the species grouping extrapolations, although it was acknowledged that some extrapolation was necessary. The extrapolation of the considered data to baleen whales (low-frequency specialists) was particularly contentious among the TWG, although it was noted that there are efforts underway to improve this. An updated final version of the draft NOAA (2015a) criteria is anticipated to be released in the next two months.

Other areas of disagreement included the use of the criteria for assessing cumulative impacts and the time window over which sound exposures should be aggregated. With regard to the latter, one TWG member noted that any one SEL is driven largely by the 5–10 loudest exposures over any given timeframe. This is of particular importance for moving sources such as a seismic vessel, whereas the cumulative exposure from stationary sources, such as pile-driving, can be significantly influenced by the total duration of the exposure.[12]

The TWG also discussed the concept of a transition from impulse to non-impulse sound, which is outlined in the draft NOAA (2015a) criteria.[13] There was general agreement that this was not supported by the available literature and only served as an added complication to the criteria. Additionally, it was noted that the criteria did not account for recovery between intermittent exposures, sensitisation or physiological exhaustion following several exposures or directional hearing, which remains an issue.[14,15]

---

[12] Although not included in the draft NOAA (2015a) criteria, some in the TWG noted that the relative movement of the animals is relevant as well; more mobile animals will be less affected by a stationary source, but increase the possibility of encountering a moving source elsewhere at some later time.

[13] It should be noted that this is no longer included in the final NOAA guidelines.

[14] Some in the TWG noted that this is not a realistic source of exposure differences for animals and sources that are changing their orientation to each other over very short times.

[15] One TWG member also noted that recent work on humans has indicated the existence of cochlear phenotypes, whereby different cochleas display similar-shaped audiograms, but have different absolute sensitivities.

Regardless of the limitations, it was agreed that the draft NOAA (2015a) criteria, including the weighting functions, are likely to become widely used as a way of assessing the number of animals likely to be subjected to hearing-related takes (ie TTS/PTS) as a result of seismic surveys. However, some TWG members noted that the New Zealand Code of Conduct could specify using these as a benchmark for comparison, but that this should not preclude assessments based on other criteria, if there was evidence to suggest that these alternatives may be more appropriate.

4.4. Behavioural responses and other sub-lethal impacts

The appropriate method for applying weighted thresholds to behavioural responses is not clear. For example, some signals may induce behavioural responses incrementally with increased weighted level, while others may be perceived as particularly threatening or attractive, leading to avoidance or strong approach at the lowest detectable levels. Social facilitation is also a consideration for these highly social animals, as it may only be necessary for one individual in a group to respond to a sound for a response to be induced in the whole group, or for some in the group to ignore a sound to cause others to abandon their initial inclinations to respond. Furthermore, it is quite reasonable to expect that marine mammals may be able to determine and consider the distance to the source, regardless of the received level. Finally, background noise levels will influence the signal excess and thus signal detectability, and ultimately also the potential response.

Some TWG members suggested that weighting functions should not be applied to behavioural thresholds or raw impulse threshold criteria like received sound pressure levels (SPL, typically measured as peak or root-mean-square levels), but only to criteria based on frequency-dependent energy flux within the inner ear, such as SEL-based criteria.

When sound becomes information rather than being treated simply as mechanical energy, loudness scaling becomes difficult: sounds with high levels of relevance or salience to the animal may elicit a response at low levels, while sounds of no relevance may fail to elicit a response until they are quite loud or nearby. The fact that different behavioural responses may occur in different individuals at different received levels, or even in the same individual at different motivational states, can perhaps be better handled by a probabilistic expression of onset, such as a dose-response function. There are alternatives to a dose-response function, but something that captures the variability in responsiveness is probably better than setting the threshold at 100% response or 50% response, which either leaves a significant portion of the population omitted from risk estimates or creates a huge overestimation of risk.

It was noted that a correlation between audibility and behavioural response thresholds has been convincingly demonstrated, at least for harbour porpoises (Tougaard et al. 2015). The assessment of behavioural reactions is admittedly more complex than for acoustic injury, however, with other factors related to the type of signal, context and previous experience of the animal needing to be factored in. However, this does not change the fundamental argument that a certain signal-to-noise ratio is required for the animal to perceive and identify the signal before it can react to it, and that this ratio will to a large degree be determined by the audibility of the signal, ie its level above the hearing threshold (for low noise conditions), which is also known as the signal excess. The argument also goes the other way: if a signal is inaudible to an animal, it cannot elicit a behavioural response in that animal. Whether audiogram weighting or loudness-based weightings are appropriate is a secondary question that needs to be settled through experiments.

For New Zealand species, there is particular concern for the Hector's (*Cephalorhynchus hectori*) and Māui dolphins, as they are classified as high-frequency cetaceans according to Southall et al. (2007), and so may be more sensitive than mid-frequency cetaceans. A precautionary approach would be to also use weighting curves from harbour porpoises for these species. However, given the distant phylogenetic relationship between porpoises and Hector's/Māui dolphins, this may not be appropriate. The conservation status of these two species may prevent extensive studies from being conducted to determine appropriate weighting curves; however, it may be viable to conduct such studies on closely related species of a more favourable conservation status, such as Commerson's dolphin. The application of surrogate hearing information when data are deficient or non-existent is a common practice (eg Southall et al. 2007; NOAA 2015a).

As in the case of effects of noise exposure on hearing physiology, one additional consideration is the duration of exposure. Behavioural responses will have very different long-term potential impacts from a single explosion that lasts some tenths of a second than from a single seismic survey that can last several months. Therefore, a metric for duration is needed in parallel with a loudness metric. One possible metric could be cumulative SEL, but metrics relating to the statistical distribution of sound levels, such as percentiles (eg L50 and L95) may be more appropriate to capture the stochastic nature of behavioural responses.

Some TWG members noted that one of the most closely studied populations in the presence of seismic survey and other noise-generating activities has been the feeding aggregation of gray whales (*Eschrichtius robustus*) off Sakhalin Island. A rapid decline in the relative abundance of whales in the Sakhalin near-shore feeding area during September 2008 was temporally coincident with a seismic survey and the resumption of on-land pile-driving. Recent analyses (Cooke et al. 2015) lend support to the hypothesis that reproductive success was indeed impacted by this disturbance, to the extent that it fell to near or below the replacement level in that year. Such detectable demographic consequences of disturbance only become apparent some years after the event, and also require a very well-studied population to allow these effects to be detected. However, others noted that, as is often the case because of inherent difficulties, other relevant environmental variables were not tracked, and that attribution of reproductive fluctuations to sound alone remains hypothetical.

Given the various uncertainties and unknowns, the TWG agreed that they were unable to provide any scientifically based advice on weighted thresholds for behavioural responses at this time. However, it was noted that some numerical weighted noise level will underpin the construction of highway noise suppression structures for humans, despite the variability in the extent of disturbance and annoyance. This number would have been set by policy makers with some legislative goal in mind. Consequently, avoidance was suggested as a possible proxy for disturbance onset in marine mammals (if disturbance was the target, rather than behavioural responses). However, the counter point was made that the duration of avoidance would also need to be considered.

Part 3: Establishing and applying impact thresholds

5. Acoustic injury

5.1. Defining 'injury'

5.1.1. Physical damage from noise exposure

As the most obvious impact of noise pollution is auditory fatigue, current efforts to regulate noise pollution often focus on identifying the sound levels that produce measurable noise-induced hearing loss. It is not yet clear how permanent hearing loss may develop from repeated episodes of TTS – whether it takes large TTS episodes of 40+ dB to produce the effect, or whether cumulative years of smaller TTS episodes might accrue over decades into the kind of old-age presbycusis we quite often see in older humans and dolphins. The difficulty is that there are other contributors to cumulative hearing loss, such as disease, ear parasites, injuries or other sound sources (man-made or natural). Thus, it may be difficult to make a case for a TTS-based metric for 'injury' from individual acute events such as a single seismic survey. Accordingly, some in the TWG felt that a single occurrence of TTS should not be considered an 'injury.' However, others felt that the possibility of neurological damage following large TTS (Kujawa and Liberman, 2009) combined with the fact that PTS may result from multiple TTS episodes means that TTS cannot be completely discounted (see below for more information). There was, however, general agreement that everyone would like to avoid TTS if possible.

While there was an agreed desire to eliminate vague wording in standards, wording that encourages progress in both understanding and mitigating the risk factors is not a bad thing. We know that our current understanding of the risks from sound is not perfect, and nor is our ability to mitigate those risks. Therefore, encouraging further investigation into the risks and the means of mitigating them should be an incentive of a regulatory process, and should be rewarded by a flexible, adaptive regulatory regime rather than discouraged by overly rigid regulatory processes. Dynamic and adaptive management is gaining increasing popularity for a reason – it makes both the regulator and the regulated parties active agents in the reduction of risk and the improvement of monitoring and mitigation capabilities.

As scientific knowledge of acoustic impacts has increased, there has been a great deal of concern about low-frequency sonar sound causing resonance of the lungs, sinuses or other air spaces, and vestibular and sensorineural effects. Some TWG members felt that despite extensive investigations using laboratory animals, human volunteers and marine mammals, no such effects have ever been found. They reasoned that the ear remains the most sensitive and thus most likely site of onset of injury to the best of our current scientific understanding. Consequently, they suggested that direct acoustic injury (PTS and blast trauma) was the injury effect that it was most desirable to mitigate and for which we have the strongest scientific data. Other TWG members argued that direct hearing damage (including TTS) and barotrauma are simply the effects that we have the strongest scientific data about (see review by Aguilar de Soto and Kight, 2016). However, they noted that it remains essential to mitigate these effects as much as possible, especially permanent damage.

Some members of the TWG argued that PTS as it is defined in the draft NOAA (2015a) criteria should be considered the best conservative indicator for the onset of risk of direct injury at this time. In support, it was noted that these criteria use an estimate of PTS that is the minimum difference (not the maximum or average) between TTS and the onset of some small amount of permanent hearing loss across a narrow frequency band.

However, these opinions were not shared by other members of the TWG, who argued that TTS is an injury in its own right and that various non-auditory impacts have also been documented. For example, human divers have been shown to experience short-term negative effects from noise exposures, such as tingling and a loss of tactile sensation due to overexposure of the pressure-sensitive corpuscles of Pacini in the dermis, pain in the digestive system due to gas vibrations, and low or moderate TTS, despite some of the affected divers subjectively describing the received sound exposures as low level (Steevens et al. 1999; Fothergill et al. 2001). Noise-exposed divers also suffered short and long-term neural damage, which was expressed as headaches, dizziness, blurred vision and other symptoms apparently linked to effects in the vestibular system. It is unknown whether marine mammals have adaptations to cope with the non-auditory physiological effects of noise, and it was noted that there has been insufficient investigation of the possibility of other injuries in marine mammals. Until such knowledge exists, these TWG members believe that there is no scientific basis indicating that hearing damage is the only possible injury caused by noise. However, they did acknowledge that the ear is the indicator of acoustic injury that is easiest to assess.

5.1.2. Damage mediated through behaviour

Irrespective of that discussion, it was noted that one potential exception to injury being defined exclusively as PTS might be the effect of mid-frequency sonar on beaked whales. There has been a long discussion about whether that effect is directly physiological, a physiological injury caused by a behavioural change, or simply a behavioural 'panic' reaction that results in animals finding themselves in shallow water or on the beach and succumbing to hyperthermia and other stranding-related effects. However, some TWG members noted that there is evidence that the same pathologies have been found in beaked whales that mass-stranded alive and those found dead some 100 km offshore following naval sonar exercises (Fernández et al. 2005, 2012), indicating that the pathologies are not due to death following stranding alone.

Some scientists believe that the stranding of beaked whales could be due to physiological issues, such as fat or gas bubbles in the tissues caused by disrupted diving patterns (Fernández et al. 2005; Cox et al. 2006). The array of pathologies observed in the beaked whales mass stranded in the Bahamas and Canary Islands in 2002 and 2004 suggests there were injuries in addition to those that are typical of the physical effects of stranding itself (Anon. 2001; Fernández et al. 2005, 2012). It has been proposed that this is due to the particular adaptations of beaked whales to extreme diving, in comparison to other cetaceans of a similar size (Aguilar de Soto & Kight 2016).

Despite these data, some TWG members argued that there is currently insufficient evidence to place confidence in any of these hypotheses. Furthermore, this effect has only been demonstrated for strandings of beaked whales and mid-frequency sonar, despite some reports linking strandings of other species to other sound sources (eg Southall et al. 2013). The sonar–beaked whale relationship has been known of for more than 15 years, and yet no comparable type or quantity of evidence is forthcoming for other species or sound sources. Other TWG members noted that the large scales over which disturbance occurs means that the existence of such relationships cannot be ruled out through the available data (eg Jepson et al. 2013). However, some TWG members noted that this is a variant of the 'proving the negative' problem, as the causes of more than half of all strandings around the world are unknown (see D'Amico et al. 2009 for a discussion of the level of coincidence in the sonar–beaked whale stranding relationship). Consequently, they suggested that, for the most part, thresholds of injurious effects are most easily based on indicators of the onset of hearing injury such as TTS or PTS, as they provide quantitative, replicable metrics for creating exposure limits, expected safe ranges and other impact-mitigating measures.

5.1.3. Injury onset

While it was agreed that TTS or PTS provide useful criteria for the onset of hearing injury for the time being (as described above), the TWG could not agree on which should be used. The TWG also agreed that it is not presently possible to rule out other effects that are likely to be classified as injury, but may be occurring at lower received levels of sound, as few studies have yet to assessed them.

5.2. Applying injury to mitigation zones

Proceeding without an agreed definition, it was suggested that direct hearing injury was the effect easiest to mitigate because it is the effect for which we have the strongest scientific data. Some TWG members noted that the information on PTS is largely derived from TTS data, which would be a more conservative level for mitigation. If it proves possible to effectively mitigate these impacts, it would have obvious benefits to both the individual and the population. The distance at which such effects may take place can be used as a metric for observer and PAM monitoring and shutdown criteria. In contrast, the ranges from the source at which behavioural effects may occur can be considerably greater than for directly injurious effects. At these larger ranges, the most commonly used mitigation methods, such as visual observers and power-downs or shutdowns, will not provide the same magnitude of reduction as they do for injurious effects on hearing. Indeed, some TWG members noted that in many cases there may be no reduction in risk. The same would be true for effects with no observable behavioural indicator, such as masking or physiological stress symptoms.

Others argued that such disturbances may not be the effect we most wish to mitigate, noting that other mitigation techniques may be required to achieve this. However, it was agreed that there is currently no way to rank the relative severity of a small (behavioural) impact on a large number of animals compared with more severe (injuring) impacts on fewer animals. Therefore, it was recommended that this should remain a priority for research. It was also recommended that DOC should define clear management objectives for the Code to allow the development of a (longer-term) management strategy to reach those objectives.

6. Disturbance

6.1. Addressing 'acoustic disturbance' under the Code

In general, it seems likely that the following impacts may occur with increasing seismic sound exposure levels:

- Behavioural responses, masking and stress responses
- TTS and potential non-hearing physiological injuries
- PTS and potential non-hearing physiological injuries

The required exposure levels mean that animals would generally need to be very close to suffer from a PTS, relatively close to suffer from a TTS, and potentially at large distances away to be subject to masking, behavioural responses or stress responses. However, this also depends on the duration of exposure, as it is possible for exposure to be accumulated up to the onset levels. With the possible exception of masking (and to a lesser extent behavioural responses), all impacts would be expected in closer 'regions' to the source.

All of these impacts represent a disruption to the normal functioning of a marine mammal and could thus potentially represent a disturbance. However, the vast majority of impacts will lie in the first category (ie behavioural, masking, etc), which relates to fitness impacts that are currently challenging to incorporate into models, impact assessments and management decisions. This difficulty does not mean they can be ignored. Furthermore, a much larger proportion of the population will be affected to this level than by TTS or higher, which suggests that it may be a much more important issue when it comes to population-level consequences and survival.

Some TWG members noted that the idea that seismic sound 'can' plausibly lead to certain effects often results in the belief that it 'will' definitely cause those effects, without any evidence to support this. Accordingly, they argued that the comment made in the review that 'any estimated impact [is] likely to be an underestimate' is simply not logical or realistic. Instead, they suggested that the comment implies an infinite threshold for the amount of mitigation effort that should be undertaken. Furthermore, while investigating plausible impacts as identified in other species in similar situations, we should not regulate activities or manage our ocean resources as if every imaginable outcome is accepted without supporting evidence. Otherwise, we will find ourselves in a situation where a plausible story about a certain outcome proffered by one person might be treated with equal weight as multiple demonstrated instances of that outcome not occurring.

However, others in the TWG argued that the available evidence was more than just 'a plausible story' and felt that such situations called for balanced management of the potential for risk. They suggested that the application of a precautionary approach with some studies showing evidence of impacts at different levels places the burden of proof on those conducting an activity that is suspected of having a plausible risk to the environment in the face of scientific uncertainty, ie it is upon them to demonstrate that such harm is not occurring. If further scientific findings emerge that provide sound evidence that no harm will result, established protections can then be relaxed.

While some TWG members noted the scientific basis for applying the precautionary principle in maintaining viable populations, others felt that decisions about this should lie with policy makers and are thus beyond the scope of this TWG. They suggested that the precautionary principle is in direct opposition to principles of scientific critical thinking and parsimony – one can be a scientist or be precautionary, but not both. Other members of the TWG indicated that the later affirmation contradicts modern scientific literature on conservation biology (eg Thompson et al. 2000). There was no resolution of these differences within the TWG, and this fundamental difference underpinned many of the disagreements outlined within this section. Substantial discussion arose over consistency in the use of anecdotal evidence, as well as the appropriate level of weight that should be given to such information when making decisions. Despite these disagreements, the TWG as a whole suggested that the Code should be structured in a way that the various uncertainties are taken into account and the various assumptions are explicitly stated to make it clear what needs to be updated and how this should proceed once relevant new information is available.

6.2. Defining 'disturbed'

It is well documented that some animals change their vocal behaviour, alter their dive behaviour, or move away or orient towards a source in response to a sound exposure or some other disturbance, even if there are inconsistencies across species, locations and contexts. The US Marine Mammal Protection Act is quite broad in defining what constitutes a Level B 'take' by harassment, and so, by those criteria, we might expect thousands of 'behavioural disturbances' from activities such as people walking on a beach where seals haul out, driving boats in and out of dolphin resting areas, and undertaking whale watching tours. However, when it comes to permitting the disturbing activity, the US regulatory agency has consistently concluded that all of these behavioural disturbances have no significant or meaningful biological consequence under either the Marine Mammal Protection Act or the Endangered Species Act (for those species covered by both legal requirements).

Some TWG members argued that what constitutes a 'disturbed' animal from a biological perspective is exactly what was laid out in the National Research Council's (NRC 2005) report on the Population Consequences of Acoustic Disturbance (PCAD, later developed into PCOD) – an effect that adversely affects life functions such as survival and reproduction. Since we cannot follow every individual through its entire life to determine if and when an acoustic disturbance altered its biological success, we use actuarial tables just like we do for humans when looking for the effects of asbestos, automobile seat belts or taking supplemental multivitamins. However, it is not yet possible to assess the population-level consequence of behavioural reactions to seismic surveys due to the practical difficulties involved in obtaining the necessary data. Accordingly, we do not know whether there is any value in attempting to mitigate behavioural disturbance, which is sometimes actively sought as part of mitigation, such as when soft starts are used in an attempt to move animals away from the loud source that will follow. However, some concern was expressed by other TWG members that it is in these situations that management should be driven by the precautionary principle, and regulation as well as policy makers should stimulate further research to better address the lack of scientific knowledge.

The TWG was unable to reach a resolution on this issue, but continued discussions on some of the various elements of 'disturbance'.

6.3. Behavioural responses

If behavioural disturbance leads to an adverse effect, this should be detectable using a well-designed and likely long-term study. However, no such studies have been completed to date for assessing the impacts of seismic surveys.

However, some TWG members argued for caution here, stating that all of the current seismic noise mitigation actions and monitoring plans do not specifically address long-term consequences of behavioural responses. Behavioural disturbance is expected to occur at large distances from the noise source, while all of the efforts to observe marine mammals are focussed on the close-range, to support power-down, shutdown and ramp-up procedures. Furthermore, the occurrence of behavioural disturbance and its potential adverse effects has never been monitored by industry. It is only in the past few years that dedicated projects (eg behavioural response studies) have been developed to better address this gap. There are many challenges in conducting experiments that adequately assess the behavioural reactions of whales to seismic noise. These include the need to obtain an adequate sample size with the necessary controls and to measure the range of variables likely to affect the observed response. Analysis is also complex. Well-designed experiments are complex and logistically difficult, and thus expensive. Consequently, caution should be taken to avoid translating the lack of evidence of a behavioural reaction into a lack of negative effects. Multiple studies have documented spatial displacement as a behavioural reaction to seismic survey noise (for a review, see Castellote & Llorens 2016) and efforts are underway to better quantify behavioural reactions under certain conditions (Dunlop et al. 2015).

Some TWG members pointed out that disturbance as evidenced through behavioural responses has been demonstrated and that the Code states an intent to reduce such disturbances. However, there was consensus in the TWG whether simple detection of behavioural changes was sufficient to make fully informed management decisions. Therefore, there is a pressing need to assess the consequences of these changes; for example, how many dives need to be lost to reduce pup weight gain in a pinniped? If beaked whales are displaced to a non-preferred foraging site, but only need a couple of additional dives to meet their energetic requirements, is that a problem? The ongoing work of the behavioural response group in the USA was noted, along with the difficulties of operationalising available data for meta-analyses of such responses.

Integration of the ultimate impact of behavioural responses into analytical frameworks remains a challenge. Behavioural responses to other sources, such as naval sonar exposure and multibeam sonar, have been linked to strandings (Evans & England 2000; D'Amico et al. 2009; Southall et al. 2013). With regard to seismic surveys, it has been suggested that behavioural responses caused ice entrapment in narwhals (*Monodon monoceros*; Heide-Jørgensen et al. 2013) or potentially caused a stranding in beaked whales (Barlow & Gisiner 2006). However, some TWG members argued that, in any case, displacement inevitably carries energetic costs and potential lost opportunity costs, noting that whether these trigger negative consequences for survival and reproduction depends on the spatial and temporal severity of the disturbance, as well as the ecological context of the disturbed population. Although there have been no direct estimates of such impacts arising from seismic surveys, estimates have been made for other sources of marine mammal disturbance (eg Williams et al. 2006). Avoidance-induced temporary habitat loss as assessed for porpoises in the presence of pingers (eg Tougaard et al. 2009) or for beaked whales in the presence of naval sonar (eg Moretti et al. 2010) is another potential problem that has not been assessed and that probably cannot be assessed in New Zealand waters at this time given the lack of information on species distributions.

Ultimately, the TWG agreed that there is plenty of evidence that animals can change their behaviour in response to any audible sound. However, disagreement persisted over exactly what this ultimately means. Much of this seemed to be based on the fact that no definition for disturbance is in place. While there was agreement that a response carrying a biologically significant impact would be considered a disturbance, determining when a biologically significant impact may occur is extremely difficult if not impossible in many cases. There was less agreement over whether quality of life should be incorporated into a disturbance definition. However, it was noted that this is often linked to health, which in turn is linked to fitness, and so the end result may be the same.

6.4. Stress responses

6.4.1. The importance of stress responses

With regard to other sub-injurious impacts, it is possible that exposure levels that are capable of inducing avoidance would also be capable of inducing a stress response – especially in cases where avoidance was not possible or desirable. Changes in the levels of stress hormones have been correlated with changes in ship noise (Rolland et al. 2012), though the significance of those changes is currently unknown. Furthermore, although indications of stress responses to acute noise exposures have been inconsistent, they have been observed (eg Romano et al. 2004), though again with unknown consequences. However, the stress-related consequences of repeated, long-term noise exposure (eg increased mortality, reduced fertility, suppressed immune function) are known to occur in humans and other mammals. It has been speculated that this may have caused the observed differences in population structure between beaked whales inhabiting a naval sonar-training area and a relatively disconnected neighbouring area in the Bahamas, where both the density and number of calves in relation to adult females were lower in the disturbed area (Claridge 2013). It is important to compare this data with that from beaked populations in other areas to gain further knowledge about how different from 'usual' the population structure observed in the naval range area was. Some TWG members felt that it was important to note that it was difficult to say how such responses in other species to other stressors at different levels and durations might relate to the specific case of seismic survey sound. However, others countered that cross-species extrapolations had been used in other important, but controversial, documents relating to this issue (eg Southall et al. 2007; NOAA 2015a).

Avoidance behaviour is often used as one measure of response, and indeed often considered a 'mitigation method' (ie animals move themselves out of 'harm's way').

However, it is difficult to determine whether an animal chooses to remain in an area because the sound is not disturbing it, or because the area is important to it (eg Rolland et al. 2012 regarding right whales, *Eubalaena glacialis*, in the Bay of Fundy). In any case, some TWG members argued that there are several lines of evidence that such a lack of response could still result in sub-lethal effects, such as stress responses (eg Wright et al. 2007), and so we cannot and should not assume that simply because an animal remains in an area it is not being disturbed. Some TWG members noted a remarkable example recorded in the 1990s, where humpback whales (*Megaptera novaeangliae*) failed to abandon a foraging area when blasting occurred in a bay and did not show any overt behavioural signs of distress; however, when one of these whales became entangled in fishing gear, pathological analysis showed evidence of blast injury (Ketten et al. 1993; Lien et al. 1993). Other TWG members argued that this example refers to an animal being in close proximity to an unexpected and rare explosion that would induce actual mechanical impact from a shock wave, rather than a typical source of sound that might be anticipated and avoided. However, some TWG members countered that example demonstrated the lack of displacement after the fact, presumably because they had a strong motivation to be there (feeding), and that the example clearly supports the notion that disturbance can occur without observing a reaction.

In response to the wider point around responses in general, some members noted that many possible non-injurious consequences are not actually realised. Accordingly, while there should be some degree of caution in the face of uncertainty, equal weight should not be given to speculated minor to non-existent non-injurious impacts and demonstrated hearing effects. Following this reasoning, it would thus be hard to 'mitigate' behavioural effects when the anticipated outcome of the mitigation is so uncertain: what changes would we expect to see if mitigating over large distances?

However, it was countered by others that the finding that whales progressively decrease diving (and thus also foraging) rates as they come closer to an array is not a minor impact (Miller et al. 2009), although the fitness consequences remain uncertain. Likewise, some TWG members noted the possible role of seismic surveys in the roughly 50% decrease in sperm whale population numbers in the northern Gulf of Mexico between 2004 and 2009, when the last survey was conducted and prior to the Deep Water Horizon spill. While such impacts more likely relate to aggregated seismic survey activity rather than individual surveys, they still need to be mitigated. However, at least one TWG member noted that sperm whale population estimates in the Gulf of Mexico have fluctuated between 590 (1994) and 1665 (2004), and yet clearly the population will not have grown threefold in one decade, just as it was unlikely to have declined 50% in the 5 years between 2004 and 2009. As stated in the 2015 US National Marine Fisheries Service (NMFS) Stock Assessment Report, 'It should be noted that since this is a transboundary stock and the abundance estimates are for U.S. waters only, it will be difficult to interpret any detected trends' (NOAA 2015b). Thus, the movement of sperm whales in and out of the US Exclusive Economic Zone (EEZ), which is the area to which surveys have been confined, obscures any real trends in the data (NOAA 2015b). This clearly demonstrates how difficult it is to assess temporal changes in the abundance of cetaceans to detect significant declines with certainty. Some members of the TWG argued that this underlines the need for a precautionary approach within realistic management needs.

6.4.2. Stress responses as a metric for disturbance

Regardless of the current state of evidence, the idea of using stress responses as a metric for disturbance was discussed, as chronic responses may have fitness impacts (Claridge 2013). The TWG agreed that a sufficiently large number of small, individually negligible effects may accumulate to have population-level impacts over time, but that stress-based metrics of disturbance are not quite available. Quantifying such accumulations will likely require dose-response functions and probabilistic methods that could incorporate motivational state and other contextual elements. However, this would require large amounts of data to cover the range of species (or even species groups) present in New Zealand waters, else the level of uncertainty would render estimates completely uninformative.

Furthermore, the technologies for assessing stress and placing the information in the context of baseline levels of stress are not yet mature enough for application, according to Mike Weise, Office of Naval Research (ONR) Marine Mammal S&T Program Manager and the sponsor of most of the current research on this topic (pers. comm.). Two papers exemplify this issue. The first, by Romano et al. (2004), looked at trained Navy dolphins that were exposed to high levels of simulated sonar sound, while the second, by Rolland et al. (2012), monitored cortisol and other stress hormones in right whale faeces and exhalations during the temporary decline in shipping traffic during the 9/11 terrorism response. Rolland et al. (2012) found a statistically significant decline in cortisol, while Romano et al. (2004) did not. The PCAD/PCOD model for consequences of disturbance incorporates stress hormones or other condition indicators as a potential indicator of the likelihood of downstream life history and population consequences. However, stress responses are natural and can be a healthy part of an animal's physiology when they are not extreme, which seriously complicates the issue, especially since comparative baseline data in cetaceans are not yet available. At this point in time it would thus be difficult to prescribe which data to collect and how to collect them – and to then put the results into a baseline context of fluctuating stress levels where a relative difference of any given amount was considered biologically significant.

While stress responses cannot yet be used as a metric of disturbance by seismic sound, they should prove a useful tool in the near future and so their use should be encouraged by regulators. Similarly, some TWG members made the point that it is still possible to manage and mitigate for an impact, even if you cannot measure it. It was also noted, however, that it is difficult, and in many cases impossible, to evaluate the effectiveness of mitigation measures.

6.5. Masking

Masking is the obscuring of signals of interest by noise. The case for using masking as a means of measuring impact comes primarily from papers by Clark et al. (2009; but see also Møhl 1980, 1981), in which a simple equal energy model for masking is applied to a scenario and any sound with more energy in the frequency band of interest than the signal is assumed to have masked the sound. This simple engineering model of masking is widely used for mechanical, engineered acoustic systems, but fails to capture the considerable capabilities for masking release of the mammalian auditory system. Clark et al. (2009) did mention some of these adaptations for masking release, which can give an animal some gain in signal detection within correspondingly louder noise, but these adaptations are seldom factored into the resulting calculations of the acoustic footprint of a noise source or effective communication range for signal detection.

Some TWG members suggested that a more realistic view of the effectiveness of various man-made sounds as maskers comes from work by Erbe et al. (2016) and Branstetter et al. (2013), who investigated phenomena such as comodulation release from masking. Similar work on ice seals is currently being carried out by Reichmuth et al. (unpubl. data) and should be published soon. The upshot is that the short, intermittent impulse of seismic sound is a less effective masker than continuous 'pink' or 'white' noise such as ship noise (Erbe et al. 2016). Even at distance, where reverberation and refraction can 'fill in' the space between primary pulse returns, the amplitude modulation of the seismic sound may fail to completely mask signals of interest that are more than 20 dB quieter than the seismic sound in some situations (Erbe et al. 2016), while in other cases, reverberation may impede detection at low frequencies. This means that the potential for full masking of the signal will be highly influenced by the substrate, slope and other parameters governing sound transmission in each habitat. It is unknown to what extent marine mammals require the full information contained in any given signal for it to be effective, making the impact of partial masking of a signal unclear.

Some TWG members believed that equal energy models of masking greatly overestimate the amount of masking that is likely to occur and thus the need to set mitigation distances for the hypothetical masking effect. They ascertained that the effectiveness of seismic sound as a masker of biological sounds is substantially lower than a continuous broadband signal of equivalent source level for normal communication or echolocation ranges, whether an animal is within tens of metres of a seismic array, where the pulses are short and discrete, or many kilometres away, where the sound has spread over time. Signal redundancy, listening out of the band of interest for harmonics, directional hearing adjustment by movement and other normal behaviours can mitigate communication and echolocation masking by seismic sound if animals make use of these.

However, others pointed out that partially masked biological signals could be inefficient at transmitting the codified information. A whale song note or a dolphin whistle might convey information that can only be interpreted if the full signal is received. In any case, whales modify their acoustic behaviour, with other possible consequences, well before masking levels are reached (eg Castellote et al. 2012; Blackwell et al. 2015).

Some members noted that seismic sound is concentrated in the low-frequency range with peak energy below 200 Hz (Tolstoy et al. 2004), making baleen whales more likely to experience considerable overlap in frequencies between seismic sounds and signals of interest for communication. However, others noted that pinnipeds also use sounds in this range (Erbe et al. 2014) and even though the peak at close range may be below 200 Hz, there is still significant energy up to a few kilohertz, which is well above ambient, out to great distances (Madsen et al. 2006).

It was also argued by some TWG members that much attention has been given to the potential audibility of seismic sound hundreds or thousands of kilometres from the source, but this audibility is confined to specific oceanographic conditions that typically occur below the dive depth of baleen whales (only the low-frequency sound is transmitted across those ranges). (This statement was disputed by other members of the TWG, however, who had recorded seismic pulses with an array towed at some 20 m depth from a ship at some 500 km from the seismic source location, which was provided by official sources; Aguilar de Soto et al. 2004). Furthermore, it was argued that this scenario begs the question of what two whales might communicate to each other across hundreds or thousands of kilometres? Accordingly, these TWG members believed that the likelihood of masking occurring at all or occurring over durations and distances that had a biological impact is close to zero, and so should not be considered a plausible outcome to be addressed by mitigative actions.

By contrast, other TWG members argued that the long-range disturbance of acoustic space for a large number of individuals should be considered the main cause for concern rather than the long-range communication between individuals. Furthermore, seismic surveys can last for months with an acoustic footprint that covers vast areas around the source, and in most cases are also recurrent in the same region (eg 4D surveys). Accordingly, these members argued that the potential for biologically significant masking effects is quite high and that the duration of the airgun array shot acoustic signature should not be considered the main feature to consider when addressing masking. While airgun pulses are obviously impulsive and short, they occur in sequences, meaning that instantaneous masking effects might also be small; however, the cumulative effect of masking over the long term may be more considerable. The first group of TWG members countered that the relevant concept here is duty cycle, or the percentage of time with noise present versus noise absent. For seismic sound, the duty cycle is 0.1%, and even with long-distance effects of multi-path and reverberation, the amount of time when seismic noise would exceed the signal would only be in the order of 1–2% (eg Breitzke et al. 2008). As is well understood from literature on human speech, and bird and frog song, the listener is able to fill in missing pieces of a signal to reconstruct the masked portions, especially when the fraction of masked signal is so small (Bregman 1990). However, the second group of TWG members were of the opinion that the percentages of time mentioned above may be higher (eg McCauley 2015), depending on the signal of interest (eg baleen whale signals would be more prone to being masked), and the source-receiver distance and sound transmission characteristics in the area. For example, Guerra et al. (2011) showed that for one seismic survey there was "…no time period between airgun shots when reverberation does not mask one's ability to detect and recognise calls."

Seismic surveys can last for months and, because of their design, airgun shots are concentrated in space along tightly defined acquisition transects. Thus, even if the source is moving, the acoustic disturbance is concentrated in time and space at the masking level. However, some TWG members noted that this will depend on the particular threshold for masking and the characteristics of the masking source, meaning that the acoustic footprint of the seismic survey will move and change shape over time, as the animal moves, changes orientation and detects signals between pulses. This problem cannot be simply modelled by a 2D disk, but must be modelled in the full 3D world of the listener, and the 4D time series over which the animal and masker are moving and changing orientation.

If masking occurs in areas where the use of sound carries specific biological importance, such as feeding or breeding habitats, the potential for negative effects is heightened. However, some TWG members noted that longer signals are actually less-easily masked, since it has been shown that some animals can reconstruct missing pieces from the parts that are not masked (see Bregman 1990; Katz et al. 2014), suggesting that this may also happen in the case of marine mammals. The most common masker of signals of biological interest are competing conspecific signals (eg other calling male humpback whales or other echolocating members of one animal's feeding group), which means that animals may have developed means to deal with masking from natural signals. Other members noted that the additional masking from days, weeks, or months of impulsive signals that can reverberation (potentially resulting in nearly continuous low-frequency noise) could further obfuscate information codified in song bouts, phrases and syllables. For all these reasons, these TWG members argued that masking of biologically significant signals by seismic survey noise should be considered as a plausible mechanism, with potential for both acute and chronic negative impacts.

In summary, the TWG acknowledges the extensive literature that demonstrates masking empirically. However, there is currently no information available on masking effects from large numbers of impulsive sounds under real-world scenarios. Accordingly, the TWG agrees that much work remains to be done before the biological significance of such masking at the population level can be determined.

6.6. Disturbance and mitigation zones

6.6.1. The merits of current mitigation for reducing disturbance

Some TWG members noted that Stone (2015a, b) analysed more than 15 years of observer data to statistically document the effectiveness of UK Joint Nature Conservation Committee (JNCC) mitigation efforts for short-range effects. The effectiveness of the JNCC guidelines (JNCC 2010) has previously been discussed by several authors (eg Parsons et al. 2009) and many countries add measures beyond them (eg Greenland: Kyhn et al. 2011). It was argued that while behavioural disturbance cannot be prevented with the current JNCC (2010) guidelines, simple actions such as posting observers and implementing shutdown criteria have been shown to greatly reduce the chances of animals coming too close to seismic sources and thus suffering impacts to hearing. Whether the implementation of guidelines to reduce behavioural reactions at a long-range scale has any meaningful impact at the population level depends on whether these disturbances have any biologically meaningful consequence in the first place.

Other TWG members noted that in many scenarios, particularly for inconspicuous species, posting observers and implementing shutdowns does little to reduce the risk of injury and has no benefit in reducing disturbance.[16] Some maintained that these mitigative measures probably have some limited benefits. However, other TWG members noted that it must be remembered that the spatial extent of this potential disturbance means that there is no relatively low-cost mitigation measure that could provide complete assurance in the face of uncertainty about whether biologically meaningful consequences may or may not be occurring. For example, it must be remembered that MMO coverage is barely enough for hearing injury and is definitely inefficient for mitigation of behavioural disruption (to whatever extent is considered biologically significant).

[16] This could be improved by increasing the probability and range of detection of these species.

These TWG members argued that the current Code does little to reduce 'acoustic disturbance', in that the central premise to the main mitigation efforts is around ensuring that acoustic injury is avoided, which is important but primarily accomplished by inducing documented and/or presumed avoidance behaviours using measures such as ramp-up. Larger mitigation zones around the seismic vessel provide greater minimisation of acoustic disturbance on paper, but are subject to lower detection rates than smaller mitigation zones simply due to the increased distance from the observation point. The effectiveness of PAM also declines sharply with increasing distance, especially for high-frequency cetaceans – plus one potential response to a disturbance is reduced vocalisation rates, further complicating PAM efforts. Even if perfect detection is achieved, the fact remains that these zones are intended to reduce acoustic injury. Behavioural responses can occur at substantially greater distances from the source. For example, recent studies suggest that bowhead whales (*Balaena mysticetus*) vary their calling behaviour in areas that are ensonified by seismic sounds (Blackwell et al. 2015).[17] The calling rates initially increased as soon as seismic sound became audible (ie around a sound exposure level of 100 dB re 1 μPa^2.s), and then plateaued before decreasing over cumulative sound exposure levels of around 127 dB re 1 μPa^2.s, until they became virtually silent at exposure levels above approx. 160 dB re 1 μPa^2.s (Blackwell et al. 2015). These TWG felt that it is important to note that the 160-dB level represents the presumed start of behavioural response in the USA, and yet changes at this represents the extreme of the vocalisation behaviour change in bowhead whales. One very plausible interpretation of these results is that the whales' initial increase in vocal rate is to maintain contact in the face of some disturbance and then they 'compete' with this noise source for some time; however, by around 160 dB, they decide they can no longer compete (ie contact calls cannot maintain contact) and simply stop vocalising to either move away from the source or wait until it subsides. The consequences of a loss of ability to maintain contact is unknown, but such contact may be crucial for maintaining mother–calf cohesion.

6.6.2. The scale of disturbance

Some TWG members noted that several other studies have reported behavioural responses by marine mammals to seismic surveys at long distances (eg hundreds or thousands of kilometres; Castellote et al. 2012; Blackwell et al. 2013, 2015; Cerchio et al. 2014; Pirotta et al. 2014). This means that an unknown number of responses are occurring well beyond the mitigation zone ranges, and so observations from the seismic vessels cannot reliably be used to assess all impacts from the source. Recent work by Robertson et al. (2013, in press) further supports this conclusion, with the finding that behavioural responses to seismic surveys make bowhead whales less detectable to aerial surveys. However, others in the TWG felt that a focus on whether or not animals react behaviourally to the sources detracts from the more important question of whether the reactions actually have biologically meaningful consequences.

These TWG members also noted that while "much is made" of the "potential" that whales could hear seismic sources "thousands of kilometres away" (eg Nieukirk et al. 2012), or

[17] It was noted that these are species that are hunted and live in an acoustically complex environment (due to ice and ice noise) and consequently may not be representative.

that failure to hear fin whales (*Balaenoptera physalus*) on a listening array might mean that they had moved away from an area (Castellote et al. 2012), there is also plenty of anecdotal evidence that animals are seemingly unaffected as they pass close by active seismic arrays (eg Stone 2015a; observations of bow-riding on the arrays; and preliminary results of the recent Behavioural Response of Australian Humpback whales to Seismic Surveys (BRAHSS) study). However, other TWG members countered that initial BRAHSS experiments were made using a 20-in^3 single airgun, yielding a source level of 199 dB re 1 μPa^2.s at 1 m (Dunlop et al. 2015), which clearly is far from being equivalent to a full array that is typically used in commercial surveys. It should also be noted, as discussed elsewhere, that a lack of behavioural reaction cannot be translated into a lack of negative effects on the individuals.

Regardless of the position on the implications of these responses, it was agreed that some disturbance does occur over large distances and that applying current mitigation efforts over such scales would be extremely expensive and may not be effective. Likewise, the reductions in disturbance that are achieved by applying current mitigation methods to only part of this area are likely negligible.

6.6.3. Alternative means for reducing disturbance

These points led to discussions on other ways to mitigate disturbance, such as by limiting the duration of surveys in sensitive areas to reduce chronic exposures and disturbance, although this may have huge costs to industry. Regardless of the approach taken, seasonal habitat use and hotspot information is critical to making informed decisions over the best timing of surveys, implementing time-area closures (with the note of caution that locations of oil interests are not as flexible as those of Navy sonar exercises), or moving towards more comprehensive ocean zoning. Similarly, the species composition in an area must be known to optimise the planning of mitigation zone sizes and other elements. Such information is not yet available to allow DOC to reduce the co-occurrence of surveys and marine mammals in ways that will reduce disturbance without unnecessarily compromising industry. Therefore, the TWG cannot recommend strongly enough that the New Zealand Government actively seeks to acquire the necessary data.

Some members of the TWG then noted that the other way to reduce disturbance is to minimise the sound that is introduced into the water. This could be done by reducing the survey area to only that required (potentially through reducing the number of lines) and minimising the array size. A simple requirement in the Code that numerical justification for array size and line spacing be provided would help to achieve this, as would questioning around the size of the survey area.

Another option is the use of alternative survey technologies, such as gravimetric surveys. Some TWG members felt that there could be some potential here, at least with regard to very preliminary surveys or in terms of reducing the total number of seismic surveys, as the more data (seismic, gravimetric or otherwise) that are available, the better the decisions that can be made on sites of interest. However, others noted that gravimetric surveys are in no way a substitute for the data provided by compressed air surveys, as they complement or supplement either other, rather than being equivalent and thus interchangeable.

A further option is to encourage efforts to minimise the total amount of surveying, by finding ways to reduce redundant or overlapping surveys, or even track lengths for any given survey. However, it was noted by some TWG members that this could be seen as obstructive and thus likely act as a huge disincentive to industry in New Zealand. In any case, the TWG noted their lack of expertise in this area and did not discuss this further.

With regard to improvements of the current mitigation methods that are outlined in the Code, improvements in poor-visibility detection options need to be incentivised in some way to better ensure that injury is not occurring. Such options may include remote PAMs or unmanned aerial vehicles (UAVs), either for direct mitigation or for the collection of data related to the effectiveness of the poor-visibility detection techniques. Such equipment could also be used over the long term (life-of-field) to address some of the persisting questions about the long-term impacts of oil industry activities. Additional lessons can be taken from the Sakhalin seismic surveys, which incorporated several less-typical elements into the mitigation plans, such as pre-survey sweeps along beaches with drones (eg Nowacek et al. 2015). Similarly, NOAA has been using drones to survey ice ahead of its research vessels to improve counts of ice seals, which typically flee to the water as the boat approaches. Similar drone use may help to identify the distances over which marine mammals are avoiding seismic surveys.

It was then noted that these considerations were starting to blend science and politics. If politicians and regulators wish to take issues such as cost into account, they can do that. The task of this TWG is to provide scientific advice on which conservation goals can be based, with more political issues then being taken into account at that point. With that in mind, the TWG agreed that the extent to which seismic survey exposure induces population-level impacts remains completely unknown at the present time. There are simply not enough data to address this subject at present.

6.6.4. Fitness impacts from disturbance

What is known is that there are measured fitness impacts in wildlife exposed to noise (Shannon et al 2015). Some TWG members thus argued that given the observed reductions in foraging rates in cetaceans (Miller et al. 2009; Pirotta et al. 2014), it is only logical to conclude that continued exposure to acoustic disturbance will lead to the same types of fitness consequences in cetaceans as have been measured in numerous studies of terrestrial mammals and birds. However, others disagreed and were not convinced that fitness consequences have been demonstrated or are likely to be demonstrated under extended monitoring. They noted that inference of fitness impacts or lost foraging is not equivalent to a demonstrated effect. However, others countered that many of the PCOD models cited elsewhere in this document are equally inferring a lack of impact and should be treated with the same caution.

By way of conclusion, the TWG agreed that there will always be an impact of seismic activity on marine mammals. However, the real question is a political one: how much of an impact is acceptable? The situation can be more accurately reflected by a continuum of impacts than through any step-function related to received levels. It is clear that all parties want to avoid things like the Navy sonar-beaked whale strandings, but there is less agreement on behavioural responses and there are no clear boundaries on the response continuum.

Similarly, how precautionary DOC needs to be in the absence of information is not a scientific issue and may need to vary from one species to another. For example, there may be a clear management objective for Māui dolphins (like the gray whales of Sakhalin) – no additional impact is acceptable. However, things are less clear for the majority of New Zealand's marine mammals, especially those without a population estimate.

Moreover, in-field mitigation needs to be practical and simple to implement: the US Navy monitors to 2 km as this is realistic. However, monitoring around seismic survey vessels can only attempt to protect a proportion of animals from hearing damage – it can do nothing to lessen behavioural responses that occur at much greater distances than can currently be monitored. Some in the TWG felt that this highlighted the need for pre-survey baseline data and increases in detection distances, such as through using novel technologies (eg Aguilar de Soto et al. 2016).

It was agreed that we should continue to look for fitness and population-level effects. It was also decided that if an effect is plausible based on comparable scenarios in birds or humans in the workplace, we should investigate such possibilities in marine mammals. One approach to assessing this could involve the removal of exposure, in a manner similar to the 9/11 shipping traffic reductions that led to the discovery that right whales exhibited a concurrent reduction in stress hormones (Rolland et al. 2012). A well-designed experimental quiet year (or half-year) in the Gulf of Mexico could be instigated to make a similar assessment for seismic surveys. In the meantime, the almost complete lack of information on the population-level consequences (if any) of exposure to seismic surveys in marine mammals makes it impossible to quantify the need to mitigate those consequences. It was, however, noted that this lack of information does not mean that such impacts do not exist, in just the same way as it does not mean that they do exist.

6.7. Area-based protection zones and sound propagation

It was agreed that a marine mammal sanctuary requires additional consideration beyond what would be applied to the larger ocean area. However, the specifics of that consideration would need to be based on the various characteristics of the sanctuary, including the species present and the reasons for sanctuary designation.

Some argued that there needs to be a substantive basis for assuming that the animals would be engaged in activities that would make them more sensitive within a sanctuary than outside it. The argument was that one cannot simply assume that because it is a sanctuary for marine mammal feeding or fish breeding, seismic sound is going to have a relatively greater or lesser effect than elsewhere, unless there is evidence (not speculation) about the differential consequence of operating within a sanctuary. However, others countered that the existence of the sanctuary represents acknowledgement that the marine mammals in the area are at increased sensitivity, engaging in important behaviours where disruption would have a greater fitness or survival impact, or at an unusually high density.

Regarding the last point, it was also noted that larger numbers of marine mammals would be at risk of injury of behavioural disturbance if a seismic survey was conducted in a sanctuary that had been designated due to high densities of marine mammals that are consistent between years. This could be addressed by seasonal or spatial restrictions on seismic survey activity to prevent surveys from being conducted in high-density areas, and by placing a certain buffer zone around the area to account for propagation beyond the seismic survey track line. However, it was noted that sanctuaries are normally defined with static permanent boundaries, while the distribution of marine mammals is not. Sanctuary-based restrictions would thus be most efficient if boundaries and timeframes were dynamic, although this would require potentially expensive monitoring of marine mammals.

Some TWG members believe that seasonal and spatial restrictions are probably the best available tool for mitigating the aggregated impacts of noise, and that seismic survey exposure should be avoided during critical periods for marine mammal, with year-round, permanent limitations implemented if necessary. They suggested that it is critically important to consider seasonal and spatial restrictions in New Zealand, as they are the only direct way to avoid any impact. As noise travels far and behavioural harassment might happen at large distances, the scale of the restrictions should be adapted accordingly and buffer zones should be considered with progressively increasing mitigation protocols. An example of a permanent limitation to acoustic pollution is the Canary Islands, where naval sonar is banned within 12 nm of the archipelago to protect beaked whales.

Part 4: Biological basis for mitigation measures

7. Start-ups/soft starts

7.1. Overview of soft starts

There is little evidence to demonstrate the effectiveness (or lack thereof) of soft starts in mitigating the overall risk of hearing effects in marine mammals exposed to seismic surveys. Recent studies have provided evidence that some marine mammal species do avoid the source during soft starts, but also that some species or individuals are attracted to either the noise or the vessel. Within and between species variability is to be expected, as some species groups or individual animals may be uninterested in the sound, or have different motivations for remaining in areas of higher noise (eg while feeding or travelling, or with/without a calf). However, the soft start is relatively easy to implement, provides a period of time for monitoring before the full array operation commences, and offers some reduction of risk for those species that generally find sound aversive, such as harbour porpoises and beaked whales or some of the population of less-responsive species. While soft starts might attract some species (sea lions and oceanic dolphins are particularly notorious for this investigative type of behaviour), the attraction also suggests that they are not bothered by the sound and/or are able to manage their exposure by staying near the surface or through some other mechanism.

7.2. Efficiency of soft starts

The premise for imposing soft starts in the first place was that they might provide sufficient time for an animal to move further away from a seismic array before being exposed to potentially injurious effects from the sudden onset of a loud sound. The effectiveness of the soft start depends on all or most individuals of a given species or population responding aversively to the sound as it ramps up, thus removing themselves from the zone of potential injury before full source levels are attained. If, however, the soft start causes animals to approach the source or to behave neutrally, the effectiveness of the soft start is reduced.

Evidence for the benefits of soft starts is still equivocal. Intuitively, it would seem that giving the animals some sort of low-level, non-dangerous cue would reduce the likelihood of sudden high-level exposure that could potentially result in injury or strong behavioural reactions. However, we also know that some animals approach loud man-made sound sources, potentially as an indication of a vessel that they might bow ride to save energy. It is also conceivable that seismic vessels, like vessels in general, may perturb the surface layers to produce mixing and bioconcentration at turbulent interfaces, which animals have learned to associate with increased food availability. Furthermore, there have been several anecdotal reports of large whales approaching the array, investigating it briefly and then moving off. We do not know if they are investigating because it sounds like a conspecific or predator, or if it is simply something unusual that needs to be ‘checked out’.

This does demonstrate though that the behavioural responses to seismic surveys are not only avoidance or silencing, but can vary, and so it may be difficult to arrive at a measure of soft-start effectiveness or behavioural impacts generally, as these may vary by locality, species or some other factors.

In a recent review of a large set of UK JNCC observer data, Stone (2015a) detected a negative correlation between the sighting ranges (proximity to the array) and the numbers of sightings, suggesting source avoidance during full array seismic survey operations and soft starts. This suggests that soft starts can cause an avoidance response for at least part of the population in a wider range of marine mammal species.

However, two subsequent lines of evidence have decreased our confidence in the value of soft-start procedures. First, new metrics for the onset of injury (see draft NOAA, 2015a, criteria) have reduced the area of concern about injury and thus the positive benefits from soft starts relative to the potential negatives, which include an increased time of exposure during the added soft-start time, which may increase the behavioural effects of the operation (though this will be minimal in comparison to the total survey duration). This effect includes the possibility that soft starts may either fail to sufficiently motivate animals to move away, leading to reduced avoidance of the source or even an approach to the source.

The time required for an animal to remove itself from repeated exposures at ranges implied by the new draft NOAA (2015a) injury criteria is very short and probably does not require a full 20–40 min soft start from a single source to the full array. However, some TWG members noted that if sensitive species/individuals avoid the sound (such as is seen in tagged beaked whales at several kilometres from low received levels of sonar exposure; Miller et al. 2015), a longer soft start may allow the animals to increase their distance from the source, reducing their exposure.

The second line of evidence that soft starts may not be as effective as originally thought comes from observer data and controlled exposure studies with soft-start or ramp-up procedures, such as the BRAHSS study of humpback whales in Australia (Dunlop et al. 2016). Similarly, studies of humpback whale responses to sonar ramp-up (Sivle et al. 2015, Wensveen et al. submitted), although employing a different source, different ramp-up protocol and different sound source movement relative to the animal(s) being observed, have produced mixed results, with more animals responding aversively during the ramp-up than were unresponsive or approached the source. These data demonstrate quite wide variability in the responses of humpback whales even after repeated exposures, ranging from the expected avoidance response to no response or even a brief approach to the vessel – and we might also anticipate considerable variation between marine mammal taxa and the behavioural contexts in which the ensonification occurs. The humpback whales that were the subjects of the BRAHSS study showed a general unresponsiveness to sound (albeit substantially downscaled), whereas other species have shown tendencies to be more responsive to sound than most (eg beaked whales, harbour porpoises and killer whales, *Orcinus orca*). Furthermore, in the case of the study of Wensveen et al. (submitted), soft starts seemed to be more effective for mother-calf pairs, which again appears to indicate that they are more effective for the most sensitive animals. Thus, soft starts may be useful for reducing exposure to these more easily disturbed animals.

For humpbacks and other species, it may be that the extent of avoidance in terms of distance and time to a mitigation gun (typically 20 in^3 in BRAHSS) is not much different than the avoidance of the vessel itself, which is announced by the noise of the propeller and engine – a continuous sound source of considerable loudness. In other words, even without any operating seismic sources, the vessel itself represents a considerable source of noise that may produce much the same 'warning' effect as a mitigation gun. Since the ramp-up procedure exceeds the typical mitigation gun source level within the first few minutes (as soon as more than one gun is active), it may well trigger stronger reactions.

Additional quantitative evidence will be needed to decide whether the soft start is an effective mitigation measure or whether vessel noise itself may provide a similar warning to keep animals at a sufficient distance from man-made sounds to avoid injury even without soft starts. Von Benda-Beckman et al. (2014) modelled the interaction between source movement, animal movement and animal responsiveness for a variety of species. Although this model was constructed for typical sonar operating conditions, it offers the opportunity to generalise to other conditions. Their results suggest that shorter ramp-up periods may be worth considering, as movement of the source can limit the effectiveness of long soft-start procedures. However, these results need to be considered in light of the time required by the animals to move away from the source.

Provisions have been made in the USA and elsewhere to exempt animals from mitigative shutdowns or power-downs when the animals deliberately approach the vessel and array, to bow-ride or for other reasons. Such provisions may also be included in the New Zealand Code of Conduct. By contrast, in Greenland, the guidelines state that if marine mammals are observed within 200 m of the array, a power-down is mandatory until the animal leaves this area:

> ***6.6.4 Use of mitigation gun when marine mammals are within 200 m***
> *If marine mammals are detected within 200 metres of the centre of the airgun array whilst the airguns are firing at full power, firing should be reduced to only the smallest airgun in the array, a so called mitigation gun, which should prevent further approach of animals to the array. Full power may be regained as soon as the animals have left the 200 m protection zone.* (Kyhn et al. 2011)

The TWG agreed with the assessment that little evidence is available to show that soft starts are effective overall, although they are likely to offer at least some reduction in overall risk to aversive species. The TWG also agreed that soft starts, or ramp-ups, have not proven to be a particularly costly mitigation measure, and so operators have generally been agreeable to employing them even if the evidence for their effectiveness remains equivocal for all species.

If we begin to see more evidence from our required mitigation monitoring that soft starts have little or no value, they could be dropped without seriously affecting the risk from seismic surveys. In any case, the length of the ramp-up and the structure of it (cf. Dunlop et al. 2016) can and should be reconsidered. The duration of the soft start is often based on the diving time of cetaceans and the argument that any animals that are underneath the seismic vessel should have sufficient time to reach the surface and swim away. Thanks to the sonar behavioural response studies (BSR) studies, we now know that the response by beaked whales to sonar sound is an underwater movement away and a slow surfacing far away from the sound source, although the responses of other species to other sound sources may vary by source and context. Therefore, this knowledge should be incorporated into a re-evaluation of the ramp-up procedure.

7.3. Applying soft starts in the Code

The TWG agreed that the current definition of a soft start ("a gradual increase of the source's power, starting with the lowest capacity gun, over a period of at least 20 minutes and no more than 40 minutes") needs to be improved. For example, a measure of the increase in source level per unit of time should be incorporated. A review undertaken in 2006 indicated that the most common approach is to follow a sound pressure level increase of 6 dB per minute or per 5 minutes, or 500 psi steps (Castellote 2007).

The International Association of Geophysical Contractors (IAGC) recently recommended that doubling the number of sources per time increment is the simplest and easiest way to approximate a 3 dB (energy) step per increment (or 6 dB pressure increment; IAGC 2015) – although, as some TWG members argued, the need for smooth incremental increases has not really been established. As Dunlop et al. (2016) noted, soft starts that attempt to increment by fractions of total array volume or estimated array loudness may actually produce leaps and plateaus, due to the variations in size between individual elements within an array and their geometric interaction as more elements are recruited. However, the TWG consider the proposed IAGC soft starts to be less likely to produce irregular increases over time because the most direct contributor to total array output is the number of elements – more so than the total array volume. The TWG considers that the simplicity of implementation will mean that this is more readily followed by the crews of the vessels than complex schemes that involve activation of different parts of the array based on complex formulas of element size and location.

Regarding the actual mechanism of effect, it is probably best to think of soft starts as a way of increasing the sound source to operational levels while allowing enough time for the animals to move a safe distance away, given the nominal speed of the animals and the speed of the vessel. Some intermediate step between no sound and the sudden initiation of a full array is desired, but whether a specific schedule of increase is more or less effective in moving the animals is probably less important than a schedule that is easy to implement by the operator and which does not lead to sudden large leaps in source level that might trigger a stronger aversive reaction than the minimum necessary to move the animal to a safe distance in a timely manner.

While achieving perfectly smooth increases in loudness is not practicable or possible, it is true that a protocol that produces long periods of no perceptible change interspersed with sudden jumps of 6–10 dB SPL or more may not produce the intended gradual incentive for animals to move away, but rather might elicit startles or pauses in the movement away. Some TWG members pointed out that an animal might alter the perceived level of the sound during the course of its normal swimming and diving, so a jump of 6–10 dB SPL may not in fact be cause for concern about startling or panicking animals. This was an additional rationale behind the IAGC recommendations, which ensure a perceptible increase per increment and increments that produce similar degrees of increase at each step – stepping from 1 to 2 to 4 to 8 to 16 elements and on up will mean that most arrays will complete six or more steps within 20–40 min (at 3.5 – 5-min increments) and that each increment except the final will be approximately twice the sound energy of the prior step (with minor variations due to the range of element sizes being recruited at each step). If the total number of elements is not a perfect multiple of two, then the last step is less than a doubling, but occurs at the end of the full soft-start period.

Procedures in which variable numbers and locations of elements need to be activated at each time step to achieve a doubling of the total air volume per step or an estimated 6 dB SPL increase per step can be difficult for the crews on the back deck to implement correctly – and as Dunlop et al. (2016) noted, may not actually achieve the intended 'smooth' increase anyway due to the complex acoustic interactions between array elements. There is also no real substantive basis for believing that stepwise increases in the source level makes any meaningful contribution to the aversive effect of the soft-start protocol – rather, it may just be the sound itself whether increasing or not, in combination with other factors (eg behavioural context, distance to the source), that produces the avoidance effect, and so elaborate measures to ensure smooth stepwise increases are not really necessary as long as the risk from each of the soft-start transmissions themselves are mitigated by the previous transmissions.

The TWG noted that the IAGC protocol is simpler and easier to explain and therefore also to implement, regardless of whether incremental increases in array loudness add appreciably to the aversive effect of the soft start or not. The procedure starts with the vessel underway at its typical speed of 4–5 kn (2.0–2.6 m/s). After 20–30 min of no detections of protected species by visual observers (or acoustic detections if PAM is in use), the soft start is initiated with the smallest volume element in the array. Another element is then added to double the number, followed by two more for a total of four, and so on until the full array is active. The time increments should be such that the full array is not active until 20+ min after beginning the soft start (eg for an 8-element array, there would be four steps of 5 min each; and for a 32-element array, there would be six steps of 3.5 min each). If at any time during the soft start an animal is observed or acoustically detected within the Exclusion Zone and is not actively approaching the array, the soft start is terminated and another 20–30 min all-clear period is required before starting the process all over again. However, some TWG members noted that the number of airguns active in an array may dramatically modify the directionality of the array, meaning that one airgun may have a horizontal acoustic footprint at the surface that is relatively similar to that of the array. This has been one of the reasons why many scientists challenge the efficacy of soft starts. An alternative method would be to use the full array but at low pressure, and gradually increase the pressure to achieve full power over a given time.

8. Other sound-based mitigation measures

8.1. Mitigation source

Many mitigation protocols allow for the ship to proceed with a single element active during turns and repairs, but only for a limited time, eg 4 hours. Once the repair or turn is complete, the soft start can be initiated without having to wait for the 20–30 min all clear period, since the single element has presumably kept the alerting process going while the full array is not operational and observers have been on duty the entire time. However, the appearance of an animal within the Exclusion Zone at any time during the ramp-up will still lead to a shutdown and a full 20–30 min observation period followed by a full ramp-up cycle for a total of 40–60 min before full array operation resumes. The TWG also agreed that it would be useful to keep one air source active so that the animals know where it is and can decide what to do accordingly. A very well thought out and precise protocol for how to apply the single mitigation source would be required to prevent misuse as a means of avoiding shutdowns at night or the required soft starts. Similarly, ramp-ups and pre-watch periods should be conserved in any case. It was also noted that this should be an airgun and not just the ship's noise – this way, animals will be conditioned to the mitigation gun and learn to associate it with the noise of the full array and hence maintain a safe distance, which may not happen with the ship noise itself. Further, in light of the diving pattern of some species (eg sperm whales have a mean dive of 40 min, or beaked whales of some 60 min) some TWG members proposed that pre-watch periods should be extended and PAM should be implemented during pre-watches given the low probability of visual detection for many species, mainly in sea states > 3.

The suggestion was made that some of the well-tested acoustic deterrent devices (ADDs) used to deter animals from marine piling sites may also be as effective, if not more effective, than a single seismic sound source for some species. It would be important to obtain further data to make decisions about this possibility.

8.2. Separating surveys

Although not part of the current New Zealand Code of Conduct, a related mitigation measure deserves comment in this light. There has been an increasing use of survey separation (or separation of surveys and other sources of noise such as drilling) as a mitigation technique. The rationales offered for imposing this type of mitigation include: (a) this will reduce the level of cumulative sound exposure for a given animal in proximity to the two sources; and/or (b) the separation will create a 'corridor' that the animals can swim through without high exposure levels from either source.

This presumes of course that all animals are swimming in the same direction and are on the same side of the line between the two sources, and that the animals understand that the separation is intended to be a corridor for their passage (the TWG notes that these are doubtful assumptions, at best). A brief look at the geometry of the situation would actually lead one to the opposite conclusion: sources should be bunched up to the extent that is safely practicable. This would make the swim around the sources to get from Point A to Point B much shorter and less risky for all animals in all positions and for all relative angles of the animal's Point A and Point B relative to the angle between the two sources. And for scenarios where one or both sources are moving, the postulated 'corridor' would itself be moving or even closing faster than the animal would be likely to swim, depending on the relative angle of the animal's track to the tracks of the two vessels.

Additionally, it should be kept in mind that sound when expressed in decibels is on a logarithmic scale, which is also how animals hear the sound. A sound must be around 3 dB SPL different from another sound to be heard as being louder or quieter, and a sound that is ten times as loud, or 10 dB louder, is only heard as being twice as loud, not ten times as loud. For sound exposure levels, a doubling of the duration or loudness of exposure, or adding a second equally loud source,[18] results in a 3 dB change in SEL.

[18] This is true only for identical sources such as airguns in an array, but is definitely not the case when considering seismic and drilling or pile-driving activities.

The TWG agreed that it might actually be better, then, to move the sources close together, making the exposure not appreciably louder in SEL and shortening the duration of exposure (if the duty cycle and spectra of the sources are identical). However, some TWG members noted that the issue of an increase in the degree of masking when more than one source is active remains unsolved for both options, and is even greater for impulsive sources. Many seismic operations require two airgun arrays (eg ping-pong operations). Some TWG members suggested that, in these cases, operators purposely alternate shots, increasing both the amount of acoustic energy introduced into the environment and the potential for confusion when exposed animals engage in evasive behaviours. However, others noted that such ping-pong operations are standard protocol, simply allowing the required duty cycle to be achieved while one array is refilling with air. Accordingly, animals at any reasonable distance outside the twinned arrays would perceive these to be in the same place. These TWG members argued that consequently the total amount of energy is not doubled and nor would animals be confused in the way implied.

Duration also influences TTS and PTS, and so shortening the duration of exposure by moving multiple sources closer together may be a more effective mitigation method than separating them and increasing the duration of exposure, even though the peak pressure that the animal receives (SPLpeak) that the animal receives may be slightly less when the sources are separated. However, although concentrating sources in space reduces their overall footprint, it might increase the amount or severity of disturbance in the concentrated area, both in terms of the received level and the degree of masking. Such decisions may need to be made on a case-by-case basis, depending on how important the area in question is to marine mammals that will be disturbed, as well as their seasonality in terms of occurrence and diversity, if at all known.

For both soft starts and source separations, the properties of sound and how ears process it can lead to non-intuitive conclusions. While it might seem intuitively reasonable to our human way of thinking that a soft start gives animals more time to move away from a potentially loud source and that separating two sources allows animals to ‘swim between’ the sources, there is no particular reason to believe that the animals perceive the soundscape in the same way, or that the resulting sound exposure will achieve the cumulative duration and loudness, or total energy flux, that we might think would occur, due to the log scale of sound perception and the effect of the mitigation measures on duration of exposure. Although acoustic propagation can easily produce areas of convergence, meaning that a gradual decrease in received levels with distance from the sources is not always obvious, in deep water these convergences generally occur at large distances, which are not the scales of concern for the types of injury that soft starts aim to reduce.

Part 5: Other issues

9. Terminology: 'impacts' or 'potential impacts'

While this seems like a point that has as much to do with final edits of the document for consistency, this issue was raised as one that relates to precision in the terminology used in this report. Some TWG members suggested that the term 'impacts' should be reserved for real, actual, demonstrated impacts and that some special terminology should be used for decisions based on 'potential impacts' to indicate that there are no data or insufficient data to be confident that there is an impact. The argument was that the separation would elicit different levels of concern by decision makers considering regulatory actions that may be based on assumed or hypothetical consequences. Basing decisions on 'potential' impacts also implies that there is a need for clarifying research to take decisions out of the realm of the speculative and into the desired realm of 'best available science' based decisions – so anywhere that we use 'potential impacts' should be a flag for a corresponding recommendation for future research/more data needed.

Some TWG members proposed that 'impacts' should be used more when referring to hearing effects, and 'potential impacts' for behavioural and other sub-lethal effects. However, other TWG members did not agree, arguing that non-hearing physiological effects and injury should also be considered impacts, as well as demonstrated behavioural responses. They believed that the word 'potential' should be restricted only to cases where studies have not been performed to resolve if there is a real impact. It was also proposed that the probability of impact should be conveyed when statistically based or model-based decisions contain an unavoidable degree of uncertainty, such as weather forecasting. The TWG was unable to resolve this issue.

10. Reducing behavioural disturbance 'as far as possible within operational constraints'

The phrasing used in the current Code does not preclude significant disturbance to a threatened population of marine mammals, with significant possibilities for population consequences. Discussions in the TWG generated two distinct views.

Firstly, some TWG members expressed concern that there is a tendency to treat evidence of a behavioural response to seismic sound as meaning that there must also be negative biological consequences in terms of health, survival and reproduction. Behavioural disturbance and biological consequences are two different phenomena.

These members felt that speculations about how behavioural changes might become biologically significant require additional evidence in order to become the foundation of regulatory actions that will have significant social and environmental consequences (see NRC 2005). While we should not stop investigating the occurrence (or lack thereof) of such effects, we should not base current decisions on speculations that lack supporting evidence, and evidence of behavioural disturbance is not evidence of adverse biological outcomes. As part of this process, use of the PCAD/PCOD model framework for relating disturbance to outcome is very helpful.

By contrast, other TWG members noted that 'operational constraints' should not preclude measures that would prevent the decline of a population. For example, defining 'significant disturbance' may not be easy in a New Zealand context, as we do not have abundance estimates for the majority of populations – indeed, even the distributions are not completely known for most. Of particular note, it is still not known exactly how far the most endangered population in New Zealand, the Māui dolphin, extends offshore. Given these uncertainties, these TWG members felt that it is not possible to provide a scientifically based alternative wording, but that additional steps should be taken by DOC and other agencies to address the data gaps and ensure that residual impacts of mitigated seismic surveys do not further threaten marine mammal populations. Moreover, these TWG members stated that the burden of proof regarding if there is, or is not, significant disturbance should lie with the proponents of activities.

Despite these disagreements, recent publications on population consequences modelling (eg Christiansen & Lusseau 2015; King et al. 2015; New et al. 2015) were highlighted as being important by the whole TWG. The TWG also advocated for the use of the tools outlined in these papers, rather than developing a blanket statement that behavioural responses can be ignored. Additional data on behavioural responses, stress and other sub-injurious impacts need to be collected. Some additional data might also be generated by pre and post-seismic survey monitoring to look for spatial displacement at distances well beyond the acquisition zone. The collection of blow and faecal samples from both periods for cortisol analysis would also be a worthy undertaking.

References

Aguilar de Soto, N.; and C. Kight. 2016: Physiological effects of noise. Ch. 8 in Solan, M. and Whiteley, N. (Eds): Stressors in the Marine Environment. Physiological and Ecological Responses and Societal Implications. Oxford University Press. 350 pp.

Aguilar de Soto, N.; K. Gkikopoulou, S. Hooker, S. Isojunno, M. Johnson, P. Miller, P. Tyack P., P. Wensveen, C. Donovan, C. Harris, D. Harris, L. Marshall, C. Oedekoven, R. Prieto, L. Thomas. 2016: From physiology to policy: A review of physiological noise effects on marine fauna with implications for mitigation. Proceedings of Meetings on Acoustics 27 (Fourth International Conference on the Effects of Noise on Aquatic Life Dublin, Ireland. 10–16 July, 2016). 14 pp.

Aguilar de Soto, N.; Rogan, E.; Ó Cadhla, O.; Gordon, J.C.D.; Mackey, M.; Connolly, N. 2004: Cetaceans and Seabirds of Ireland's Atlantic Margin. Volume III – Acoustic surveys for cetaceans. Report on research carried out under the Irish Infrastructure Programme (PIP): Rockall Studies Group (RSG) projects 98/6 and 00/13, Porcupine Studies Group project P00/15 and Offshore Support Group (OSG) project 99/38. 51 pp. http://www.ucc.ie/research/crc/papers/Vol%20III%20-%20Cetacean%20acoustics%20Final.pdf.

Anon. 2001: Joint interim report on the Bahamas marine mammal stranding event of 15–16 March 2000 (December 2001). NOAA unpublished report. 59 p. Available at: http://www.nmfs.noaa.gov/ pro_tres/overview/Interim_Bahamas_Report.pdf.

Ashe, E.; Noren, D.P.; Williams, D.P.; 2010: Animal behaviour and marine protected areas: incorporating behavioural data into the selection of marine protected areas for an endangered killer whale population. *Animal Conservation 13(2)*: 196–203. http://onlinelibrary.wiley.com/doi/10.1111/j.1469-1795.2009.00321.x/abstract.

Barlow, J., and R. Gisiner. 2006: Mitigating, monitoring and assessing the effects of anthropogenic sound on beaked whales. *Journal of Cetacean Research and Management 7*: 239–249.

Bregman, A.S. 1990: Auditory Scene Analysis: The Perceptual Organization of Sound. Cambridge, Massachusetts: MIT Press, 773 pp.

Breitzke, M., O. Boebel, S. El Naggar, W. Jokat, and B. Werner. 2008: Broad-band calibration of marine seismic sources used by R/V Polarstern for academic research in polar regions. *Geophys. J. Int. 174*: 505–524

Blackwell, S.B., C.S. Nations, T.L. McDonald, C.R. Greene Jr, A.M. Thode, M. Guerra, and A.M. Macrander. 2013: Effects of airgun sounds on bowhead whale calling rates in the Alaskan Beaufort Sea. *Mar Mamm Sci. 2013;29(4)*:E342–E365. doi: 10.1111/mms.12001

Blackwell, S.B., C.S. Nations, T.L. McDonald, A.M. Thode, C.R., D. Mathias, K.H. Kim, Greene Jr, and A.M. Macrander 2015: Effects of Airgun Sounds on Bowhead Whale Calling Rates: Evidence for Two Behavioural Thresholds. PLoS ONE 10(6): e0125720. doi:10.1371/journal.pone.0125720.

Branstetter, B.K., J.S. Trickey, K. Bakhtiari, A. Black, H. Aihara, and J.J. Finneran. 2013: Auditory masking patterns in bottlenose dolphins (*Tursiops truncatus*) with natural, anthropogenic, and synthesized noise. *Journal of the Acoustical Society of America 133*: 1811–1818.

Castellote, M. 2007: General Review of Protocols and Guidelines for Minimizing Acoustic Disturbance to Marine Mammals from Seismic Surveys. *Journal of International Wildlife Law and Policy 10*: 273–288.

Castellote, M., and C. Llorens. 2016: Review of the Effects of Offshore Seismic Surveys in Cetaceans: Are Mass Strandings a Possibility? In: The Effects of Noise on Aquatic Life II. Volume 875. Advances in Experimental Medicine and Biology. Springer-Verlag, New York. 133–143.

Castellote, M., C.W. Clark, and M.O. Lammers. 2012: Acoustic and behavioural changes by fin whales (*Balaenoptera physalus*) in response to shipping and airgun noise. *Biological Conservation 147*: 115–122.

CBD. 2014: Background document on the development of practical guidance and toolkits to minimize and mitigate the significant adverse impacts of anthropogenic underwater noise on marine and coastal biodiversity.

Cerchio, S., S. Strindberg, T. Collins, C. Bennett, and H. Rosenbaum. 2014: Seismic Surveys Negatively Affect Humpback Whale Singing Activity off Northern Angola. PLoS ONE 9(3):e86464. doi:10.1371/journal.pone.0086464

Christiansen, F., and D. Lusseau. 2015: Linking Behavior to Vital Rates to Measure the Effects of Non-Lethal Disturbance on Wildlife. *Conservation Letters 8(6)*: 424–431. 10.1111/conl.12166

Clark, C.W., W.T. Ellison, B.L. Southall, L. Hatch, S.M., van Parijs, A. Frankel and D. Ponirakis. 2009: Acoustic masking in marine ecosystems: Intuitions, analysis, and implication. *Marine Ecology Progress Series 395*: 201–222

Claridge, D.E. 2013: Population ecology of Blaineville's beaked whales (*Mesoplodon densirostris*). PhD thesis, University of St Andrews, UK.

Cooke, J.G., D.W. Weller, A.L. Bradford, O. Sychenko, A.M. Burdin, A.R. Lang, and J.R.L. Brownell. 2015: "Updated Population Assessment of the Sakhalin Gray Whale Aggregation based on the Russia-US photoidentification study at Piltun, Sakhalin, 1994–2014". Western Gray Whale Advisory Panel WGWAP-16/17.

Cox, T.M., T.J. Ragen, A.J. Read, E. Vos, R.W. Baird, K. Balcomb, J. Barlow, J. Caldwell, T. Cranford, L. Crum, A.D' Amico, G.D' Spain, A. Fernandez, J.J. Finneran, R.L. Gentry, W. Gerth, F. Gulland, J. Hildebrand, D. Houser, T. Hullar, P.D. Jepson, D.R. Ketten, C.D. MacLeod, P. Miller, S. Moore, D.C. Mountain, D. Palka, P. Ponganis, S. Rommel, T.K. Rowles, B. Taylor, P.L. Tyack, D. Wartzok, R.C. Gisiner, J.G. Mead, and L. Benner. 2006: Understanding the impacts of anthropogenic sound on beaked whales. *Journal of Cetacean Research and Management. 7(3)*: 177–187.

D'Amico, A., R.C. Gisiner, D.R. Ketten, J.A. Hammock, C. Johnson, P.L. Tyack, and J. Mead. 2009: Beaked whale strandings and naval exercises. *Aquatic Mammals 35*: 452–472.

DeAngelis, D.L. and Grimm V. 2014: Individual-based models in ecology after four decades. F1000Prime Reports 6:39. Published online 2014 Jun 2. doi: 10.12703/P6–39.

Department of Art, Heritage and the Gaeltacht 2014: Guidance to Manage the Risk to Marine Mammals from Man-made Sound Sources in Irish Waters. 57 p. Available from: www.npws.ie/marine/bestpracticeguidelines/.

DeRuiter, S.L., B.L. Southall, J. Calambokidis, W.M.X. Zimmer, D. Sadykova, E.A. Falcone, A.S. Friedlaender, J.E. Joseph, D. Moretti, G.S. Schorr, L. Thomas, and P.L. Tyack. 2013: First direct measurements of behavioural responses by Cuvier's beaked whales to mid-frequency active sonar. Biology Letters 9:20130223. doi:10.1098/rsbl.2013.0223

Dunlop, R.A., M.J. Noad, R.D. McCauley, E. Kniest, D. Paton, and D.H. Cato. 2015: The Behavioural Response of Humpback Whales (*Megaptera novaeangliae*) to a 20 Cubic Inch Air Gun. *Aquatic Mammals 41(4)*: 412–433.

Dunlop, R.A., M.J. Noad, R.D. McCauley, E. Kniest, R. Slade, D. Paton, and D.H. Cato. 2016: Response of humpback whales (Megaptera novaeangliae) to ramp-up of a small experimental air gun array. *Marine Pollution Bulletin 103(1-2)*: 72–83. http://dx.doi.org/10.1016/j.marpolbul.2015.12.044.

Ellison, W.T., B.L. Southall, C.W. Clark, and A.S. Frankel. 2012. A New Context-Based Approach to Assess Marine Mammal Behavioural Responses to Anthropogenic Sounds: Marine Mammal Behavioural Responses to Sound. *Conservation Biology 26*: 21–28. DOI: 10.1111/j.1523-1739.2011.01803.x.

Energinet, D.K. 2015: Marine mammals and underwater noise in relation to piledriving – Working Group 2014. 21.01.2015, Revision: 2, Endelig udgave [Final Edition]. 20 p. Available at: https://ens.dk/sites/ens.dk/files/Vindenergi/underwater_noise.pdf.

Erbe, C., R. Williams, D. Sandilands, and E. Ashe. 2014: Identifying Modeled Ship Noise Hotspots for Marine Mammals of Canada's Pacific Region. PLoS ONE 9(3):e89820 doi:10.1371/journal.pone.0089820 (and also The PLOS ONE Staff (2014) Correction:

Identifying Modeled Ship Noise Hotspots for Marine Mammals of Canada's Pacific Region. PLoS ONE 9(11):e114362. doi:10.1371/journal.pone.0114362)

Erbe, C., C. Reichmuth, K. Cunningham, K. Lucke, and R. Dooling. 2016: Communication masking in marine mammals: A review and research strategy. *Marine Pollution Bulletin 103(1–2)*: 15–38.

Evans, D.L., and G.R. England. 2001: Joint interim report Bahamas marine mammal stranding event of 15–16 March 2000. US Department of Commerce and US Navy, Washington, D.C.

Fernández, A., J.F. Edwards, F. Rodriguez, A. Espinosa de los Monteros, P. Herraez, P. Castro, J.R. Jaber, V. Martin, and M. Arbelo. 2005: 'Gas and fat embolic syndrome' involving a mass stranding of beaked whales (family Ziphiidae) exposed to anthropogenic sonar signals. Vet. Pathol. 42: 446–57

Fernández, A., E. Sierra, V. Martín, M. Méndez, S. Sacchinni, Y. Bernaldo de Quirós, M. Andrada, M. Rivero, O. Quesada, M. Tejedor, and M. Arbelo. 2012: Last "atypical" beaked whales mass stranding in the Canary Islands (July, 2004). *Journal of Marine Science: Research & Development 02*. doi:10.4172/2155-9910.1000107

Finneran, J.J. 2015: Noise-induced hearing loss in marine mammals: A review of temporary threshold shift studies from 1996 to 2015. *The Journal of the Acoustical Society of America 138(3)*: 1702–1726.

Fothergill, D., J. Sims, and M. Curley. 2001: Recreational scuba divers' aversion to low-frequency underwater sound. Undersea & Hyperbaric Medicine 28(1).

Goldbogen, J.A., B.L. Southall, S.L. DeRuiter, J. Calambokidis, A.S. Friedlaender, E.L. Hazen, E.A. Falcone, G.S. Schorr, A. Douglas, D.J. Moretti, C. Kyburg, M.F. McKenna, and Tyack, P.L. 2013: Blue whales respond to simulated mid-frequency military sonar. Proceedings of the Royal Society B: Biological Sciences 280:20130657. http://dx.doi.org/10.1098/rspb.2013.0657.

Guerra, M.; Thode A.; Blackwell, S.; Macrander M. 2011: Quantifying seismic survey reverberation off the Alaskan North Slope. *The Journal of the Acoustical Society of America 130(5)*: 3046–3058.

Heide-Jørgensen, M.P., R. Guldborg Hansen, K. Westdal, R.R. Reeves and A. Mosbech. 2013: Narwhals and seismic exploration: Is seismic noise increasing the risk of ice entrapments? *Biological Conservation 158*: 50–54.

IAGC (International Association of Geophysical Contractors) 2015: Mitigation Measures For Cetaceans during Geophysical Operations. http://www.iagc.org/uploads/4/5/0/7/45074397/2015-02_iagc-mitigation_measures_for_cetaceans.pdf.

Jepson, P.D., R. Deaville, K. Acevedo-Whitehouse, J. Barnett, A. Brownlow, F.C. Clare, N. Davison, R.J. Law, J. Loveridge, S.K. Macgregor, S. Morris, S. Murphy, R. Penrose, M.W. Perkins, E. Pinn, H. Seibel, U. Siebert, E. Sierra, V. Simpson, M.L. Tasker, N. Tregenza, A.A. Cunningham, and A. Fernández. 2013: What Caused the UK's Largest Common Dolphin (Delphinus delphis) Mass Stranding Event? PLoS ONE 8(4):e60953. doi:10.1371/journal.pone.0060953

JNCC (U.K. Joint Nature Conservation Committee) 2010: JNCC guidelines for minimising the risk of injury and disturbance to marine Mammals from seismic surveys August 2010. Joint Nature Conservation Committee, Aberdeen, UK

Kaschner, K., N.J. Quick, R. Jewell, R. Williams, and C.M. Harris. 2012: Global Coverage of Cetacean Line-Transect Surveys: Status Quo, Data Gaps and Future Challenges. PLoS ONE 7(9):e44075. doi:10.1371/journal.pone.0044075

Katz, J., M. Chasen, C. English, L.J. Hood, and K.L. Tillery. 2014: Handbook of Clinical Audiology, 7th ed. Lippincott, Williams and Wilkins, NY NY, ISBN 9781451191639

Ketten, D.R., J. Lien, and S. Todd. 1993: Blast injury in humpback whale ears: evidence and implications. *The Journal of the Acoustical Society of America, 94(3): 1*849–1850.

King, S.L., R.S. Schick, C. Donovan, C.G. Booth, M. Burgman, L. Thomas, and J. Harwood. 2015: An interim framework for assessing the population consequences of disturbance. *Methods in Ecology and Evolution. 6(10)*: 1150–1158.

Kugler, K.., L. Wiegrebe, B. Grothe, M. Kossl, R. Gurkov, E. Krause, and M. Drexl. 2014: Low-frequency sound affects active micromechanics in the human inner ear. *Royal Society Open Science 1*: 140166–140166. http://rsos.royalsocietypublishing.org/content/royopensci/1/2/140166.full.pdf.

Kujawa, S.G. and Liberman, M.C. 2009: Adding insult to injury: cochlear nerve degeneration after "temporary" noise-induced hearing loss. *Journal of Neuroscience, 29(45)*: 14077–14085.

Kyhn, L.A., D. Boertmann, J. Tougaard, K. Johansen, and A. Mosbech. 2011: Guidelines to environmental impact assessment of seismic activities in Greenland waters. 3rd revised edition, Dec. 2011. Danish Center for Environment and Energy, Roskilde, Denmark

Leaper, R., S. Calderan, and J. Cooke. 2015: A Simulation Framework to Evaluate the Efficiency of Using Visual Observers to Reduce the Risk of Injury from Loud Sound Sources. *Aquatic Mammals 41(4)*: 375–387. DOI 10.1578/AM.41.4.2015.375.

Lien, J., S. Todd, P. Stevick, F. Marques, and D. Ketten. 1993: The reaction of humpback whales to underwater explosions: orientation, movements, and behavior. *The Journal of the Acoustical Society of America, 94(3)*: 1849–1849.

Madsen, P.T., M. Johnson, P.J.O. Miller, N. Aguilar Soto, J. Lynch, and P. Tyack P. 2006: Quantitative measures of air-gun pulses recorded on sperm whales (*Physeter macrocephalus*) using acoustic tags during controlled exposure experiments. *The Journal of the Acoustical Society of America 120*: 2366. DOI: 10.1121/1.2229287.

McCauley, R.D. 2015: Offshore Irish noise logger program (March to September 2014): analysis of cetacean presence, and ambient and anthropogenic noise sources. Report produced for Woodside Energy, DRIMS-#10574111, CMST Curtin University, R2015-01, 110 pp. 72 Fig.

Miller, P.J.O., P.H. Kvadsheim, F.-P.A. Lam, P.L. Tyack, C. Cure, S.L. DeRuiter, L. Kleivane, L.D. Sivle, S.P. van IJsselmuide, F. Visser, P.J. Wensveen, A.M. von Benda-Beckmann, L.M. Martin Lopez, T. Narazaki, and S.K. Hooker. 2015: First indications that northern bottlenose whales are sensitive to behavioural disturbance from anthropogenic noise. *Royal Society Open Science 2*: 140484-140484. doi:10.1098/rsos.140484

Miller, P.J.O., M.P. Johnson, P.T. Madsen, N. Biassoni, M. Quero, and P.L. Tyack. 2009: Using at-sea experiments to study the effects of airguns on the foraging behavior of sperm whales in the Gulf of Mexico. Deep Sea Research Part I: Oceanographic Research Papers 56: 1168–1181. DOI: 10.1016/j.dsr.2009.02.008.

Miller, P.J.O. R.N. Antunes, P.J. Wensveen, F.I.P. Samarra, A. Catarina Alves, P.L. Tyack, P.L., P.H. Kvadsheim, L. Kleivane, F.-P.A. Lam, M.A. Ainslie, and L. Thomas. 2014: Dose-response relationships for the onset of avoidance of sonar by free-ranging killer whales. *The Journal of the Acoustical Society of America 135*: 975–993. DOI:http://dx.doi.org/10.1121/1.4861346

Møhl, B. 1980. Marine mammals and noise. *Arctic Seas Bulletin 2(8)*: 1–2.

Møhl, B. 1981. Masking effects of noise; their distribution in time and space. In N. M. Peterson (Ed.), The question of sound from icebreaker operations. Proceedings of a Workshop 23 and 24 February 1981. (pp. 259–268). Toronto: Arctic Pilot Project.

Moretti, D., T.A. Marques, L. Thomas, N. DiMarzio, A. Dilley, R. Morrissey, E. McCarthy, J. Ward, and S. Jarvis. 2010: A dive counting density estimation method for Blainville's beaked whale (*Mesoplodon densirostris*) using a bottom-mounted hydrophone field as applied to a Mid-Frequency Active (MFA) sonar operation. *Applied Acoustics 71*: 1036–1042. doi:10.1016/j.apacoust.2010.04.011

New, L.F., A.J. Hall, R. Harcourt, G. Kaufman, E.C.M. Parsons, H.C. Pearson, A.M. Cosentino, and R.S. Schick. 2015: The modelling and assessment of whale-watching impacts. *Ocean & Coastal Management 115*: 10–16.

Nieukirk, S.L., D.K. Mellinger, S.E. Moore, K. Klinck, R.P. Dziak, and J. Goslin. 2012: Sounds from airguns and fin whales recorded in the mid-Atlantic Ocean, 1999–2009. *Journal of the Acoustical Society of America 131(2)*: 1102–1112.

NOAA (National Oceanic and Atmospheric Administration) 2015a: DRAFT Guidance for Assessing the Effects of Anthropogenic Sound on Marine Mammal Hearing: Underwater Acoustic Threshold Levels for Onset of Permanent and Temporary Threshold Shifts. Revised version for Second Public Comment Period, July 23, 2015.

NOAA (National Oceanic and Atmospheric Administration) 2015b: DRAFT US Atlantic and Gulf of Mexico Marine Mammal Stock Assessments: 2015. http://www.nmfs.noaa.gov/pr/sars/pdf/atl2015_draft.pdf.

Nowacek, D.P., C.W. Clark, D. Mann, P.J. Miller, H.C. Rosenbaum, J.S. Golden, M. Jasny, J. Kraska, and B.L. Southall. 2015: Marine seismic surveys and ocean noise: time for coordinated and prudent planning. *Frontiers in Ecology and the Environment 13*: 378–386. DOI:10.1890/130286.

NRC (National Research Council) 2005: Marine mammal populations and ocean noise: determining when noise causes biologically significant effects. The National Academies Press, Washington, D.C.

Parsons, E.C.M., S.J. Dolman, M. Jasny, N.A. Rose, M.P. Simmonds, and A.J. Wright. 2009: A critique of the UK's JNCC seismic survey guidelines for minimising acoustic disturbance to marine mammals: Best practise? *Marine Pollution Bulletin 58*: 643–651.

Pirotta, E., K.L. Brookes, I.M. Graham, and P.M. Thompson. 2014: Variation in harbour porpoise activity in response to seismic survey noise. Biology Letters. DOI:10.1098/rsbl.2013.1090.

Prideaux, G., and M. Prideaux. 2015. Environmental impact assessment guidelines for offshore petroleum exploration seismic surveys. Impact Assessment and Project Appraisal 34(1): 33–43. DOI:10.1080/14615517.2015.1096038.

Robertson, F.C. W.R. Koski, and A.W. Trites. In Press. Behavioural responses affect distribution analyses of bowhead whales in the vicinity of seismic operations. Marine Ecology Progress Series. doi:10.3354/meps11665.

Robertson, F.C., W.R. Koski, T.A. Thomas, W.J. Richardson, B. Würsig, and A.W. Trites. 2013: Seismic operations have variable effects on dive cycle behavior of bowhead whales in the Beaufort Sea. *Endangered Species Research 21*: 143–160.

Rolland, R., S.E. Parks, K.E. Hunt, M. Castellote, P.J. Corkeron, D.P. Nowacek, S.K. Wasser, and S.D. Kraus 2012: Evidence that ship noise increases stress in right whales. Proceedings of the Royal Society B 279: 2363–2368.

Romano, T.A., M.J. Keogh, C. Kelly, P. Feng, L. Berk, C.E. Schlundt, D.A. Carder, and J.J. Finneran. 2004: Anthropogenic sound and marine mammal health: Measures of the nervous and immune systems before and after intense sound exposure. Can. J. Fish. Aq. Sci. 61: 1124–1134.

Salt, A.N., and T.E. Hullar. 2010: Responses of the ear to low frequency sounds, infrasound and wind turbines. *Hearing research 268*: 12–21.

Shannon, G., M.F. McKenna, L.M. Angeloni, K.R. Crooks, K.M. Fristrup, E. Brown, K.A. Warner, M.D. Nelson, C. White, J. Briggs, S. McFarland, and G. Wittemyer. 2015: A synthesis of two decades of research documenting the effects of noise on wildlife. Biological Reviews. doi: 10.1111/brv.12207

Sivle, L.D., P.H. Kvadsheim, C. Curé, S. Isojunno, P.J. Wensveen, F.-P.A. Lam, f. Visser, L. Kleivane, P.L. Tyack, C.M. Harris, and P.J.O. Miller. 2015: Severity of Expert-Identified Behavioural Responses of Humpback Whale, Minke Whale, and Northern Bottlenose Whale to Naval Sonar. *Aquatic Mammals 41(4)*: 469–502.

Steevens, C.C., K.L. Russell, M.E. Knafelc, P.F. Smith, E.W. Hopkins, and J.B. Clark. 1999: Noise-induced neurologic disturbances in divers exposed to intense water-borne sound: Two case reports. *Undersea & Hyperbaric Medicine 26(4)*: 261–265.

Southall, B.L., A.E. Bowles, W.T. Ellison, J.J. Finneran, R.L. Gentry, C.R. Greene, Jr., D. Kastak, D.R. Ketten, J.H. Miller, P.E. Nachtigall, W.J. Richardson, J.A. Thomas, and P.L. Tyack. 2007: Marine mammal noise exposure criteria: Initial scientific recommendations. *Aquatic Mammals 33*: 411–521.

Southall, B.L., T. Rowles, F. Gulland, R.W. Baird, and P.D. Jepson. 2013: Final report of the Independent Scientific Review Panel investigating potential contributing factors to a 2008 mass stranding of melon-headed whales (Peponocephala electra) in Antsohihy, Madagascar. Available from: https://iwc.int/2008-mass-stranding-in-madagascar.

Stone, C. J. 2015a: Marine mammal observations during seismic surveys from 1994–2010. JNCC report, No. 463a.

Stone, C.J. 2015b: Implementation of and considerations for revisions to the JNCC guidelines for seismic surveys. Joint Nature Conservation Committee Report No. 436b.

Taylor, B.L., M. Martinez, T. Gerrodette, J. Barlow, and Y.N. Hrovat. 2007: Lessons from monitoring trends in abundance of marine mammals. *Marine Mammal Science 23*: 157–175

Thompson, P.M., B. Wilson, K. Grellier, and P.S. Hammond. 2000: Combining Power Analysis and Population Viability Analysis to Compare Traditional and Precautionary Approaches to Conservation of Coastal Cetaceans. *Conservation Biology 14*: 1253–1263.

Tolstoy M.D., J.B. Diebold, S.C. Webb, D.R. Bohnenstiehl, E. Chapp, R.C. Holmes, and M. Rawson. 2004: Broadband calibration of R/V Ewing seismic sources. Geophysical Research Letters 31:L14310. doi:10.1029/2004GL020234. 4 p

Tougaard J., J. Carstensen, J. Teilmann, H. Skov, and P. Rasmussen. 2009: Piledriving zone of responsiveness extends beyond 20 km for harbor porpoises (*Phocoena phocoena* (L.)). *The Journal of the Acoustical Society of America 126*:11. DOI: 10.1121/1.3132523.

Tougaard, J., A.J. Wright, and P.T. Madsen. 2015: Cetacean noise criteria revisited in the light of proposed exposure limits for harbour porpoises. *Marine Pollution Bulletin 90*: 196–208. doi:10.1016/j.marpolbul.2014.10.051. http://www.sciencedirect.com/science/article/pii/S0025326X14007358.

Tyack, P.L., W.M.X. Zimmer, D. Moretti, B.L. Southall, D.E. Claridge, J.W. Durban, C.W. Clark, A. D'Amico, N. DiMarzio, S. Jarvis, E. McCarthy, R. Morrissey, J. Ward, and I.L. Boyd, 2011: Beaked Whales Respond to Simulated and Actual Navy Sonar. PLoS ONE 6:e17009. doi:10.1371/journal.pone.0017009

von Benda-Beckmann, A.M., P.J. Wensveen, P.H. Kvadsheim, F.-P.A. Lam, P.J.O. Miller, P.L. Tyack, and M.A. Ainslie. 2014: Modeling ramp-up effectiveness at mitigating impact of sonar sounds on marine mammals. *Conservation Biology 28(1)*: 119–128. (doi:10.1111/cobi.12162).

Weir, C.R.., and S.J. Dolman. 2007: Comparative Review of the Regional Marine Mammal Mitigation Guidelines Implemented During Industrial Seismic Surveys, and Guidance Towards a Worldwide Standard. *Journal of International Wildlife Law & Policy 10*: 1–27. DOI: 10.1080/13880290701229838.

Wensveen, P., P. Kvadsheim, F.-P.A. Lam, P.L. Tyack, A.M. von Benda-Beckmann, L. Sivle, F. Visser, C. Curé, and P. Miller. Submitted. Effectiveness of ramp-up of naval sonar studied in humpback whales.

Werner, S. 2012: Determination of noise exposure criteria – the German approach. Federal Environment Agency (UBA), Presentation in Stralsund, January 24th, 2012: Available at: www.bfn.de/fileadmin/MDB/documents/themen/erneuerbareenergien/Tgng_offshore2012/2_3_werner.pdf.

Williams, R., D. Lusseau, and P.S. Hammond. 2006: Estimating relative energetic costs of human disturbance to killer whales (*Orcinus orca*). *Biological Conservation 133*: 301–311

Williams, R., C. Erbe, E. Ashe, A. Beerman, and J. Smith. 2014: Severity of killer whale behavioural responses to ship noise: A dose–response study. *Marine Pollution Bulletin 79(1–2)*: 254–260. http://www.sciencedirect.com/science/article/pii/S0025326X13007376.

Wright, A.J., S.J. Dolman, M. Jasny, E.C.M. Parsons, D. Schiedek, and S.B. Young. 2013: Myth and Momentum: A Critique of Environmental Impact Assessments. *Journal of Environmental Protection 04*: 72–77. DOI:10.4236/jep.2013.48A2009.

Wright, A.J., Aguilar de Soto, N.; Baldwin, A.L.; Bateson, M.; Beale, C.M.; Clark, C.; Deak, T.; Edwards, E.F.; Fernández, A.; Godinho, A.; Hatch, L.T.; Kakuschke, A.; Lusseau, D.; Martineau, D.; Romero, M.L.; Weilgart, L.S.; Wintle, B.A.; Notarbartolo-di-Sciara, G.; Martin, V. 2007: Anthropogenic Noise as a Stressor in Animals: A Multidisciplinary Perspective. *International Journal of Comparative Psychology 20*: 250–273.

Appendix: Peer review

Given the scientific nature of this particular Technical Working Group report, DOC solicited peer review of the material contained within. What follows is the substantive feedback obtained through this process. Editorial comments received have already been incorporated into the text. Changes to the document in response to the reviewer comments have been noted.

DOC appreciates the additional clarification and opinion on the various statements and arguments contained within the pages of this report, especially given the timeline under which the original was produced. DOC also notes that the majority of the additional opinions expressed by the reviewers is already captured in the discussions between the different TWG members themselves. However, all the additional feedback was considered during revision of the Code.

DOC thanks the reviewers for their time and feedback, in alphabetical order by surname:

- Susanna Blackwell, Greeneridge Sciences, Inc., Santa Barbara, California
- Shane Guan, Office of Protected Resources, National Marine Fisheries Service. Silver Spring, Maryland

Reviewer 1

General comments

Overall it's a good summary of the current state of knowledge (or lack thereof). The few comments that I had were quite minor.

Specific comments

1) Section 6.5, third paragraph – Regarding the statement "the effectiveness of seismic sound as a masker of biological sounds is substantially lower than a continuous broadband signal of equivalent source level for normal communication or echolocation ranges …" I'd like to point out that at close ranges the masking ability of a seismic operation is actually considerable, and the masking sound is not the airgun pulses themselves but rather the reverberation between the pulses. Guerra et al. (2011, cited in the document) showed that seismic surveys can "increase the background noise (ie between airgun pulses) over natural ambient levels by 30–45 dB within 1 km of the activity, by 10–25 dB within 15 km of the activity, and by a few dB at 128 km range".
2) Section 6.5, sixth paragraph – I disagree with the statement that "this audibility is confined to specific oceanographic conditions that typically occur below the dive depth of baleen whales". Greeneridge Sciences as well as colleagues of mine have placed recorders in many places on the continental shelf off the North Slope of Alaska, over an area hundreds of kilometres wide. All our recorders (in depths of 20–55 m, some much deeper) have recorded thousands/tens of thousands of airgun

pulses over the last decade, sometimes nearly continuously for the entire deployment season. Some of the pulses detected on these shallow recorders came from ~1,300 km away.

3) Section 6.5, seventh paragraph – the statement "… the cumulative effect of masking over the long term is very considerable" is bothersome in my opinion, as it makes a strong statement (that the lay reader will pick up on) but without explaining anything.

<u>**DOC COMMENT**: This was been revised by the TWG to "the cumulative effect of masking over the long term may be more considerable" to soften the strong statement.</u>

4) Section 6.5, seventh paragraph – re the estimate 0.1%: it could be higher than that when you move somewhat away from the ship: at a few tens of km, when the pulse duration is close to 1 s, the duty cycle could be about 5%.

5) Section 6.5, seventh paragraph – "Guerra et al. (2011) showed that for one seismic survey there was 'no time period between airgun shots when reverberation does not mask one's ability to detect and recognise calls'" This statement is really meaningless (or erroneous) if you don't specify the distance from the seismic ship and the size of the array for which the statement applies! At some distance from the seismic ship there will no longer be any masking by reverberation, but that is not clear from the statement.

6) Section 6.6.4, last paragraph – "It would be possible to deploy these systems before, during and after the survey to assess the vocal responses of marine mammals." Such studies are very difficult to interpret (speaking from experience!) because you cannot tell the difference between an animal that deflects and leaves the area without modifying its calling behaviour and another animal that stops vocalising but stays put.

<u>**DOC COMMENT**: This paragraph was removed by the TWG.</u>

7) Section 7.2, eighth paragraph – "… or whether vessel noise itself may provide a similar warning to keep animals at a sufficient distance from man-made sounds to avoid injury even without soft starts." I find that argument very poor. Most likely, every marine mammal on the planet has been subjected to vessel noise, and for many it is a sound they hear often if not nearly all the time. It is therefore unclear to me how such a ubiquitous sound could then be expected to forewarn an animal of the imminent presence of airgun pulses.

8) Section 7.2, ninth paragraph – "Provisions have been made in the USA and elsewhere to exempt animals from mitigative shutdowns or power-downs…" An issue with such an exemption, in my opinion, is the scenario of a seal approaching an array either with its head in air or only in the top 50 cm of water – the animal will experience severely reduced sound levels and may not realise how high the amplitudes are until it dives.

9) Section 7.2, tenth paragraph (quoted text) – "… firing should be reduced to only the smallest airgun in the array, a so-called mitigation gun, which should prevent further approach of animals to the array." If an animal is approaching an array firing at full power, why would then a mitigation gun "prevent further approach of animals to the array"?) Maybe I am not understanding the intended meaning of the sentence. I

thought the role of the mitigation gun was to eliminate (or at least limit) injury in the curious animal … so it would make more sense to me if the sentence said "… so-called mitigation gun, which should avoid (or reduce) potential injury in the curious animal."

10) Section 7.3, last paragraph – re: "An alternative method would be to use the full array but at low pressure, and gradually increase the pressure to achieve full power over a given time." I agree that from the perspective of the complex way in which sound from adjacent airguns interact with each other and then propagate away from the array, it would make much more sense to use the full array all the time, at variable pressures.

11) Section 8.2, whole section – whether the footprint of two sound sources and their overall effect on the local marine mammals is greater with the sources spread apart or bunched together probably needs to be examined on a case by case basis, as it will depend on a number of variables, such as the nature of the sound source and the animal of interest. A large oil company in the US recently made a case for the "bunched together" scenario being less disruptive to walruses during drilling, but the supporting documents are not at this point in the public domain.

Reviewer 2

General comments

Overall I think it's a very thorough report, and it addresses a broad array of issues. Some of the issues do not seem to have consensus among panel members, which is fine, but the reasons for disagreement don't seem to be fleshed out clearly (eg regarding whether TTS should be considered as injury).[19]

Specific comments

1) Section 3.1, forth paragraph– re: "One proposed option for 'safe' levels (upon which 'safe' ranges might be based) was the newly revised draft US National Oceanic and Atmospheric Administration (NOAA) regulatory criteria (NOAA 2015a; 'draft NOAA (2015a) criteria') for direct hearing injury (other types of injuries are not assessed)." The draft NOAA criteria also include TTS, which they do not consider a form of injury. However, reducing risk of auditory injury (as defined by NOAA) also reduces TTS occurrence.

2) Section 3.1, last paragraph – I am not sure I completely understand why the Code would need to provide incentives. I think in general the operators are aware of the mitigation measures that could affect the operations. For example, shutdown measure would delay the complete of project, which is a strong economic motivation for the operator to seek smaller airgun array (if operationally feasible) so the shutdown zone would be small – thus less chance of shutting down.

[19] An effort was made to restructure the document following this comment to make the reasons for the disagreements clearer.

DOC COMMENT: This paragraph was removed by the TWG.

3) Section 3.2, footnote 3 – re: "However, it also needs to be noted that the effectiveness of PAM also relies on factors beyond human control, such as whether an animal is vocalising or not." PAM effectiveness also depends on marine environment that affects sound propagation (reverb, etc).

4) Section 3.3, first paragraph – Southall et al. (2007) provides a 10-tier category system to rate different types of behavioural responses for 'biological significance'.

5) Section 3.4, last paragraph – regarding the lack of spatially explicit abundance and density estimates in New Zealand, I would suggest a recommendation that request the New Zealand Government to invest more in marine mammal stock assessment in its waters.

6) Section 4.1, first paragraph – re: "It was agreed that weighting functions provide a useful way of broadly accounting for the hearing sensitivities of different species across different frequency ranges, especially for the assessment of cumulative exposure using some sort of sound exposure level (SEL) metric (it is less important for considering peak pressure impacts)." Why?

   DOC COMMENT: This question was put to the TWG, but no response or additional details were provided.

7) Section 4.2, second paragraph – re: "The latest draft NOAA (2015a) criteria weighting functions are another step forward, but there are currently no experimental data that correlate the weighted levels with degree of injury (or the severity of behavioural reactions) across the range of species they are being applied to." I would argue that weighting function not be applied to assess behavioural responses, since it's highly subject to the contexts.

   DOC COMMENT: The text has been edited to remove mention of behavioural responses, "The latest draft NOAA (2015a) criteria weighting functions are another step forward, but there are still extrapolating beyond our current knowledge on very few species to all marine mammals. Some in the TWG noted the inherent risk of this extrapolation, which urgently needs to be addressed with newer empirical data."

8) Section 4.2, fourth paragraph – re: "For example, a curve that is based on measurements across 1/3 octave bins may be better supported biologically…" Since the weighting curves are developed using audiogram (or estimated hearing sensitivity), so I suspect that 1/3rd octave bands are already taken into consideration?

   DOC COMMENT: This paragraph was removed by the TWG.

9) Section 4.2, fifth (now fourth) paragraph – ABR does not provide accurate measurements in low-frequency regime as the process is subject to system noise, which is mostly in the low-frequency regime.

10) Section 4.4, first paragraph – as mentioned earlier, I do not agree using weighted function for behavioural disturbance analysis. Behavioural response is highly context-based.

11) Section 4.4, fourth paragraph – re: "It was noted that a correlation between audibility and behavioural response thresholds has been convincingly demonstrated, at least for harbour porpoises (Tougaard et al. 2015)" and "… if a signal is inaudible to an animal,

it cannot elicit a behavioural response in that animal." To some degree this may be right, but being not audible is a special case in most situations when assessing behavioural response. The signal has to be extremely low, or completely outside the animals' hearing range, to be undetected. In the latter case, I think a simple box-cut would define which frequency of signal would not be detected for behavioural response, rather than applying the weighting function.

12) Section 5.1.1, first paragraph – while I agree the statement that TTS should not be considered as injury, I think the assessment on noise induced hearing loss (PTS) should be based on healthy animals. This is also the norm for human hearing studies. While it is understood that many animals in a population would not have perfect hearing sensitivity due to a variety of reasons, since the audiogram of the species is based on healthy individuals, so should the impact assessments be also.

13) Section 5.1.1, last paragraph – re: "...other members of the TWG, who argued that TTS is an injury in its own right ..." I do not agree. TTS is a condition that we frequently experience when exposed to elevated noise.

14) Section 6.4.1, first paragraph – re: "... inducing avoidance ..." Aside from avoidance, when an animal has to tolerate an external stressor (eg noise) in order to survive (feed or other biologically important function), it also induces physiological stress response.

15) Section 6.4.2, last paragraph – re: "While stress responses cannot yet be used as a metric of disturbance by seismic sound, they should prove a useful tool in the near future and so their use should be encouraged by regulators." But it is difficult to measure stress levels in wild animals/populations.

16) Section 7.1, first paragraph – re: "There is little evidence to demonstrate the effectiveness (or lack thereof) of soft starts in mitigating the overall risk of hearing effects in marine mammals exposed to seismic surveys." Studies show that, at least, also cetaceans are able to reduce hearing sensitivity following a warning sound (Nachtigall et al. 2015). This stapedial reflex mechanism of auditory system is also found in other mammalian species, including humans. Perhaps this can be considered as a benefit of ramp-up/soft start?