RiCORE
ricore-project.eu

WP 5	Deliverables **5.2 & 5.4**

Guidance on effective Adaptive Management and post-consent monitoring strategies

Deliverables 5.2 & 5.4

PROJECT COORDINATOR

David Gray (Robert Gordon University)

TASK LEADER

Finlay Bennet (Marine Scotland Science)

AUTHORS

Finlay Bennet (Marine Scotland Science), Ross Culloch (MaREI, University College Cork), Adrian Tait (Marine Scotland Science)

SUBMISSION DATE

29 | June | 2016

Citation

Bennet, F., Culloch, R. and Tait, A. 2016. Guidance on effective Adaptive Management and post-consent monitoring strategies. Deliverables 5.2 & 5.4., RiCORE Project. 45 pp.

RiCORE
ricore-project.eu

Contents

RiCORE Project Synopsis

The aim of the RiCORE project (Risk based Consenting for Offshore Renewable Energy) is to establish a risk-based approach to consenting where the level of environmental survey required is based on the environmental sensitivity of the site, the risk profile of the technology and the scale of the proposed project. The project, which has received funding from the European Union's Horizon 2020 research and innovation programme, will run between January 1st 2015 and June 30th 2016.

The consenting of offshore renewable energy is often cited as one of the main non-technical barriers to the development of this sector. A significant aspect of this is the uncertainty inherent in the potential environmental impacts of novel technology. To ensure consents are compliant with EU and national legislation, such as the Environmental Impact Assessment and Habitats Directives, costly and time consuming surveys are required even for perceived lower risk technologies in sites which may not be of highest environmental sensitivity.

The RiCORE project will study the legal framework in place in the partner Member States to ensure the framework developed will be applicable for roll out across these Member States and further afield. The next stage of the RiCORE project is to consider the practices, methodologies and implementation of pre-consent surveys, post consent and post-deployment monitoring. This will allow a feedback loop to inform the development of the risk-based framework for the environmental aspects of consent and provide best practice. The project will achieve these aims by engaging with the relevant stakeholders including regulators, industry, and EIA practitioners, through a series of expert workshops and developing their outcomes into guidance.

A key objective of the project is to improve consenting processes in line with the requirements of the Renewable Energy Directive (specifically Article 13-1) to ensure cost efficient delivery of the necessary surveys, clear and transparent reasoning for work undertaken, improving knowledge sharing and reducing the non-technical barriers to the development of the Offshore Renewable Energy sector so that it can deliver clean, secure energy.

EXECUTIVE SUMMARY

This report considers policy approaches and scientific methods of relevance to post-consent monitoring (PCM). The purpose is to provide guidance on how PCM can be approached to enable decision makers to become progressively better informed by cost-effectively reducing the scientific uncertainties associated with the putative impacts of Offshore Renewable Energy (ORE) on biodiversity interests. In this regard a key concern is the potential for data-rich and information-poor (DRIPy) post-consent monitoring, which is an undesirable outcome. The purpose of this report is to help ensure it arises less often. Uninformative post-consent monitoring risks polarizing debate with respect to the putative impacts associated with ORE, and may lead to considerable delays in the establishment of commercial scale projects. There are a number of technical reasons why this outcome arises relating to the specification of the monitoring study question, the study design and inattention to appropriate effect sizes and issues associated with statistical power and significance. Where regulatory institutions have overly risk-averse approaches to consenting and staff who lack the specialist technical skills needed to plan post-consent monitoring, it is more likely that studies will lack statistical power to detect meaningfully reduce scientific uncertainty. Each of the key issues are considered and recommendations made in the context of Adaptive Management approaches. More open and transparent risk-based approaches to post-consent monitoring are needed if regulators are to meaningfully reduce the key scientific uncertainties in a manner that is more effective than existing practice.

1. Introduction

A key aspect of environmental assessment and regulation of new technologies using risk-based approaches is a requirement for informative monitoring of devices that have been deployed in the marine environment. The learning outcomes can be expected to inform future decision making as the progression from single devices, small arrays to full scale commercial deployment takes place. Policies that support risk-based decision-making therefore require a coherent strategy for enabling this monitoring, and it has been suggested in RiCORE workshops that they are not currently well established.

Other reports in the RiCORE series primarily focus on site survey, data collection and analysis for the purpose of site characterization that informs a strategic approach to risk-based decision making, and supports decision-making in the face of scientific uncertainties. This report is concerned with guidance to ensure adequate standards of post-consent monitoring by consideration of policy approaches and scientific methods that promote post-consent monitoring that can enable decision makers to become progressively better informed by reducing the scientific uncertainties. The approaches considered are also intended to ensure that the maximum amount of information can be gleaned in a cost-effective manner from limited data.

2. Objectives

There are two key objectives of this report. Firstly, Section 3 provides a guide to Adaptive Management as a policy tool that provides a clear purpose for reducing scientific uncertainty through post-consent monitoring. The particular needs, experiences and context of marine renewable energy are discussed. The Survey, Deploy and Monitor Policy is considered an example of a policy approach that by enabling proportionate use of pre-consenting surveys to inform consent decisions then allows post-consent monitoring to meet the goals of Adaptive Management. Secondly, Section 4 provides a guide to setting up post-consent monitoring to reduce scientific uncertainty. The section discusses application of key scientific principles that are required in order to design genuinely informative post-consent monitoring studies.

3. Guidance for effective Adaptive Management

3.1 Introduction

The definition of Adaptive Management can often be open to various interpretations. This report takes the definition and associated Technical Guidance produced by the United States Department of Interior (US DOI) (Williams *et al,* 2009) to be the most definitive.

Adaptive Management: a process used to manage "resources that are responsive to management interventions but subject to uncertainties about the impacts of those interventions" and has been described as "a structured process of learning by doing, and adapting based on what is learned." The goal of adaptive management is to reduce scientific uncertainty.

The US DOI provides guidance for adopting Adaptive Management practices that are applicable to the scientific uncertainties associated with the impacts of all human activities on natural resource systems. Consequently, there is a need to more carefully consider the specific context of ORE devices and the issues associated with reducing scientific uncertainties in relation to impacts on biodiversity afforded statutory protection under EU nature conservation Directives.

There are a number of examples of regulators approaching the issue of Adaptive Management by requiring developers to produce Adaptive Management Plans (Le Lièvre *et al*, 2016), and an increasing body of evidence that these plans are often ineffective at achieving the goal of meaningfully reducing scientific uncertainties. The contention that the following sections is based upon is that Adaptive Management is most appropriately viewed as an over-arching philosophy that is the only approach that enables improved decision-making over time through reducing scientific uncertainties.

3.2 Clarifying the purpose of post-consent monitoring

Post-consent monitoring: Post-consent includes the post-deployment period plus the additional time that exists between a regulator granting consent and devices being deployed. This period between post-consent and deployment is potentially important as it may provide opportunities to design studies that can gather more meaningful information than is possible where data collection is limited to the post-deployment period.

A first-step in considering how Adaptive Management should be applied is to reconsider the purpose of doing post-consent monitoring, which ultimately provides the information necessary for successfully implimenting an AM approach to reducing scientific uncertainties and improving decision making. The position taken in this document is that the most useful purpose of post-consent monitoring, in the context of the scaling up to commercialisation of ORE devices, is for informing the decision making processes for other future projects. Often the apparent purpose of post-consent monitoring differs from this. In the worst case scenario, the purpose of post-consent monitoring may not be clearly defined, merely being imposed as a step following Environmental Impact Assessment (EIA) as a condition attached to the consent. Alternatively, it may be set in the context of specifically needing further information to provide assurance that certain levels of change have not been exceeded, or the potential relaxation of initially stringent mitigation that was considered appropriate on a precautionary basis. In some cases, it may be to identify unforeseen impacts. Culloch *et al.* (2015) have demonstrated that it is likely to be unrealistic that monitoring at a population level will allow informed inferences to be made about the levels of impact associated with ORE devices. It is also clear that monitoring objectives constrained by tolerable effect sizes that are negligible makes it more likely that monitoring programmes lack sufficient statistical power to reach meaningful conclusions. This aspect is discussed in more detail in Section 4 of this document.

Where initial assessments have necessarily been informed by a precautionary approach, owing to scientific uncertainties regarding the impacts, a useful goal of post-consent monitoring can be to reduce the uncertainties within analytical assessment frameworks being used, which, in turn, can make them more realistic and robust over time. The purpose is to inform future decision making through provision of information that will allow assessment and regulation of new proposals. Such a focus on ‘learning by doing’ is the key component of Adaptive Management strategies to enable improved decision making in the face of scientific uncertainties (Williams *et al,* 2009). The recommendation here is that the emphasis is on learning which has the greatest level of utility. This is likely to mean that the learning can be applied by government institutions and their advisors in order to provide more informative approaches to assessing the impacts of future projects. The expected adaptive change is to government policies associated with assumptions and methods that inform decision making. An alternative form of learning is to make adjustments to specific projects, and whilst this may have value, its utility will be relatively more limited in scope. Figure 1 below illustrates learning outcomes that lead to either adjustments of the existing project or learning applied to future plans and projects.

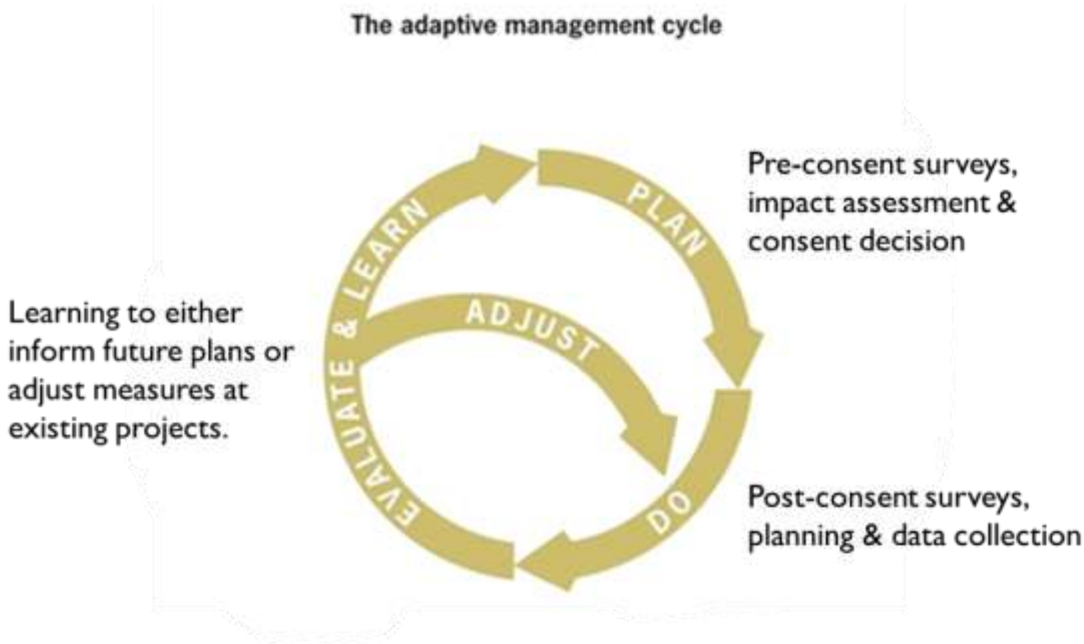

Figure 1: Diagram of the Adaptive Management cyclical process in the context of consenting ORE projects

3.3 Towards Adaptive Management for offshore renewable energy

The Adaptive Management cycle can also be illustrated using Figure 2, below.

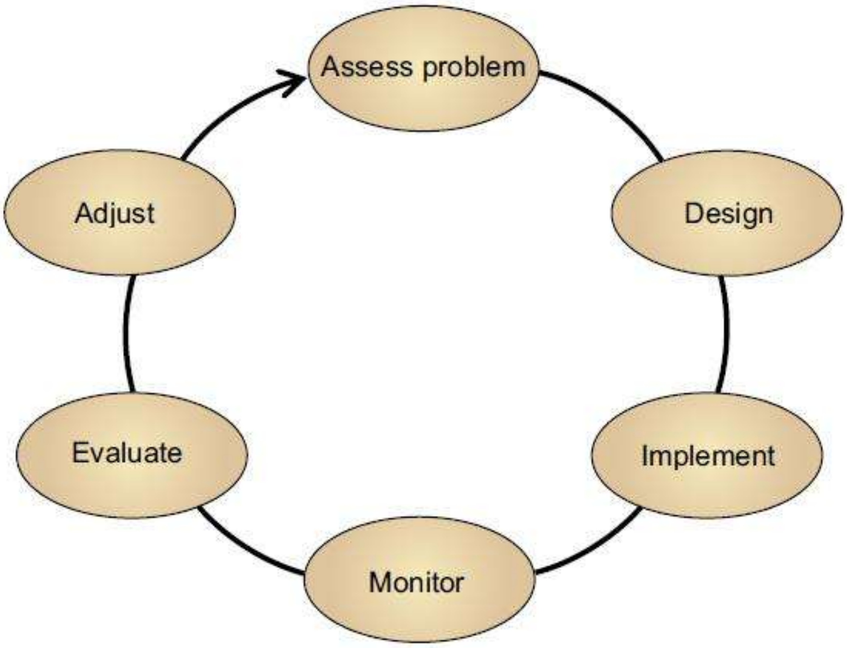

Figure 2: Diagram of the Adaptive Management cyclical process, taken from Williams *et al* (2009)

Consideration of the relationship between each of the generic steps in the Adaptive Management cycle with the specific steps associated with consenting ORE can inform approaches that are most suitable.

- **Assess problem**: When a new proposal for developing ORE arises this will require assessment and decision making. There will be scientific uncertainty arising from the assessment that risk making the conclusions of the assessment open to debate. In the context of managing natural resources using Adaptive Management, the effects of the ORE proposal on biodiversity interests are likely to be a focus of the assessment. The environmental assessment may be highly qualitative or be based on analytical assessment frameworks. The assessment framework provides important context for informing options around which questions can be meaningfully addressed, post-consent, in order to reduce uncertainty. Whilst it can be anticipated that regulators will take a precautionary approach to assessing impacts, debate can still be expected to become polarized around the degree of precaution that is appropriate. Some critics may claim an overly-risk averse approach is taken by regulators with respect to the assessment of impacts of the ORE proposal, whilst others will point to the uncertainties and call for additional risk-aversion.

- **Design:** In the context of licensing of ORE proposals, this is taken to be the design of the project proposal and associated mitigation measures, rather than the design of the monitoring programme. Aspects of the design of the project and associated mitigation will be informed by the environmental assessment. The Design phase is complete when the decision maker has either consented or rejected the proposal. The design of the monitoring programme often begins before a consent decision is made but is not completed until the next stage.
- **Implement:** can be taken to refer to deployment of the ORE devices that have been assessed, but in the context of Adaptive Management also refers to the work in tandem required to prepare for implementation of the post-consent monitoring. The importance of developing the question-led approach is discussed in Section 4 of this document.
- **Monitor:** The data gathering phase of Adaptive Management may begin before deployment of the ORE devices and will continue after it has been deployed.
- **Evaluate:** The analysis and presentation of the results of the monitoring study.
- **Adjust:** Adjustments can be to policies or to some specific aspect of the Adaptive Management plan associated with deployment of the ORE devices (e.g. changes to the mitigation or compensation associated with the project). Given that meaningful monitoring will have reduced uncertainties associated with the impact assessment of ORE devices the most important adjustment can be expected to be in the way the results are used to inform future assessments. Learning outcomes that can be applied to the project (e.g. changes to the mitigation or compensation measures) are considered to be of secondary importance in this context where the principal reason for undertaking Adaptive Management is to improve future decision making regarding the impacts of ORE devices. The onus is on government institutions to ensure the learning outcome is applied in this manner.

The Technical Guidance identifies the scientific attributes of structured decision making processes that lend themselves to an Adaptive Management approach. Post-consent monitoring can be used as a part of the overall Adaptive Management approach. The approach can be applied where the scientific uncertainties that can be reduced through monitoring relate either to qualitative judgements within assessments or rigorous quantitative approaches that inform assessments of impact. In general terms, comprehensive quantification of the uncertainty associated with variable covariates going into modelling will result in a large amount of scientific uncertainty associated with output. This scenario can lead to overly precautionary assumptions with respect to multiple input parameters resulting in uninformative assessments. When coupled with risk-averse approaches to the protection of biodiversity the outcome can be an extremely limited tolerance for the activity combined with extremely limited opportunities to reduce scientific uncertainties.

In the context of licensing MRE projects, decision making relies upon impact assessments that are informed by a combination of qualitative assessment and/or mechanistic modelling of effect pathways (e.g. collision, displacement, barrier, smothering, removal). Some assessments might be highly qualitative in nature, with very limited consideration of numerical data. In more quantitative assessments, a combination of effect pathways may be assessed to account for the cumulative consequences of the effects on populations of animals or upon ecosystems.

When Adaptive Management is being considered, it is important that regulators are mindful that they may not need to reduce scientific uncertainties when making future decisions. The need to reduce the uncertainty associated with putative effects of ORE devices can be deferred where the decision maker already has a sufficient level of confidence. This deferral is sometimes described as retiring the issue. The term deferral is used here to reflect the fact that scientific uncertainty will be expected to still exist with respect to the issue but the utility of reducing the uncertainty for the purpose of decision making is not currently apparent and therefore monitoring to reduce the uncertainty is not a current priority. Those effects whose uncertainties are currently deferred are considerable (Copping *et al*, 2016) and increasing in number. This is a reflection of a maturing regulatory system responding to new evidence as it becomes available. As a consequence, the majority of potential effects associated with ORE devices can be sufficiently mitigated and otherwise managed using standard, agreed approaches. Scenarios where decision making is based upon established standard practice using analytical assessment frameworks and are associated with general consensus around the appropriate responses to scientific uncertainties would not be expected to provide the focus for Adaptive Management. It is the small number of key effects, whose uncertainties are a limiting factor for decision making, that are the most appropriate focus. From the outset monitoring should be targeted towards specific key questions.

Before designing any post-consent monitoring for the purpose of Adaptive Management there are a wide range of issues to be addressed. Agreeing the question to be answered by monitoring, or hypothesis to be tested, the type of data that are to be gathered to robustly answer the question, the techniques to be used to gather the data, and any stakeholders that are involved the post-consent monitoring. All these issues apply equally to any monitoring programme that is for Adaptive Management, irrespective of the topic area or receptor group. Whilst the nascent nature of marine renewable energy makes it the ideal candidate for Adaptive Management, the onus is on the regulatory community to provide leadership that can meaningfully address the technical challenges associated with doing good-quality post-consent monitoring. A good example of collaboration between industry and government bodies is the Joint

Industry Programme approach under the auspices of the Ocean Energy Forum. Developers may lack, or be unable to commit, the necessary resources and expertise required to design and implement such studies. Use of demonstration sites where key questions can be addressed through collaborations bringing together government institutions, academia and expertise based within the ORE industry is likely to provide the greatest value.

Prioritising the questions that should be the focus of Adaptive Management should be informed by the pre-consent Environmental Statement or Appropriate Assessment which are assessments of impact informed by analytical frameworks. Government institutions who have ultimate responsibility for undertaking and advising decision makers on these assessments should be encouraged to develop cultures and approaches that enable Adaptive Management approaches, starting with the manner in which assessments inform the approach. Good practice points towards undertaking this work at an early stage as part of the consenting process, in preference to leaving them entirely unconsidered until the post-consent phase is entered into. More objective decisions can be made about the relative value of reducing scientific uncertainty where the uncertainty relates to quantified parameters in analytical assessment frameworks that use probabilistic approaches. This means being open and transparent within the EIA about the inherent scientific uncertainties, which will enable decision makers to use this information to inform their risk-appetite. This is important given that any risk-based approach must ultimately balance potential financial costs associated with undertaking the activity with potential costs to biodiversity. Current established practice can involve the use of qualitative approaches to assessment or deterministic analytical assessment techniques that do little to provide information on the inherent scientific uncertainties. Coupled with this is a tendency to attribute greater scientific certainty to assessments in an attempt to reconcile the assessment process with partial interpretations of court rulings. An example of this is with respect to implementation of Appropriate Assessments under Article 6(3) and the Waddenzee judgement (Case C-127/02) which makes reference to the need for the assessment to reach conclusions that are “beyond reasonable scientific doubt” with respect to impacts on integrity. Superficially, assessments that avoid quantifying the scientific uncertainty are considered by non-specialists to be easier to reconcile with this judgement, even though they are less informative for risk-based decision making and thus implementing Adaptive Management.

A key consideration for regulators is a growing body of evidence that post-consent monitoring programmes are often not effective at providing results that meaningfully reduce scientific uncertainty and thereby provide information that can give greater confidence to decision makers with respect to future project proposals (MMO, 2014). The name that can be given to post-consent monitoring studies that entail the collection of large quantities of data that do not provide useful information is that they are Data Rich but Information Poor (DRIPy). There are a range of reasons why this outcome can arise, including:

1. Post-consent monitoring studies whose design is not led by questions (hypotheses) that relate to key uncertainties associated with the initial assessment. The risk that questions are poorly specified is greater where initial assessments are highly qualitative and do little to inform which specific parameters might be the focus of post-consent monitoring to reduce scientific uncertainty.
2. Study designs that provide data for which analysis is not able to reduce the scientific uncertainty in a meaningful way. A range of study designs exist and, in general terms, more information can be gleaned from those that allow spatial changes over time to be attributed to the effect mechanism of interest with greater confidence.
3. Study designs that do not provide sufficient data to meaningfully answer the hypothesis with the desired level of confidence. Studies that are underpowered were identified as a key concern by the MMO (2014) with respect to post-consent monitoring of offshore wind farms in the United Kingdom. For smaller scale marine renewable projects, such as wave and tidal, the likelihood of suffering from underpowered data collection are potentially of greater concern because the smaller spatial scales of interest add to the challenge of creating study designs capable of distinguishing the effect mechanism of interest from background variation of the wider system.

Each of these issues can result in the outcome of DRIPy monitoring, and some simple management strategies and recommendations for addressing this concern are the focus of the following sections of this report.

3.4 Summary recommendations for effective Adaptive Management

Adaptive Management is seen as a systematic approach for improving natural resource management, with an emphasis on learning about management outcomes, dissemination of new knowledge and incorporating what is learned into ongoing management. It can be viewed as a special case of structured decision making, which deals with an important subset of decision problems that relate to scientific uncertainty and for which recurrent decisions are needed and an inevitable consequence is that the uncertainty about management outcomes is high.

To achieve the goals associated with Adaptive Management, key changes are likely to be required at an institutional level within Member States. Many of these changes are cultural as well as technical, and involve changing attitudes towards understanding and rationalising risk-based decision making in order to address the challenges presented by scientific uncertainties of the impacts of ORE devices on biodiversity features. Risk-based decision-making requires a collaborative and inclusive approach between regulators, developers and through stakeholder engagement with other interested parties. It cannot be delegated by regulators by giving developers the task to produce Adaptive Management Plans for each of their projects. The evidence collated through the RiCORE workshops indicates this approach is unlikely to provide learning outcomes that can be applied beyond specific projects circumstances to inform future decision making as the industry moves to commercial scales. Where the regulator is not closely engaged with considering their risk-appetite this can re-inforce existing tendencies towards the precautionary principle, and result in overly risk-averse approaches to the conservation of biodiversity. This view is consistent with the emerging draft findings from the Nature Directives REFIT process (Milieu Ltd, 2015). To achieve the aims of Adaptive Management, regulators and their advisory institutions need to support other stakeholders towards an understanding of how risk-based decision making can be re-balanced so that it remains precautionary whilst enabling uncertainties to be meaningfully reduced.

Consenting of ORE devices by regulators needs to be made in the context of scientific uncertainty regarding the level of impact on species afforded protection under the Birds and Habitats Directives, as well as other legislation. Given the high mobility of many of these species, and the manner in which they use large areas of the marine environment, project locations will have a degree of connectivity with the qualifying interests of protected areas even when they appear relatively remote from those areas. Potential effects arising at the project locations will be assessed to have potential impacts on the populations of protected areas. For Adaptive Management to be applied whilst avoiding an adverse effect on integrity of Special Protection Areas and Special Areas of Conservation, there is a clear need for conservation objectives to be reconciled with the need to tolerate a level of potential effect if the scientific uncertainty about the magnitude of the effect is to be meaningfully reduced. This will require a suitably balanced approach to the conservation objectives of statutory protected areas in combination with case-by-case consideration of risk-appetite.

In the context of licensing requirements for more, Adaptive Management cannot be applied without there being sympathetic policy goals in the wider context. Approaches that seek to implement Adaptive Management through requiring developer led plans to identify monitoring are less likely to have synergistic connections with the wider policy context, and are therefore less likely to be effective. The recommendation is that approaches to Adaptive Management that are effectively coordinated by government institutions will have more opportunity of ensuring that the necessary wider policy context needed to support the approach is in place. In this respect the key piece of wider context is likely to be the conservation objectives for protected species and habitats.

Some key take home message are that:

- The Adaptive Management approach should be viewed as an over-arching policy goal of government institutions that should be reconciled with the precautionary principle on a case-by-case basis.

- A key issue for Adaptive Management to meaningfully improve decision making, and be cost-effective, is to ensure post-consent monitoring is not DRIPy.
- Government institutions are likely to be more effective at applying Adaptive Management principles if they have the in-house expertise to develop risk-based approaches to balancing the use of the precautionary principle with the goals associated with reducing scientific uncertainty.
- Government institutions should move towards more open and transparent approaches for assessing scientific uncertainties in order to make better informed decisions, enable Adaptive Management and be more accountable for their level of risk-appetite.
- Coordinated collaborative approaches to post-consent monitoring are likely to be necessary in order to reduce key uncertainties. The coordinating role that Government institutions can provide can facilitate this.
- Structured approaches to Adaptive Management should adopt question-led approaches to monitoring, with careful consideration of the value of the study design and the risks associated with statistically underpowered studies. It is important that monitoring plans are focused and avoid trying to achieve too much at once.

4. Guidance for informative post-consent monitoring

4.1 Introduction

This section is divided into discrete parts, each offering guidance to enable informative post-consent monitoring that supports future consenting decisions with respect to ORE. Firstly the question-led approach is described, explaining its purpose with discussion on effective implementation. Next, the options with respect to study design that can be adopted to answer the question are considered, in the context of ORE. Latter sections then consider the relevance of statistical power and inference, along with meta-analysis, and their roles in achieving the goal of informative post-consent monitoring is again considered in the context of ORE.

4.2 The question-led approach

Critical to the success of any post-consent monitoring programme will be the clarity and relevance of the study question for informing decision making and the ability of the data gathered to meaningfully answer the question. The purpose of adopting a question-led approach is to maximise the likelihood the outcome will avoid being DRIPy from the outset. The section provides some simple hints and tips to help implement the question-led approach.

Conceptually the question-led approach is very simple. It can be summarised by the following 3 steps:

1. Define the question, based on a hypothesis.
2. Specify the data collection methods and analysis techniques needed to meaningfully answer the question by testing the hypothesis.
3. Decide to either implement the proposed monitoring campaign, define a more appropriate question or choose to not undertake post-consent monitoring.

4.2.1 Step 1 – defining the question

Specifying the question to be asked is often challenging, and may take considerable iterative communication to refine. Refining the question may require identification and description of the potential effect pathway, or impact, under consideration. Or it may relate to an assumed parameter in a mechanistic model of the effect pathway (e.g. an assumption associated with modelling collision risk). Part of the process of defining the question is also likely to entail clarifying the possible outcomes associated with answering the question. Where the goal is to improve future decision making it is likely the question will relate to assumptions in EIA or Appropriate Assessment (AA) documents.

A well defined question will maximise the chances that the monitoring programme will meaningfully reduce scientific uncertainty. The change that is of interest to regulators of ORE often relates to differences in the abundance and/or distribution of a receptor specific interest as a consequence of the activity. The following examples are progressively more informative with respect to the answers they can provide, and the answers would be expected to inform future assumptions in EIA/AAs regarding rates of displacement:

1. Has there been a specified change in animal numbers at the location over the period?
2. Has there been a specified redistribution of animals at the location over the period?
3. Has there been both a specified change and a redistribution of animals at the location over the period?
4. Can the specified change be associated with a particular location and/or activity?

Does the question require comparison? Questions that are highly relevant to Adaptive Management might not entail before-after comparison but will still meet the key criteria of having a clear link with the aim of reducing scientific uncertainty associated with the use of analytical assessment frameworks in future decision making. Examples are questions that provide information to enable additional factors to be parameterized in the analyses that form assessment frameworks by creating a baseline of understanding e.g.

1. How does animal density at the location vary with tidal state?
2. What is the dive-depth profile of animals?
3. What is the dive duration of animals?

Decision makers and managers should bear in mind that comparisons pre- and post-deployment or with and without activity, for example, are not always required when trying to reduce uncertainty associated with future decision making. Questions that do not entail comparison with and without the activity may be of greater priority at the early stages of creating assessment frameworks when the analytical techniques being used are poorly parameterized owing to limited baseline data. It may also be the case that these sorts of questions can be answered more cost-effectively than attempting to measure relatively small changes that are difficult to detect arising from comparison of before-after presence of devices. If the monitoring done post-consent is to be focused on these questions, then it would be considered inappropriate for the decision making process for current project proposals to be delayed pending the information.

For practitioners of Adaptive Management, it will be essential that a process is established to ensure that all stakeholders are able to contribute to the question-led approach. This helps ensure that the data that are gathered are going to be as useful as reasonably possible. Key points include:

- Ensure the question is clearly defined at the outset; it is essential to get this right rather than hoping *post hoc* analysis of data will be answer a useful question. Insufficiently well-defined questions are a key reason why post-consent monitoring studies can be DRIPy.
- Is there a clear link between answering the question and meaningfully improving future decision making? If the question is only applicable to a single project it will not provide wider learning outcomes for future projects, making it less useful.
- Establishing an iterative process with input from stakeholders can provide a good approach to ensuring weaknesses associated with poorly defined questions are identified and addressed.
- Being more specific will generally increase the likelihood that the question will be meaningfully answered. For example, what is the effect size of interest? Are you concerned with the risk of false positives and false negative error rates? Over what spatial and temporal periods does the question relate to? Clarifying these aspects can ensure the appropriate levels of data are gathered. Getting statistical advice at the outset is recommended.

Quotes that reflect the value of defining the question

"To call in the statistician after the experiment is done may be no more than asking him to perform a post-mortem examination: he may be able to say what the experiment died of" Sir Ronald Aylmer Fisher

"The combination of some data and an aching desire for an answer does not ensure that a reasonable answer can be extracted from a given body of data" John Tukey

4.2.2 Step 2 – specify the data collection and analysis methods

It is often the case that different techniques are capable of gathering the data to answer the question. For example, surveys along transects to estimate cetacean and/or seabird abundance and distribution at sea can use different platforms e.g. boats, observers in planes or digital equipment on a plane. Culloch *et al* (2015) review the utility of different techniques for monitoring at proposed ORE locations pre-consent, and their findings apply as equally to post-consent monitoring as they do to site characterisation.

In the context of developing the question-led approach, a simple piece of guidance is to guard against situations where the approach to post-consent monitoring becomes technique-led, based on the specialist knowledge and skills of the person tasked with designing the monitoring. Ideally the question-led process should consider various methods that can be used to answer the question with the most suitable being chosen.

4.2.3 Step 3 – implement, refine or discard the monitoring programme

The iterative nature of the question-led approach means it should remain open to refinement. This is because the practicalities of meaningfully answering questions through post-consent monitoring are often not fully apparent at the outset of defining a monitoring programme through the question-led approach. A concern that often arises is that insufficient data will be collected by a monitoring programme to meaningfully answer the question. To address this, the question-led approach might be refined in a number of ways and the trade-offs that arise should be carefully considered by decision makers:

- The temporal and spatial scale and associated sampling intensity could be increased which would add cost.
- The effect size of interest could be increased which will reduce the costs associated with sampling (e.g. a 10% decrease in abundance could be changed to a 50% decrease in abundance), and expose the receptor to greater risk of change.
- The false positive and false negative error rates may be made less stringent. Increasing the false positive rate will make it more likely the results will falsely conclude that change occurred, which in turn will make it more likely that management action is falsely assumed to be required to regulate the observed change. Increasing the false negative rate will make it more likely that no change is detected by the study, even if an effect is occurring. In other words, a decrease in abundance, for example, would go undetected, which may result in the erroneous conclusion that monitoring/mitigation is not required or can be reduced.

Ultimately a pragmatic approach will need to be taken by decision makers that balances the goals associated with reducing scientific uncertainty, enabling human activities and avoiding unacceptable impacts on the environment. Clear specification of the various attributes discussed above should inform those scenarios where it will not be realistic to undertake post-consent monitoring that can meaningfully reduce scientific uncertainty associated with future decision making. It may be clear that the level of sampling will cost too much or take too long, or that the level of risk to a receptor is too great to allow meaningful monitoring to be undertaken. Whatever the underlying reasons, if the conclusion is that meaningful monitoring cannot be delivered then there is no value in implementing the monitoring programme. This can be demonstrated to be a legitimate and reasoned conclusion where the decision is transparently accounted for under a question-led approach that includes stakeholders in the decision making process. The implication for the project would be that monitoring was not required by the decision-maker.

4.3 The study design

It is imperative that studies are designed in a manner that allows the scientific uncertainty to be reduced in a meaningful way. It is often easier to design a study that can answer the question, yet, with additional consideration, it can be possible to answer the question in a more useful manner that can inform future developments. With respect to questions that relate to changes in presence and distribution arising from an activity, a range of approaches to study designs exist. In general, more information can be gleaned from monitoring programmes that address spatial changes over time. Some are more suitable for some receptors than others. In very general terms the more mobile the receptor of interest the more likely it is that more advanced designs will be required in order to provide adequate information with confidence. Types of design that inform comparison are discussed below with a series of figures illustrating each example. The examples show the abundance and distribution of a mobile receptor with blue indicating absence, green low density and red highest density. Numbering is indicative of latitude and longitude.

4.3.1 Before-after

The comparison between the study area before the event and after the event is shown in Figure 3. This design provides information to answer the study question in relation to a single factor of the difference before with after. A fundamental limitation of this study design is that it does not provide information through the use of a control site on the likelihood that the change is attributable to the mechanism of interest, as opposed to natural stochastic influences, for example.

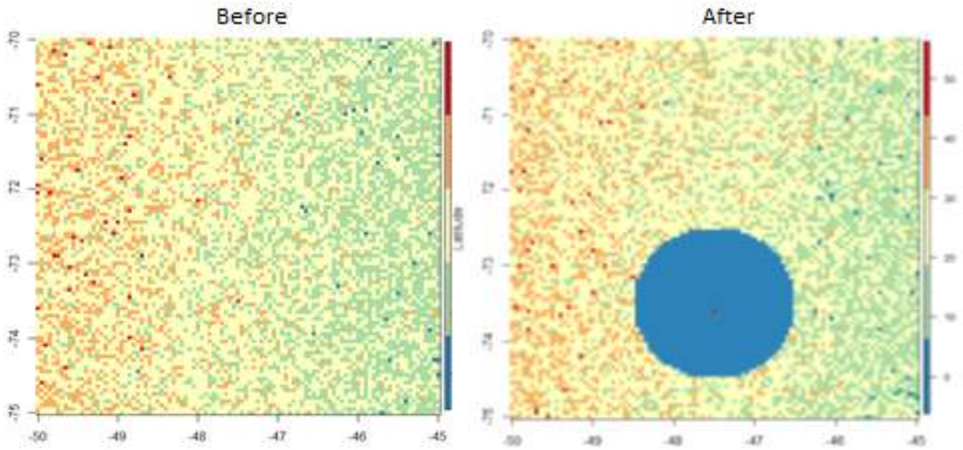

Figure 3 Before-after study design (blue circle identifies the predicted? impact area)

4.3.2 Control-impact

Studies that compare the study area after the event with a control area are illustrated in Figure 4 (outlined in red). This design provides information to answer the study question in relation to a single factor of the difference between control and the impact areas. A limitation of this study design is that it does not provide information on the status of either area before the event, and it requires the assumption of knowing where the impact area is exactly.

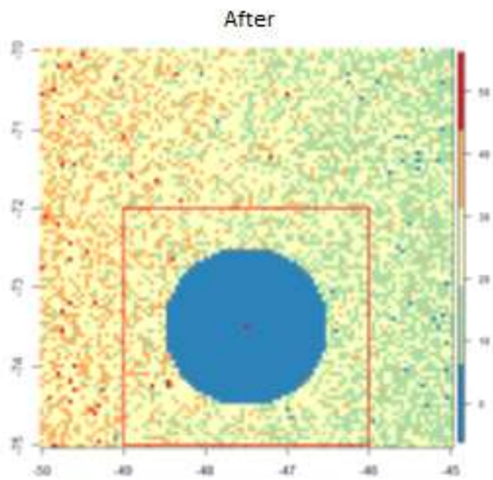

Figure 4. Control-impact study design (red square delineates the control area)

4.3.3 Before-after-control-impact

The Before-after-control-impact (BACI) study design, illustrated in Figure 5, is able to test against each of the two factors individually addressed by the previous designs: differences between before and after, and between control and impact. Its ability to address the limitations associated with each of the previous two designs makes it considerably more robust. However, BACI analysis are unable to provide information on effects that occur across a gradient within the study area (e.g. a redistribution resulting from a gradient of displacement).

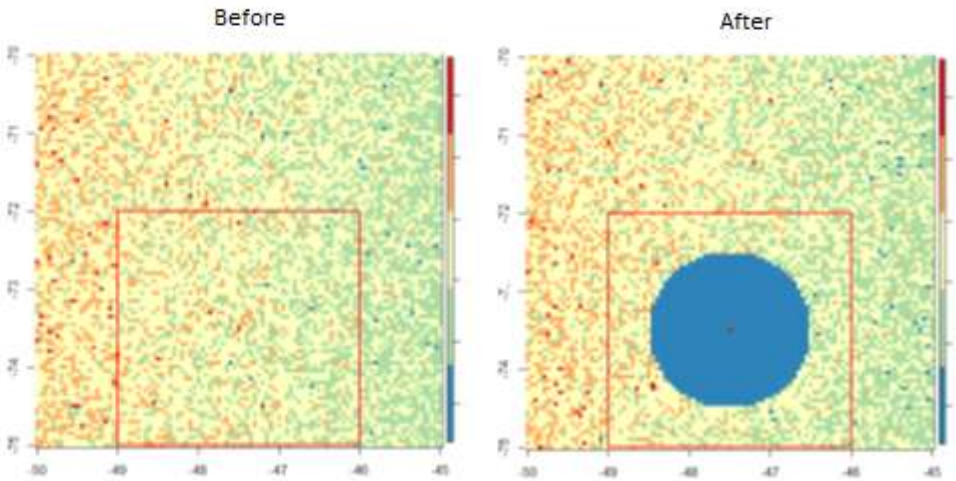

Figure 5. Before-after-control-impact (BACI) study design

4.3.4 Before-after gradient

The before-after gradient analysis illustrated in Figure 6 allows a gradient of effect to be measured at (standardised) intervals from a point location, often the source of the potential impact. The advantage of this approach over the BACI approach is that changes in distribution throughout the study area are quantified. However, it functions best where there is one source/point as it cannot make inferences regarding multiple different gradients across a site.

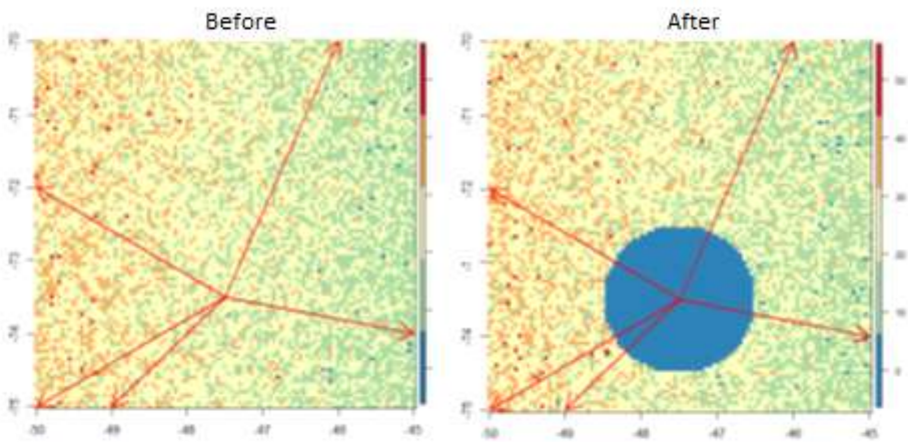

Figure 6. Before-after gradient (BAG) study design (red lines indicate gradients from the point source)

4.3.5 Spatial analysis

A comprehensive form of spatial analysis is illustrated in Figure 7. This is an advanced technique which is also highly specialised. Detailed worked examples of spatial analysis are presented in the MRSea modelling package guidance notes (Scott-Hayward *et al* 2013). This approach is likely to give the best quality of information with respect to identifying the impacts of MREDs, as it has the potential to perform well at describing both the underlying ecological processes across a site and at identifying spatially explicit change.

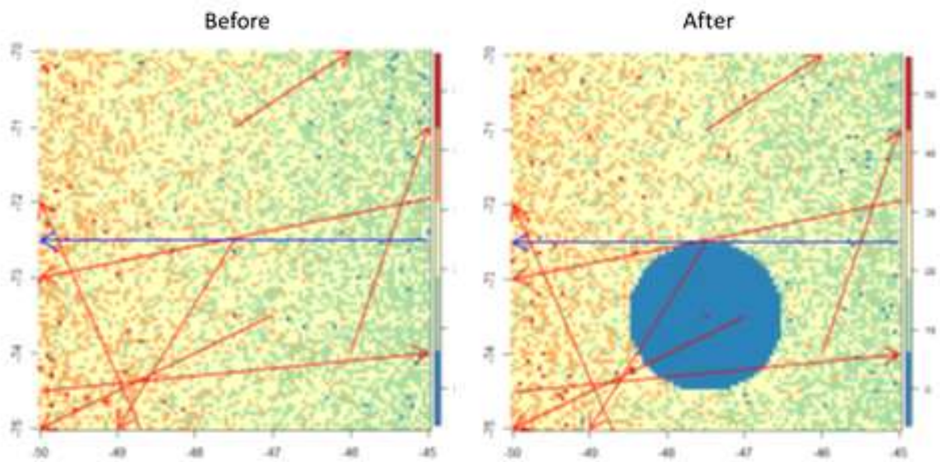

Figure 7. Spatial analysis study design (blue line is a gradient associated with the underlying surface)

4.4 Further consideration of a question-led approach and the resulting study design

This section focuses on examples of marine fish monitoring, but the issues discussed apply equally to other marine receptors. When considering post consent monitoring of offshore renewables on marine fish species, it is often the case that we are concerned about monitoring specific species of concern, as identified through the pre-consent impact assessments. The aim of this is to better understand the extent to which an offshore development is responsible for change. The utility of post-consent monitoring in the context of Adaptive Management is that the overall goal is to reduce the scientific uncertainty associated with potential changes identified at the pre-consent assessment stage.

Below, we take a look at Scotland's first established offshore wind farm, Robin Rigg in the Solway Firth, which has fulfilled its post-consent monitoring requirements for marine fish species. In addition, we consider the findings of a more general review of monitoring programmes by Lindeboom *et al* (2015), which also considered the lessons learned to-date, and provided recommendations for the future. Conclusions are drawn on existing experience and potential reasons why post-consent monitoring may not always achieve the goal of meaningful learning from a question-led approach. Although the examples discussed are for offshore wind, the conclusions apply equally to post-consent monitoring of other activities including new technologies such as wave and tidal renewable energy projects, and are therefore pertinent to the RiCORE project.

The aim of the Robin Rigg Marine Environment Monitoring Programme (MEMP), set in place to comply with condition 6.4 of the Section 36 consent under the Electricity Act, was: *"to detect and/or predict direct and indirect adverse impacts, likely to have a significant effect on the marine environment, arising from the pre-construction, construction, operation and decommissioning of the wind farm".* The MEMP states that: *"The remit of the Monitoring Programme will be to allow changes to the physical and ecological environment caused by the construction and operation of the wind farm to be recorded principally in areas where there is some uncertainty in the effects of the wind farm on the receiving environment, where those effects are potentially damaging. The monitoring programme should be designed so that if potentially adverse significant impacts are predicted which can be reasonably attributed to the wind farm, mitigation measures can be adopted in time to avoid irreversible significant impacts."*

For non-migratory fish, the key potential ecological pathways identified within the MEMP were split into two considerations, those during construction and / or decommissioning and those during operation:

Construction and / or decommissioning

- Augering of pile shafts and trenching cables;

- Piling of turbine foundations; and
- Accidental spillage of hydraulic fluids, lubricants, fuel etc.

Operational

- Presence of sacrificial anodes on foundations;
- Presence of cables on sea bed;
- Noise of turbines; and
- Physical presence of turbine foundations.

The analysis undertaken on the fish and epibenthic data used species assemblages, abundance and size frequency gained from trawl surveys within, and in the vicinity of, the wind farm site, to identify any temporal or spatial trends. Data was also used to determine whether the construction and operation of the wind farm may be linked to these trends. Although considered at the time to be a standard approach, which was supported by scientific peers, in practice the monitoring design collected data in a before-after reference area with no control. The purpose was to detect potential changes in overall abundance and distribution. It did not seek to answer more specific questions relating to the potential ecological pathways described and it did not use a control site to provide added information and confidence in results.

The year 3 operational report provides a comparative discussion on the survey results pre-construction (baseline), during construction and post construction considering temporal and spatial differences, as well as the impact of the wind farm vs. natural fluctuations. The overall conclusion states that it "*provides evidence that broad scale changes in fish and invertebrate communities are unlikely to occur at a magnitude beyond natural spatial and temporal variation*". The report also finds that "*the results of the MEMP non-migratory fish surveys highlights the difficulty in identifying impacts in fish assemblages resulting from the construction and operation of offshore wind farms*".

The year 3 report considers it is difficult, within the experimental design used, to attribute the differences that are detected to the presence of the wind farm, considering them more likely to be attributable to natural variation. In a study of offshore wind park monitoring programmes from across the North Sea, lessons learned and recommendations for the future, Lindeboom *et al* (2015), writes that basic monitoring has to be rationalised at the level of the likelihood of impact detection, the meaningfulness of impact size and representativeness of the findings. Targeted monitoring is crucial and should continue to be applied to disentangle processes behind observed changes. The example of overarching artificial reef effect caused by wind farms is provided. Among the lessons learned reported by Lindeboom *et al* (2015), is that basic monitoring by itself is not sufficient to disentangle specific cause-effect relationships, especially in systems with a high natural variability. This latter point is consistent with the findings of the Robin Rigg year 3 report, which states that it is inherently difficult to disentangle natural drivers from anthropogenic drivers such as the construction and operation of an offshore wind farm.

Whilst the Robin Rigg MEMP was within the context of its peers at the time, and addressed the requirements of the Scottish Ministers, it can also be described as an example of DRIPy monitoring, which has a very limited utility for future decision making. The underlying reason for this example being DRIPy is that the before-after study design was unable to meaningfully disentangle cause-effect relationships, and this risk was not identified when the question to be monitored was agreed upon. This raises the issue of what research questions can provide meaningful answers for decision making through post-consent monitoring studies. It is clear that more general and broader questions (such as 'has there been a significant change at the location'?), may not provide information that will contribute to helping reduce the scientific uncertainties associated with future decision making regarding the impacts of renewable energy devices on fish ecology. This applies equally to other receptors.

Dolan *et al* (2016) write that conducting monitoring studies with adequate power in variable habitats can be costly. Given that underpowered effort can also result in DRIPy monitoring, this is a further important consideration. The key lessons learned are that the question-led approach must not merely concern itself with detecting change, but with detecting change that is attributable to cause-effect relationships. It is essential thatsuch relationships are meaningfully monitored with sufficient statistical power using a sufficiently informative study design to enable improved future decision making. A fuller discussion of statistical power, its importance and use are provided in the following section.

4.5 Statistical power and inference from data

4.5.1 Introduction

Open and transparent decision making with respect to the desired levels of confidence in the results of monitoring can be informed by an understanding of statistical inference. Doing so will increase the likelihood that post-consent monitoring provides meaningful data and that the work is supported and understood by stakeholders. This section provides a technical overview of relevant issues.

4.5.2 Relevance

Statistical inference is used in science to best interpret and predict what goes on in the real world. Results obtained from a sample do not always reflect the status of the population of interest from which the sample is drawn because sampling error is always present; but the results can aim to be informative. It follows that, our ability to determine whether or not a difference measured in a study is simply due to chance. Measuring a change with 100% certainty is not possible, unless a complete census is undertaken. However, it is possible to estimate the degree of uncertainty associated with detecting that difference.

Once limits are set as to how much uncertainty is deemed acceptable, a suitable analytical approach is employed to test the null hypothesis, i.e. “there is no difference” or “there is no effect”. The outcome of the test is either to reject the null hypothesis in favour of the effect tested for, or accept the null hypothesis of no difference.

Whenever the null hypothesis is being tested, four possible situations occur (Table 1). When the results reflect ‘truth’, the null hypothesis is correctly either rejected or accepted. When the results obtained do not reflect what is taking place in the population, then the null hypothesis is erroneously rejected (‘Type I error’) or accepted (‘Type II error’).

		TRUTH	
		Change Absent	Change Present
ESTIMATE	Change Absent	Correct classification, True Negative, (1-α)	False Negative, Type II Error, β
	Change Present	False Positive, Type I Error, α	True Positive Correct classification, Power (1-β)

Table 1. The four alternative outcomes in tests of hypotheses

Clearly both types of error have important consequences and therefore both should be considered. The means of assessing and setting acceptable levels of risk associated with these errors is based on probability. The null hypothesis is rejected whenever the test shows that the probability of having done so incorrectly is below an arbitrary critical level. This is called alpha (α), or false positive rate. Convention often results in the value at which the rate false positives are considered “statistically significant” is at 0.05 i.e. 5%; it is common to interpret the result of such a study to be a true reflection of the population, even though that is not the case 5% of the time, or in 1 out of 20 occasions.

When the null hypothesis is not rejected, there is a risk that a type II error is occurring, giving a false negative result. The probability of Type II error is called beta (β), or false negative rate, and is used most commonly as (1-β) i.e. the probability of avoiding a Type II error. This is referred to as the statistical power of the study. When a study has low power, there is a high probability that no significant difference (as dictated by the false negative rate) will be found, even if it existed.

Both errors are also functions of effect size, sample size and variance. How each of these can be used in the design of monitoring activities is briefly outlined below.

- Effect size is the magnitude of change to be detected (e.g. a 10% reduction in abundance of a species over a specified area and period of time). Effect size may be increased 'experimentally' by increasing the difference in the treatments. When using statistical power to determine sample size, power is always calculated relative to a chosen 'effect size', i.e. the probability that a difference of a given magnitude would result in a statistically informative test. Choice of appropriate effect size is not straightforward; it is linked to choice of parameters and type of change to be measured, and tolerance of change based on a combination of biological, social-political, and statutory considerations in addition to what is statistically desirable.
- Sample size (i.e. the number of samples taken); once significance, power and effect size are defined, it is then possible, using power analysis, to calculate the necessary sample size to achieve them.
- Variance (at times referred to as 'noise' in the data) is a measure of variability in the data. It can be reduced by improving methodology and improving experimental / sampling design. However, in ecology 'noise' is often unavoidably large given natural fluctuations and complexity of interactions at different spatial and temporal scales.

Increasing the false positive rate (e.g. from 0.05 to 0.2) will increase the probability of Type 1 error from 5% of the results to 20% of the results. The same applies for the false negative rate. Both rates are inversely related to one another so that for a given sample size if one rate is increased, the other is reduced. Setting the level for false positives will determine the false negative rate; for example, choosing a more lenient level for the false positive rate will result in increased power and a reduced false negative rate (and *vice versa*). Setting a more stringent standard for the false positive rate will make it more likely that false negatives will occur. The ratio represents the balance between the rates of false positives and false negatives.

4.5.3 The consequences of reaching false positive and false negative conclusions

The consequences of the errors associated with false positives and false negatives can be elucidated through a hypothetical example relating to the effects of ORE, as follows.

A tidal stream channel formed by the narrows between the mainland and a nearby island has the potential to be used to generate electricity from tidal stream generators. However, the area is also used by harbour seals which are a protected species under EU law, whose population is potentially sensitive to the effects of mortality arising from collision of individuals during the operation of tidal turbines. In order to reduce uncertainty associated with these potential effects and better inform appropriate management measures in the longer-term, it has been decided to gather data on the spatial use of animals in the vicinity of operational turbines in order to estimate differences in distribution at various spatial scales.

The aim of the study is to enable the probability of collision to be estimated based upon data on the behaviour of animals in proximity to the rotating turbines. The null hypothesis to be tested is that the presence of operational turbines makes no difference to the distribution of seals. If the null hypothesis is true, then the true rate at which collisions occur is not changed by the response behaviour. The experimental design used has sought to account for natural variation in the behaviour and distribution of harbour seals in the area of interest. If a false negative conclusion of no change (Type II error) was reached, then no avoidance behaviour would have been recorded and the rate of collisions would be estimated to be different to the true rate. Harbour seals would be assumed to be colliding with the turbine more frequently than in reality. In terms of management consequences, it may lead to a conclusion that there is concern with respect to the assumed level of damage to the population even although less damage to the population is actually taking place. In this example the cost to the population of seals would be more acceptable but the cost to the tidal industry would be greater than necessary; being either a reduced number of turbines deployed or costly and unnecessary mitigation.

When a false positive occurs, it means that the study has detected a difference (of given magnitude) where that difference in reality, does not exist. The study would erroneously conclude that the collision rate can be safely relaxed from the previous level, when the truth is the collision rate is at the same level as previously assumed before the study. It follows that the management response will be to either permit more tidal turbines than is appropriate, or to permit the development without mitigation measures, which are required to adequately protect the seal population. In this example, the costs of such an error have a negative consequence on the seal population. If seals were attracted to the operational device then the costs associated with each type of false conclusion would be reversed.

4.5.4 Information from power analysis

Power analysis is a calculation of the sample size required to detect a specified effect size, given thresholds for effect determination and desired power to detect the effect. It can also be used to estimate the power to detect change given a presumed effect size and assumed sample size. As such, power analyses can enable a better informed approach to decision making regarding the level of effort required to detect meaningful changes post-consent. They are particularly useful in the context of assisting in studies designed to have sufficient power to detect changes and thus can help avoid undertaking monitoring that is unable to identify statistically meaningful changes. For example, it can be used to address an established concern that genuine changes in abundance and distribution are not likely to be found even though the possibility exists that they are changes that the regulator may consider to be meaningful (MMO, 2014). Whilst simple in concept, methods of power analysis are complicated by additional features of data such as non-linear relationships and spatio-temporal autocorrelation. The power to detect change can be affected by both the spatio-temporal model being used to detect change and the data itself, making the specifics of data-model combinations an important consideration.

In the renewables context, power analysis is likely to have utility at various spatial scales. It can be used to quantify the power to detect site-wide changes and the power to detect more local changes in distributional patterns, e.g. in and around turbines. Analysis could be used to inform data collection for various scenarios e.g.

- No change during operation
- Decline in abundance during operation, but no redistribution
- No decline in abundance, but a redistribution
- Decline in abundance and a redistribution

To date, regulators have paid relatively little attention to the issue of statistical power with respect to post-consent monitoring. This should change where practical. For example, the Scottish Government is now more frequently working with statutory advisors and developers with the goal of enabling consideration of statistical power to inform planning for post-consent monitoring.

Levels of significance and power established following convention are often unbalanced. Established practices of setting the false positive rate at 0.05 and giving no consideration to what the statistical power should be will be expected to result in a tendency to design studies with too few samples to meaningfully inform false negative rates. This situation may be more likely to arise where cost is driving decision making regarding the study design. Where it is set, the false negative rate is often 0.2, meaning that the rate of false positives are 4 times more than the number of false negatives. In some cases, this may be considered reasonable or even imperative. A useful parallel (Peterman & M'Gonigle, 1992) is that of the judicial system where the use of the higher standard of proof in criminal prosecutions (proof must be 'beyond any reasonable doubt') makes it less likely that an innocent person will be convicted (false positive rate), but also more likely that a guilty person will go free (false negative rate). However, should this also be the case for monitoring in our context? Should different scenarios and levels of risk dictate different requirements for significance and power? If the core principle is that decision makers should understand the influence of their appetite for risk on the design of monitoring undertaken, should this provide the logical starting point for determining the rate of false positives and false negatives?

A case can be made to vary the rates of false positives and false negatives according to the balance of their consequences, and the potential for the study to improve overall levels of confidence, and in relation to the overall cost of the monitoring activity. As proposed by several authors (DiStefano, 2003; Mapstone, 1995), the ratio between the two should be decided by evaluating the relative consequences of each error in terms of the respective economic and environmental costs involved. The decision on a ratio is not exclusively a scientific or a statistical one, but ultimately a societal one that may be informed by wider considerations such as whether or not statutory protection is afforded to the species.

Decision making on the acceptable rates of false positives and negatives is likely to be informed to some degree by the cost of monitoring. Trade-offs will always exist between the benefits of understanding the relationship between an activity and the receptor it impacts upon in order to inform management policies and decisions. Within the overall context of monitoring to reduce scientific uncertainty, resources could be channelled towards those monitoring activities where risks of committing errors have high consequences either from the economic or environmental perspective. The acceptable rates for false positives and false negatives should be defined with the necessary caveat that monitoring an indicator with very limited precision or power is potentially a waste of limited resources.

Finally, all stakeholders including scientists, developers and environmental organisations, together with decision makers, should provide views on acceptable rates for false positives and false negatives. This process can inform a wider understanding of the cost-effectiveness of different designs to maximise the levels of significance and power within the constraints of both monitoring budgets and tolerance levels of change to protected populations. In the context of post-consent monitoring for the purpose of adaptive management, it is useful to refer to the ultimate purpose of monitoring in order to evaluate the consequences of each error and to use this information to set levels for significance and power accordingly. This provides clarity on how monitoring will be used to inform future decision making and can provide a transparent rationale for how the risks associated with the different errors may influence the acceptable rates for false positives and false negatives.

4.5.5 Levels of significance and power

A study at the European Marine Energy Centre investigated the displacement rate of seabirds and marine mammals as a consequence of ORE devices, (DMP, 2010). Specifically, the impact indicator was based on the ability to estimate the change in the abundance and distribution of seabird and marine mammal species below a given threshold, as a consequence of the introduction of ORE devices. It is an example of power analsysis of land-based survey data. In this case the threshold level for the effect is given as a 50% change in relative abundance across the site. The false positive rate is set at 5%. The consequence of a false positive conclusion (i.e. erroneously assuming that the level of displacement is greater than the threshold and unacceptably impacting seabird populations) is expected to carry direct consequences for the future management of the activity. False negatives are set at under 10% for key seabird species, and at that level of sampling effort, they remain at over 50% for key marine mammal species. Clearly the reality of a false negative conclusion still remains for the key seabird species, i.e. erroneously assuming the displacement rate is below the threshold; which may result in negative consequences that are greater than a 50% change in relative abundance. For marine mammals, given that false negative conclusions can be expected more than half the time, it is clear that the results of the survey are unlikely to be meaningful in the context of informing decision making regarding future management of the activity.

Early and direct consultation with stakeholders on acceptable rates for false positives and false negatives is beneficial given the close link between the results of the statistical analyses and action taken in terms of future management. It can ensure that the costs to industry and that the risks to biodiversity are made explicit to all, thereby ensuring that the most appropriate future management scenarios are considered. As part of the process, the overall cost of monitoring will have to be balanced against the overall need for evidence. Let us assume a case where costs are both balanced equally between false negative and false positive rates and both are set at a low level. To maintain a balanced ratio for the false negative rate, high power is also a requirement, which in turn will mean more resources to pay for the adequately large sample size. If a high degree of confidence is sought that the results of monitoring are a true reflection of the impacts they cause then initial views might be that both rates are set to 0.05 or even less. Setting such stringent standards for monitoring may not be realistic given the high levels of natural variability associated with natural systems at the spatial and temporal scales of interest (Culloch *et al*, 2015).

Defining what is regarded as 'significant' in terms of environmental change is a complex issue. However, wherever possible, the regulator should define what is considered to be a significant environmental change in terms of the species of interest found at a location, taking account of the quantitative analysis that is to take place. It may then be possible to define the minimum observation effort needed to detect such a change at an acceptable level of confidence through the use of power analysis. This is likely to make future data gathering as efficient as it can be in terms of time and resources required. By targeting the available resources towards the key parameters of environmental concern, then EIAs will be more informative.

Sequentially testing for an effect through iterative settings of the rates for false positives and false negatives is one approach to employing power analysis that could be applied to post-consent monitoring at MRE sites. Such approaches are commonly used in medical clinical trials (Dmitrienko *et al*, 2003 & Westfall, 2001). For example, if it is obvious that the monitoring will not detect the effect size of interest, or if the effect size is detected with the degree of confidence required more quickly than initially anticipated, then the monitoring programme can be stopped early on. Cost savings might be feasible by setting either the false positive or false negative rate to more relaxed levels if the risk of committing an error is managed in a manner that still addresses the key concerns of stakeholders.

In the boxed example considered below on the effect of displacement of seabirds around OREs, the impact of a given displacement rate can be assumed to be energetic costs to individual foraging birds that in turn could have consequences for their individual survival and ability to rear chicks successfully. A false positive result from monitoring would have costs for industry, as the results would indicate a threat to populations of seabirds from displacement that did not exist in reality. A false negative result would have costs to seabirds, as the result would indicate that the level of impact was less than the true level of impact. Consequently, the regulator may choose to set the false positive threshold at a level agreed to be appropriate to afford sufficient confidence that the seabird populations are adequately protected. A false negative rate of 0.2 is considered pragmatic and has been used by regulators in Scotland.

Sequential testing example

The monitoring is concerned with answering the question of whether the displacement rate of a seabird species within a 1km area of an ORE array is at most 40%. Displacement is assumed to result in reduced adult survival and chick productivity rates. The EIA assessment indicated that up to a 40% displacement rate represented a safe magnitude of effect that was coherent with the conservation objectives of the species at their protected breeding site. Conventionally, a null hypothesis would be regarded as a change that the experiment is unable to measure (i.e. no change). However in this example the null hypothesis that we expect to reject is that the change will be greater than 40%:

$$H_0\text{: displacement} > 40\%$$

$$H_1\text{: displacement} <= 40\%$$

In order to give confidence that the results will ensure the conservation of the species the level at which the statistical power of the test is measured is for a displacement rate set at 30%. This sets the level for the test that is precautionary with respect to the threshold established by the hypotheses. The statutory advisor considers this is necessary to sufficiently guard against the risk that the study fails to identify a displacement rate of more than 40%. Power analysis is used to ensure sufficient data are collected. It is considered appropriate to set the power of the test at 0.8. To minimize time and cost a one-tailed test* is undertaken, testing only for false positives. In other words, if the results indicate the displacement rate is a 30% reduction this would be a false result on 1 in 5 occasions.

The false negative rate has not been controlled by the one-tailed test. This means that if the results do not allow the null hypothesis to be rejected, the likelihood that this is owing to a false negative result is unknown. In such a scenario, additional data would be gathered to allow sequential testing to consider the risk of a false negative result, without reducing the overall power of the experiment of 0.8. For example the test to measure the displacement rate could be sequentially applied at first to a false negative rate of 0.5, and then if the null hypothesis is still not rejected, to a more stringent level such as 0.05. Implementation of management measures to ensure protection of the population from an assumed displacement rate of more than 40% would occur if the point estimate for the displacement rate and the associated uncertainty supported action with sufficient confidence.

Sequential testing is a statistically advanced technique, which is not straightforward to implement in practice. Controlling the power needed for an entire sequential testing procedure is likely to be informed by simulation. It will also require the regulator to have a clear decision making procedure with “stop rules” at pre-defined points.

It is worth noting that cost-benefit analysis may find that extending a monitoring programme over multiple years may increase costs owing to factors such as cost of hiring equipment over several seasons. In which case a sequential approach would not reduce costs.

* A one-tailed statistical test can be used if only one of the two types of potential false result is assumed to be of interest. A benefit of one-tailed tests is that they will require fewer data to acquire the necessary level of statistical power compared to a two-tailed test. Consequently designing the testing procedure to ensure the required information is provided by a one-tailed test will provide the greatest potential cost savings. This is achieved by ensuring the hypothesis relates to the more acceptable of the two possible outcomes for the environmental receptor that is being tested.

4.6 Meta-analysis

4.6.1 Introduction

Meta-analysis is a statistical framework for synthesizing and comparing the results of multiple studies, which have all looked at the comparable study questions, often finding different results. It provides a means of answering broader questions than the individual studies were designed to do. It has the potential to be particularly useful in the context of noisy ecological systems and holds potential to provide information on the changes associated with ORE even if the individual experiments have relied upon small sample sizes due to time and cost, and therefore have low statistical power. A review by Harrison (2011) on the utility of meta-analysis, key issues, and a “to do” list for meta-analysis identifies key aspects that are relevant to the context of post-consent monitoring of ORE. This section considers the most relevant parts of that paper.

Where a number of studies have considered the same hypotheses but reached different results, a simple reading of the contrasting studies may give the reader no basis for understanding if any patterns explain true effects or are a result of variation in how the studies were conducted and/or low sample sizes. By establishing a formal statistical framework, meta-analysis can provide 2 things:

- A method to calculate the mean effect, across all the studies
- To explain variation across different studies arising as a consequence of defined ecological factors that may serve to affect the intensity of the effect at different study sites.

Meta-analysis is possible where effect sizes, the associated variance of the estimate, and study effort have been reported. It should not be conflated with highly qualitative analysis of previous studies that merely count the numbers that have reached a conclusion of significance. This cannot be considered robust, as it lacks statistical power. Even if a large number of studies have failed to detect an effect, this will tell us very little about the effects that have inevitably occurred. Excluding studies with small sample sizes should be considered formally. Notwithstanding these concerns, by formally combining the results across studies it can be feasible to increase power and address the problems associated with each individual study.

4.6.2 Requirements for informative meta-analysis

Harrison (2011) provides a ‘to do’ list which provides a useful starting point for collaborative approaches that will enable meta-analysis:

a. Ensure any study is informed by a thorough literature search.

b. Critically appraise studies to screen their suitability (applicable hypotheses, valid methods, sufficient information to enable calculation of effect size estimate with variance and record of effort).
c. Choose an appropriate metric and measure of effect size and calculate as effect size for each study.
d. Establish a master data base ensuring each study considers the same variables.
e. Use meta-analytic methods to reach a conclusion which ideally explains any variation between individual studies.
f. Assess robustness and power of analysis.

Harrison considers that the key to making meta-analysis as stress free as possible is organization and planning. This is particularly pertinent to organisation of data across studies to ensure variables can be formally compared. With respect to post-consent monitoring of ORE the current authors are not aware of any attempts to systematically establish, at a European scale, consistent approaches in relation to specific research questions that enable more robust meta-analysis techniques to be applied. Whilst this holds promise to support attempts to address key research questions with greater confidence it clearly requires pro-active collaboration between regulators and associated institutions in Member States.

In summary, where individual studies that are concerned with identifying a relationship between a cause and an effect size have produced inconsistent results, or results with a low level of confidence then meta-analysis is a well-established technique for providing a more generalised estimate. Pooling the results from even a relatively small number of studies has the potential to provide the necessary data for meta-analysis. In principle this provides stakeholders with a more robust understanding of the mechanisms and inter-actions that are of particular importance in understanding the risks associated with the impacts of ORE. Given that the manner in which predators use strong tidal areas is known to be highly variable and site specific, the approach could be used to disentangle some of the key patterns that drive behaviour, thereby allowing predicative modelling of potential effects such as collision risk to become more informative.

In conclusion meta-analysis has the potential to extract as much information as possible from a set of empirical studies. It may allow relatively powerful and robust results to be reached relative to alternative approaches. To be successful a meta-analysis requires a high degree of planning and organisation, identifying data to be gathered that are of the same variables at different study sites. Without consistent parameterisation data sets cannot be standardised in a manner that allows their collective analysis.

4.7 Discussion on approaches to statistical power and meta-analysis

Statistical significance and power, and use of meta-analysis have been reviewed and approaches to setting levels of significance and power have been considered.

In designing a programme for post-consent monitoring, a case has been made in earlier sections to adopt a question-led approach and to move towards study designs that increase the amount of information they can provide. In addition, consideration of the benefits that can accrue through varying the threshold for false positives from the value of 0.05 that is often chosen by convention. More priority should be given to agreement for false negative rates. It is recommended here that this is done on a case-by-case basis that is informed by levels of risk-appetite, noting that the value of 0.2 is often considered appropriate. Whatever values are chosen in practice it is recommended that decision making is informed by a two stage process. First, the consequences of each error occurring are evaluated in terms of costs to stakeholders and to biodiversity and the ratio between the rates for false positives and false negatives is set to reflect that. This allows the amount of data gathered to be agreed up front and to be as cost and time efficient as possible. Further savings may be achieved through adoption of iterative approaches that combine the use of power analysis with sequential setting of either false positive or false negative rates to further manage cost effective data collection.

Given monitoring activities for the impacts of ORE upon biodiversity are varied, and the way in which monitoring results provide useful information is also varied, differences in approaches are expected depending on specific circumstances. In principle given the variability of natural systems, the relative paucity of robust datasets on the mechanisms of effect and the associated costs, we might expect monitoring activities for ORE to use pragmatic effect sizes and associated levels of power and significance to allow outcomes that are measurable within the desired levels of confidence. In practice, when monitoring activities are undertaken to identify changes associated with ORE, we expect a case-by-case approach will allow the necessary flexibility according to perceived costs to stakeholders and biodiversity. However, we recommend the case-by-case specification should be arrived at through a question-led approach. Direct consultation with stakeholders at the onset of planning should be favoured to ensure that the close link between evidence and management be realised.

Meta-analysis provides further opportunity to provide meaningful results across multiple ORE sites in the context of adaptive management learning. However, in order to achieve this, the degree of planning and organisation would require considerably more active collaboration between regulators in Member States than is currently the case.

5. Conclusions

Monitoring potential impacts at MRE sites is likely to be extremely challenging given the relatively small spatial scale of current sites in combination with natural stochastic variation that will also inevitably influence how animals use and respond to the marine environment. There is a growing body of evidence showing that standard approaches to post-consent monitoring often result in DRIPy studies that are unable to meaningfully inform future decision making. There is a need for regulators to take stock of alternative risk-based approaches and the most suitable ways of achieving the goals of Adaptive Management. The focus should be on the role of government institutions and advisors to decision makers who should reconsider policies to promote an open and transparent approach to monitoring that is able to meaningfully reduce key scientific uncertainties and improve decision making. This needs to be done in association with careful consideration of the associated conservation objectives for features of biodiversity importance. Adaptive Management can provide an over-arching approach, within which the detail of implementation that is capable of providing good quality monitoring results should be planned on a case-by-case basis using some key guiding principles:

- Adopting a more open and transparent approach to accounting for scientific uncertainty within assessments to inform risk-appetite of decision makers.
- Adopting a question-led approach to post-consent monitoring that fully engages stakeholders.
- Ensuring the study design can meaningfully answer the question.
- Informing study design through careful selection of acceptable effect sizes, agreed rates for false positives and false negatives, and use of sequentially flexible approaches for minimising costs.
- Across Member States continued sharing of expertise and experience as good practice develops is recommended
- It is good practice to always present the mean effect size and associated variances
- Using demonstration projects in conjunction with meta-analysis techniques can help to more efficiently address shared priorities and would promote adoption of standardised approaches.

6. Acknowledgements

The authors acknowledge the advice and feedback of associates in the RiCORE project: David Gray, Andy Grinnall, and Ian Broadbent of Robert Gordon University, Teresa Simas of WavEC, Juan Bald of AZTI Tecnalia, Pierre Mascarenhas of E-Cube; colleagues at their respective institutions including Anne Marie O'Hagan, Cecile Le Lièvre and Mark Jessopp of University College Cork, Ian Davies, Rob Fryer Jared Wilson and Kate Brookes of Marine Scotland Science; Phil Gilmour and Paul Smith of Marine Scotland. Francesca Marubini is acknowledged for producing an upublished report on statistical power and significance that this report was informed by, and Charles Paxton gave permission for me to use his illustrations of study designs. Last but not least valuable advice on content was provided by attendees at RiCORE workshops and the wider community concerned with the contents including Ian Hutchison of Aquatera, Andrea Copping and Luke Hanna of Pacific North West Laboratory and George Lees of Scottish Natural Heritage.

7. References

Copping, A., Sather, N., Hanna, L., Whiting, J., Zydlewsk, G., Staines, G., Gill, A., Hutchison, I., O'Hagan, A.M., Simas, T., Bald, J., Sparling C., Wood, J., and Masden, E. 2016. Annex IV 2016 State of the Science Report: Environmental Effects of Marine Renewable Energy Development Around the World.

Culloch, R., Bennet, F., Bald, J., Menchaca, I., Jessopp, M. & Simas, T (2015). Report on potential emerging innovative monitoring approaches, identifying potential reductions in monitoring costs and evaluation of existing long-term datasets. Deliverable 4.3. RICORE Project. 61 pp.

Di Stefano, J. 2003. How much power is enough? Against the development of an arbitrary convention for statistical power calculations. *Functional Ecology* **17**: 707-709.

Dolan, T. E., *et al*. "Statistical power to detect change in a mangrove shoreline fish community adjacent to a nuclear power plant." *Environmental monitoring and assessment* 188.3 (2016): 1-16.

Dmitrienko, Offen, Westfall. (2003). Gatekeeping strategies for clinical trials that do not require all primary effects to be significant. *Statistics in Medicine*. 22, 2387-2400

DMP Statistical Solutions UK Ltd. 2010. Power analyses for the visual monitoring scheme at the Billia Croo site. Report prepared for SMRU Limited

Harrison, F., 2011. Getting started with meta-analysis. *Methods in Ecology and Evolution* **2**: 1-10.

Le Lièvre, C. and O'Hagan, A.M, (2016), Deliverable 2.3 Legal feasibility of implementing a risk-based approach. RICORE Project. 47 pp.

Lindeboom, Han, *et al*. "Offshore wind park monitoring programmes, lessons learned and recommendations for the future." *Hydrobiologia* 756.1 (2015): 169-180.

Long, C. 2016. Analysis of the possible displacement of bird and marine mammal species related to the installation and operation of marine energy conversion systems *Scottish Natural Heritage Commissioned Report No.* – Draft report due to be published 2016

Mackenzie, M.L., Scott-Hayward, L.A.S., Oedekoven, C.S., Skov, H., Humphreys, E., and Rexstad E. 2013. Statistical modelling of seabird and cetacean data: guidance document. University of St. Andrews contract for Marine Scotland.

Mapstone, B.D. 1995. Scalable decisions rules for environmental impact studies: effect size, Type I and Type II errors. *Ecological applications* **5**(2): 401-410.

Milieu Ltd. 2015. Draft Emerging Findings - Evaluation Study to support the Fitness Check of the Birds and Habitats Directives 72pp

MMO (2014). Review of post-consent offshore wind farm monitoring data associated with licence conditions. A report produced for the Marine Management Organisation, pp 194. MMO Project No: 1031. ISBN: 978-1-909452-24-4.

Peterman M. R., M'Gonigle M. 1992 Statistical power analysis and the precautionary principle. *Marine Pollution Bulletin* **24** (5): 231-234.

Scott-Hayward, L.A.S., Oedekoven, C.S., Mackenzie, M.L., Walker, C.G. and Rexstad E. 2013. User Guide for the MRSea Package: Statistical modelling of bird and cetacean distributions in offshore renewables development areas. University of St. Andrews contract for Marine Scotland; SB9 (CR/2012/05).

Vos P, Meelis E., Ter Keurs W. J. 2000. A framework for the design of ecological monitoring programs as a tool for environmental and nature management. *Environmental Monitoring and Assessment* **61**: 317-344.

Westfall, Krishen. (2001). Optimally weighted, fixed sequence, and gatekeeping multiple testing procedures. *Journal of Statistical Planning and Inference*. 99, 25-40

Williams, B. K., R. C. Szaro, and C. D. Shapiro. 2009. Adaptive Management: The U.S. Department of the Interior Technical Guide. Adaptive Management Working Group, U.S. Department of the Interior, Washington, DC.