Monitoring levels required in European Fisheries to assess cetacean bycatch, with particular reference to UK fisheries.

August 2003

Final Report to DEFRA (EWD)

Simon Northridge[1] and Len Thomas[2]

[1] Sea Mammal Research Unit, Gatty Marine Laboratory, School of Biology, University of St Andrews
[2] Centre for Research into Ecological and Environmental Modelling, The Observatory, School of Mathematics and Statistics, University of St Andrews

Table of Contents

Introduction

Article 12 of the European Union's Council Directive 92/43 of 21st May 1992, on the Conservation of Natural Habitats and of Wild Fauna and Flora (the 'Habitats Directive'), states that Member States "shall establish a system to monitor the incidental capture and killing of the animal species listed in Annex IV (a)". Annex IV includes all cetaceans.

The nature and scale of this monitoring is not specified, and perhaps as a result, little such monitoring has been achieved in the intervening 10 years (SEC 2002). In this paper we have tried to address the question of how much monitoring is needed to satisfy the Directive's intent.

Firstly, it must be recognised that monitoring schemes are intended to sample fisheries at some meaningful level, covering enough fishing to answer the management question posed without over-burdening the fishery, the monitoring authority or the Member State with unnecessary expense. In most circumstances a sufficient answer can be obtained by monitoring a small proportion of fishing operations and extrapolating to all the unobserved operations. Only very rarely would it be necessary to monitor a majority of fishing operations.

The Directive gives a little help in determining what management question the required monitoring might try to address. Article 12 states that: "in the light of the information gathered, Member States shall take further research or conservation measures as required to ensure that incidental capture and killing does not have a significant negative impact on the species concerned". The assumption must therefore be that monitoring should be carried out at a level that will enable Member States to identify situations where incidental catches in fishing operations may have 'a significant negative impact on the species concerned'. The Directive does not however give guidance on what such an impact might be.

There are currently at least two widely cited reference levels of impact on cetacean populations. One of these is the proposal first suggested by the Scientific Committee of the International Whaling Commission (the inter-governmental body whose Scientific Committee is widely recognised as the leading authority on cetacean population assessment and related issues) that any incidental capture that exceeds 1% of the estimate of total abundance for a particular species or stock, should be cause for concern and warrant further investigation (Anonymous 1996). If it is assumed that takes of less than 1% are not of concern, then this provides one starting point from which to examine how to allocate observer resources.

A second objective has been established under ASCOBANS (ASCOBANS 2000). The Meeting of Parties in Bristol in 2000 defined an unacceptable level of bycatch as one that exceeds 1.7% of the best available estimate of abundance. Clearly it is therefore also necessary to know with some degree of certainty if bycatch rates exceed this level.

With these two criteria in mind we examine first some previous attempts to address sample size requirements.

Previous attempts to address sample size requirements

Several previous studies have identified appropriate levels of sampling in order to estimate bycatch of cetaceans or marine mammals. Barlow (1989) provided guidelines for the implementation of the US Marine Mammal Protection Act to monitor fisheries with a known bycatch of marine mammals. Under this US legislation the level of coverage recommended to monitor marine mammal mortalities in fisheries with frequent incidental catches is between 20% and 35%. However, if insufficient funds are available a less intensive level of coverage may be considered if it can be shown to provide a statistically reliable estimate of marine mammal mortality.

In the case considered by Barlow existing estimates of marine mammal mortality were already available for several years for California gillnet fisheries. Barlow sought to determine how much sampling would be required in subsequent years. He used simulations in which he randomly assigned each simulated haul 0, 1, 2 or 3 bycaught mammals based on the observed catch rates. He also fixed the total number of operations in his simulation, but varied the proportion of hauls 'sampled' from 5% to 35% of the total number of operations. For each simulated sampling regime he computed the mean and variance for each of 500 estimates of mortality, and the CV (coefficient of variation: standard error of the estimate divided by the estimate itself; a standardised measure of statistical reliability). He produced a graph of the CV of the mortality estimates for harbour porpoises, harbour seals and sea lions in gillnet fisheries for each of 7 putative sampling regimes (from 5% to 35% of total effort). A CV of 10% was achieved at sampling rates of around 9% of total effort in the case of harbour seals and sea lions in set nets but around 35% in the case of harbour porpoises in set nets, where the total number of animals involved was much smaller. Barlow argued that a CV of 10% represents a very precise estimate, and that a CV of 15-20% might be acceptable.

Smith (1991) took this argument further. He noted that the management unit of interest is the ratio of the bycatch to the population size (B/P). Smith pointed out that the CV of the population estimate tends to be large (typically 20-30%). Improving the precision of the bycatch estimate (that is, decreasing the CV of the bycatch estimate) has an increasingly small effect on the overall CV of the B/P ratio. Thus the law of diminishing returns takes effect once the CV of the bycatch estimate goes below that of the population estimate.

In the North Sea, for example, the CV of the population estimate for porpoises is around 14% (Hammond *et al*. 2002). It would therefore make sense under this sampling objective to allocate sampling effort in relevant North Sea fisheries to achieve estimates of mortality with a CV of approximately this same level.

Computing the CV of a mortality estimate is of course only possible where previous data allow one to establish the statistical distribution of bycatch events. In Barlow's example this was done empirically by using the observed proportions of hauls with, 0,1,2,3 or more animals. In many cases there are no pre-existing data and this is not possible and the allocation of sampling effort is then problematic.

Furthermore, Barlow also pointed out that whereas obtaining a precise estimate of mortality may be a useful goal in some circumstances, in others it may not be biologically meaningful. In the case where an animal is very abundant and its bycatch rate is known to be small, there is little point in over-sampling a fishery simply to obtain a precise estimate of a number that is too small to be of concern.

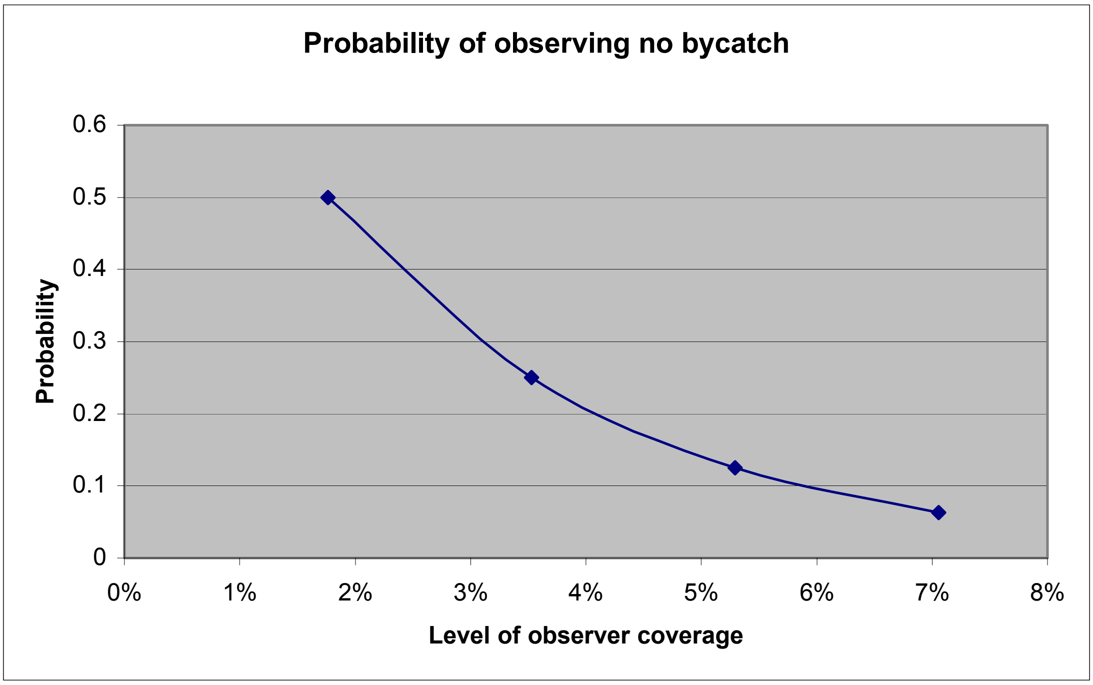
As an example Barlow suggested that if around 10 common dolphins were taken per year from a population of 100,000 a CV of perhaps 100% (corresponding to an upper confidence limit on the mortality estimate of 30) would be sufficient for management purposes.

Whereas targeting a particular CV for the estimate of bycatch mortality may therefore be one goal, it is not always the most appropriate one. Wade (1999) pointed out that for a fishery that has not previously been observed (or by extension one for which there is no reliable estimate of bycatch rate), it is more appropriate to ensure that sampling levels are sufficient to make the probability of observing zero takes (bycatches) very small when the true number of takes is high enough to be of concern. Wade argued that any first-time observer programme should be designed so that the probability of observing zero takes is very small if the real take level approaches anything that might be of management concern. In the European context this would mean any takes of more than 1% or perhaps 1.7% of the abundance estimate.

The way that Wade did this was to assume that bycatch events follow a binomial distribution, that is that each operation has an equal and independent chance of taking one mammal. He then went on to apply this to a particular (unnamed) scenario where there were an estimated 5668 days at sea and a bycatch 'limit' of interest of 39 mammals. If the fishery was responsible for as many as 39 mortalities in this case the bycatch rate would have been 0.0069 animas per day at sea. It is then possible to compute what would be the probability of observing at least 1 bycaught animal (or take) for a given level of observer coverage, and indeed what the probability of observing no animals would be, given the proposed take rate. If the sampling coverage is set to ensure that at least one animal is observed under a proposed mortality rate and none is then observed, it can be concluded with some degree of certainty that the real take rate is smaller than the proposed rate. Wade's results are shown in Figure 1, and the methods are described in more detail in the Technical Annex to this report.

Clearly the probability of observing one or more animals is the inverse of these probabilities and for this particular fishery, sampling at a rate of 7% of observer effort would provide a 94% probability of observing at least one the 39 animal assumed to be taken. As sampling levels increase so it becomes increasingly less likely that no takes will be observed if the real take rate is as much as 39 animals per 5668 days at sea. It is simple to change the numbers for total effort and presumed bycatch rate to others and do these same calculations for other specific fisheries (see Technical Annex, Section A for details).

Figure 1: Probability of observing no bycatch in a fishery with an assumed bycatch rate of 0.0069 mammals per day and 5668 days at sea in total (From Wade 1999) and with an assumed binomial distribution of bycatch



Such an approach works only where the probability of a bycatch follows the binomial distribution. If bycatch events are clumped then this may not be the most appropriate distribution to use, but in cases like that of harbour porpoise bycatch in gillnet fisheries, where catches of single animals are much the most frequent events, and where spatial and temporal clumping have been hard to demonstrate, this is probably a reasonable assumption upon which to base an initial sampling strategy.

Application To The EU Situation

Within the EU two broad types of fishery have been identified as being of interest in the context of cetacean bycatch, these being pelagic trawl fisheries and gillnet or setnet fisheries. Among the various UK pelagic trawl fisheries several have been sampled with no observed cetacean bycatch, while one has been found to have a fairly high rate of cetacean bycatch. Among the UK gillnet/setnet fisheries differing bycatch rates have been recorded among different sectors. All recorded bycatches in gillnets have so far been of harbour porpoises, with the exception of those in the Celtic Sea where some common and white-sided dolphins have been reported taken in set nets. All recorded bycatches in UK pelagic trawl fisheries have so far been identified as common dolphins.

Summary information on recent levels of fishing effort and observed bycatch levels are given in Table 1 for UK pelagic trawls and Table 2 for UK gillnet fisheries.

Table 1: UK Pelagic Trawl Fisheries.
Fishing Effort (2002) and Bycatch Observations

Fishery	Days at sea 2002	Average no of operations (tows) per day[1]	Estimated no of operations in 2002	No of observed operations[2]	No of observed cetacean takes	Estimated bycatch rate (cetaceans per tow)
Mackerel	3477	0.4	1450	34	0	0
Herring	1583	0.9	1367	44	0	0
Sprat	517	1.4	747	14	0	0
Horse mackerel	218	0.9	189	0	0	0
Bass	176	2.1	363	223	89	0.399
Pilchard	176	1.1	188	9	0	0
Others	458	1.8	814	26[3]	0	0

[1] Estimated from those logbook/landings records where number of tows were given
[2] Observed during the period 1999-2003; excludes tows with selection grids in use
[3] Includes tows for anchovies, blue whiting, smelt and sprat

Table 2: UK Gillnet Fisheries.
Fishing effort (2002) and Bycatch Observations

Target (Species of greatest value in landings)	Days at sea 2002	Hauls per day (as observed)	Assumed hauls in 2002	Observed Hauls (1996-2000)	Observed porpoises bycaught	Estimated bycatch rate (porpoises per haul)
Sole	8672	4.5	38742	373	4	0.0107
Crabs and Lobster	4515	5.4	242445	0	0	
Cod	3978	6.5	25802	3066	20	0.0065
Monkfish	3416	5.4	18343	133	0	0.0000
Other gadoids	3088	5.4	16583	0	0	
Bass	2506	3.5	8828	54	0	0.0000
Plaice	2095	5.4	11250	0	0	
Hake	1665	5.4	8941	0	0	
Dogfish	1553	4.1	6382	353	8	0.0227
Turbot	1152	5.8	6708	142	4	0.0282
Salmonids	1068	2.5	2652	149	0	0.0000
Shark	902	5.4	4843	0	0	
Skate	887	4.8	4258	298	15	0.0503
Cuttlefish	691	5.4	3710	0	0	
Deepwater spp.	553	5.4	2969	0	0	
Monkfish	531	3.4	1821	0	0	
Other	430	5.4	2309	0	0	
Herring	423	2.4	1022	16	0	0.0000
Mackerel	301	5.4	1616	0	0	
Crayfish	243	4.0	976	97	3	0.0309
Sardine	182	5.4	977	0	0	
Mullet	174	2.5	435	5	0	0.0000
Sprat	163	5.4	875	0	0	
Shellfish	111	5.4	596	0	0	
Other flatfish	45	5.4	242	0	0	
Cephalopod	17	5.4	91	0	0	
All Species	39361	5.4	195217	4686	54	0.0115

The two types of fishery and the two species have very different bycatch distributions. Whereas in most cases porpoises are recorded as single animals, common dolphins taken in pelagic trawl fisheries are usually taken in groups. This introduces an added level of complexity into the trawl situation as one needs to know or estimate not only the probability of a trawl operation catching one or more animals but also the expected number of animals taken in such an operation.

Among the gillnet fisheries, almost all that have been monitored have had a positive bycatch rate. The number of porpoises taken in these fisheries ranges from 1 in 20 hauls (skate fishery) to 1 in 154 hauls (cod fishery) with an overall mean among the sample of 1 in 87 hauls. Sampling in the mullet, herring and bass gillnet fisheries has been below the number of hauls expected to see a single animal, but in the monkfish and salmon fisheries there has been sufficient sampling to suggest that absolute catch rates are lower than the mean for all fisheries. Much of the difference in bycatch per haul may be explained by differences in soak time and net lengths (short on both counts in the salmon fishery for example) but it is not practicable to try to estimate the soak times and net lengths for the entire fleet, so here we retain catch per haul as the basic metric. Porpoises are generally caught as single animals, with an occasional capture of 2 animals in the same net.

For pelagic trawl fisheries, most have yielded no observations of bycatch, whereas in the pelagic trawl fishery for bass, observed bycatches have been fairly high (1 operation in 7 has a dolphin bycatch of 1 or more animals). For the bass fishery, which has already been sampled at around 30% of fishing effort over the past 3 years, mitigation measures are about to be implemented which, it is hoped, will reduce this bycatch rate by an order of magnitude. The sampling requirements for the bass fishery are therefore likely to differ from those of the remaining pelagic trawl fisheries. Dolphins are generally taken in groups, with a mean of around 4 animals for the bass trawl fishery.

It therefore makes sense to consider at least three fishery categories. The first category covers gillnet fisheries, for which we might wish to ensure that sampling is sufficient to be sure that less than 1%, or less than 1.7%, of the estimate of porpoise abundance <u>for a specific area</u>, is being taken per year. The second category involves most pelagic trawl fisheries, where the current assumption based on sampling is that catch rates are low, and where we may wish to ensure that bycatches are no more than 1% of the abundance estimate. The third category is the bass fishery, where we may also wish to ensure that catch rates are consistent both with some upper level that might be regarded as a maximum allowable take from a conservation perspective, and a lower level that represents the expected bycatch after mitigation measures are in place.

In all of these cases the essential question involves determining some upper confidence value of an estimate of total bycatch, and determining how much observer effort is likely to be required to achieve such an estimate. There are immediately several further questions that need to be addressed. These include the confidence level that is required, and indeed the level of certainty of the estimate of that level: estimates of the bycatch rate and associated confidence intervals will vary from one sample to another taken from the same overall population, and it is therefore necessary to specify how often the sampling regime should yield a given confidence interval. Above all, one also needs to determine the nature of the statistical distribution from which bycatch records are drawn.

The appropriate statistical distributions

The nature of the statistical distribution of bycatch events fundamentally affects all the answers about levels of certainty and estimates of confidence intervals. It is therefore important to be sure that an appropriate distribution is used. Using the existing observer data from the bass trawl fishery and those from the gillnet fisheries we fitted several different distributions to the data by numerically maximising the log-likelihood. The details of this process are given in the Technical Annex, but we found that the best fit to the dolphin/trawl data was obtained when using a Zero-Inflated Geometric (ZIG) distribution, while a simple binomial model was considered best for porpoise gillnet data (see Technical Annex, section B3)

We also examined dolphin bycatch data from observations made on board Irish pair trawl vessels fishing for tuna in the summer months (data kindly provided by Dominic Rihan, BIM). Here the ZIG distribution provided the best fit.

While the binomial model fitted the existing porpoise data well, both Zero Inflated Poisson and ZIG distributions also fitted the data well, raising the possibility that the ZIG distribution might be used as a general one for cetacean bycatch data. The Poisson distribution did not fit the observed data well (See Technical Annex section B3). For this exercise we used the binomial distribution rather than the ZIP or ZIG distributions, because it enables an easy calculation of the confidence limits, and one that provides a non-zero confidence interval for sets of observations without any bycatch (see also Technical Annex Section C). This is despite the fact that the data violate the assumption of a binomial model that the maximum number of animals taken per operation is one, as the difference made by a few hauls with two animals is very small.

The important, and convenient features of both the binomial and ZIG distributions are that they are described by just two (ZIG) or one (Binomial) parameters. In the case of the ZIG distribution these parameters are p the probability of observing zero bycatches and q, which is the inverse of the mean group size in a capture event. For the binomial, where the expected group size is one animal, the distribution is defined by the single parameter p.

Further sampling, and a larger dataset, might lead to better fits of alternate models, or may suggest that the ZIG distribution should be used for porpoise bycatch data too, but for the present we will assume that the porpoise bycatch data can be modelled as a binomial process and that the dolphin bycatch data can be modelled as a ZIG process.

The parameter estimates (with standard errors) for the three datasets that we used to specify these distributions are as follows:

Table 3: Bycatch parameter estimates

Fishery	Distribution	Parameter estimates	
Bass pair trawl	**ZIG**	**P=0.91443 (0.02045)**	**Q=0.253960**
Irish tuna pair trawl	**ZIG**	**P=0.930877 (0.0111)**	**Q=0.22361**
Gillnet / porpoise	**Binomial**	**P=0.98834**	

Once we have distributions that fit the observed data we can also compute CVs and confidence intervals for bycatch rate estimates.

The 'target' bycatch levels

In order to address sampling needs we also need to define some critical level of mortality that we wish to be 'sure' is not exceeded by the fishery (ignoring for now what we mean by 'sure'). This level, as we have seen, can be taken in the short term to be 1% or 1.7% of abundance estimates. The 1% level reflects a level above which we should be concerned, whereas the 1.7% limit represents an absolute level that must not be exceeded.

The fisheries and cetaceans concerned are distributed in areas that are shared with other nation's fishing vessels, and there is therefore the problem of how we define what a UK take limit should be in the context of multi-national fisheries. One strategy, adopted by the UK's Small Cetacean Bycatch Response Strategy, is to allocate the take limit among the nations involved in proportion to each of their total catches or fishing effort. A second problem is that we do not have abundance estimates for most cetacean species in the Irish Sea or to the West of Scotland, and the only existing abundance estimates for other parts of the UK's waters are now nearly ten years old.

With these caveats in mind we can address each of the three fishery types in turn.

<u>Bass Fishery</u>
The bass fishery is reported to take common dolphins. Our observers have identified no other species, and skippers and crews also maintain all the animals taken appear to be the same species.

If we assume that the estimate of 75,500 common dolphins made by the SCANS survey (Hammond et al. 2002) for the Celtic Sea in July 1994 represents the number of animals present in ICES Divisions VIIefgh (the Celtic Sea and Western Channel) during the winter months at present, then a total take of more than 755 (1%) would be a cause for concern, according to the IWC criteria. We also know that there may be an annual take of 200 common dolphins in the UK and Irish hake gillnet fishery. Under this assumption a take of more than 555 dolphins by the pelagic pair trawl trawl fisheries would be a cause for concern.

The UK pair trawl fishery for bass in ICES divisions VIIefgh involves some 8 boats (4 pairs). In 2001 they landed 94 tonnes of bass. The French fleet operating in the same area involves some 60 boats (30 pairs) and in 2001 they landed over 353 tonnes of bass. So even though the UK fishery is only 14% of the French fleet by number of boats, they land 27% of the total declared landings in France. At present the bass fishery is the only pelagic trawl fishery known to have a significant bycatch in this area. If we allocate the UK sector of the fleet 14% of the 1% limit of 555 animals, this would result in an allocation of 77 animals, and we can use this as an 'arbitrary' figure for an upper catch limit to frame calculations of observation requirements. At the same time, mitigation measures are being adopted by this fishery, so that it may be anticipated that bycatch rates in future years will be very much lower. We might therefore take another arbitrary figure of 10 animals per year as a target not to be exceeded in the future (note that this figure has no rational basis and is only proposed here to illustrate subsequent calculations).

Other Pelagic Trawl Fisheries

For other pelagic trawl fisheries the target levels will depend on the area under consideration. For pelagic trawl fisheries in the North Sea, the most vulnerable cetacean population is probably the white-beaked dolphin, with an estimated abundance of just 7800 in the North Sea, yielding a 1% limit of 78 animals per year. The proportion of herring, mackerel, sprats and other species taken by UK vessels among those of other nations in the North Sea could not be ascertained, so we currently have no rational basis for allocating a UK bycatch take limit for this species in the North Sea.

Bottlenose dolphins are also present in this area in even smaller numbers, but may not overlap in their distribution much with the pelagic trawl fleet. With a population size of around 130 animals (Wilson *et al* 1999), clearly a 1% limit would mean that no more than 1 animal a year could safely be taken by any fishery. At such low population levels fishery monitoring is probably not the best way to disburse scarce conservation resources, as the required coverage levels would be close to 100% of the fisheries concerned. Direct population monitoring may be a more efficient means of monitoring population level impacts.

On the western or Atlantic coasts of the Scotland, the primary delphinid species are whitebeaked dolphins over shelf waters and Atlantic white-sided dolphins primarily in deeper Atlantic waters. There are no abundance estimates for either of these species in this area at present, making it difficult to provide a rational take limit for this area.

In Area VII to the south and west of Ireland and England, Atlantic white-sided dolphins, striped and common dolphins predominate. There are estimates (as previously mentioned) of 75,000 common dolphins for the Celtic Sea and a further 62,000 from the 1993 MICA project for an area further offshore than but slightly overlapping with the SCANS survey area, as well as an estimate of 72,000 striped dolphins for the same area (Goujon et al. 1993). Goujon (1996) provided a joint estimate of 120,000 common dolphins for the Celtic Sea and adjacent offshore areas for common dolphins.

Again assuming limits of 1% of these abundance estimates suggests bycatch levels of concern of 1200 common dolphins and 720 striped dolphins, but there is no information on white-sided dolphin abundance in these areas. The species with the lowest abundance estimate will determine the take limit and so define the sampling regime requirements in any observer scheme.

Given the lack of any obvious way of defining take limits for most of these species and area combinations, we have resorted to the expedient measure below of assuming a take limit of 100 dolphins (of any or all species) for each of the five primary pelagic trawl fisheries listed in Table 1 (the bass fishery already having been considered above), for the present at least, regardless of where they are operating. Again this take limit figure is used simply to illustrate the sample size calculations.

Gillnet Fisheries

We assume that the species of concern is the harbour porpoise (ignoring for the present the known take of dolphins in gill/tangle net fisheries in the Southwest). In the North Sea the abundance estimate produced by SCANS was 268,000 animals (Hammond, Berggren et al. 2002), producing a 1.7% limit of 4556 and a 1% limit of 2680. Assuming that the UK is responsible for around 15% of gillnet catches in the North Sea, UK bycatch limits are 683 and 402 respectively. For the Celtic Sea the abundance estimate from SCANS was 36,000, producing nominal catch limits of 612 and 360 at the 1.7% and 1% levels respectively. On the assumption that the UK is responsible for 30% of the gillnet catches in this area, proposed UK bycatch limits are 184 and 108 animals. For the Irish Sea and the West of Scotland there are no abundance estimates, so we adopt values of 510 and 300 (the 1.7% and 1% levels respectively) for the West of Scotland, and 153 and 90 for the Irish Sea. These figures are arbitrary values.

A procedure for estimating sampling requirement

Once we have determined the appropriate statistical distribution for catches, it is possible to simulate a sampling regime drawing from that distribution as defined by the parameters *p* and *q* for the ZIG distribution, or in the case of the binomial just *p*. However, this presupposes that we already know what the value of *p* is – or that we know the proportion of operations that will/will not catch cetaceans. In fact in most cases we do not know this, and what we therefore need to do is to postulate a range of possible values, based on experience. The simplest way to do this is to propose a range of possible total bycatch values for the fishery in question. We can then simulate an observer programme whereby the total number of fishing operations is pre-determined, but both the proportion of operations sampled (H_o) and the total kill (K) are allowed to vary in steps. Thus, given a value for *q*, for any given value of K it is possible to determine the expected number of operations resulting in 0,1,2,3,4 …animals being taken. For each combination of H_o and K we then sample the appropriate expected hauls 1000 times and compute the 90% one-sided upper confidence limit (UCL) on each of the resulting 1000 estimates of total bycatch for each H_o and K value.

This calculation is done for the ZIG distribution assuming that the mean bycatch estimate is lognormally distributed, and therefore using the lognormal one-sided UCL. This means that when we observe no bycatches, the UCL is zero, but there is no simple alternative to this. In the case of the binomial distribution we can do better by calculating CLs using an approximation based on the F-distribution (Zar 1996, p524) that provides us with non-negative UCLs even when no bycatches are observed (see also the Technical Annex Section C).

When we have calculated 1000 UCLs for each combination of Ho and K we can then present the results to show the proportion of these UCLs that exceed the target bycatch level (as defined above), or the level of certainty that we have for the UCL estimates. Given a graph of the results it is possible, for any proposed value of Ho, to see how sure one can be that the resulting upper confidence limit on the bycatch estimate will exceed the target level for the full range of plausible catch rates.

There are several assumptions here that need to be made explicit. First we assume that the statistical distribution and the parameter estimates adequately describe the bycatch process. Second we assume that we can guess what a likely range of total bycatch values might be. Third we assume that the number of fishing operations can be predicted. Fourth we assume that it is the upper 90% confidence limit on the bycatch estimate that is the appropriate one to compare to our reference level, and fifth we have to adopt some level of certainty that our calculated confidence interval is correct. We have adopted a certainty value of 90%. Of course the bycatch limit also needs to be specified, and as shown above, in several cases we can do no better than guess what this should be.

Examples drawn from UK fisheries

<u>The bass fishery</u>

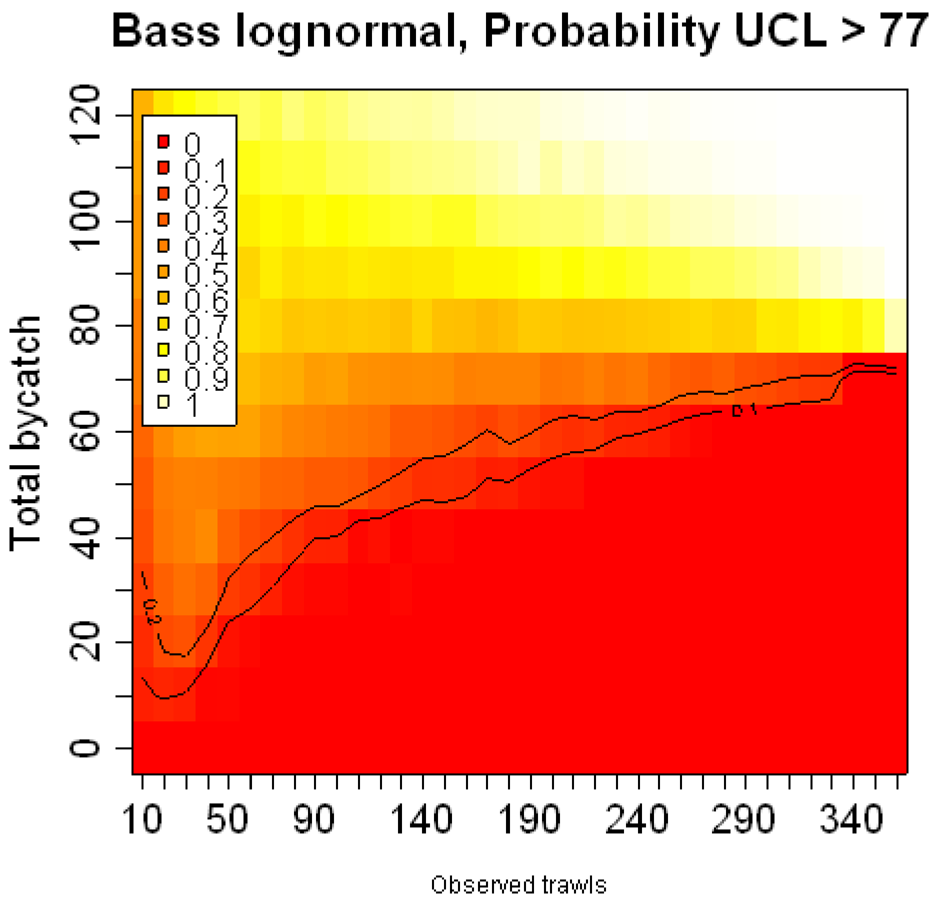
The bass fishery has been sampled fairly intensively over the past three years with coverage of around 30% of effort. The official statistics suggest that the total number of hauls in 2002 was 363. We assume that effort will continue at the same level in future years. It is expected that the bycatch rate in this fishery will fall considerably after the introduction of mitigation measures in the coming season. If the bycatch mitigation measures are successful or largely successful, we might further expect that the q parameter in the ZIG distribution, the inverse of the number of animals taken per haul, might increase from 0.24 to 1, in which case we would be looking at a binomial distribution of catches (each haul either has one animal or none) and we could therefore also model the bycatch under these circumstances as a binomial process.

We might therefore wish to know what levels of coverage will be needed either to ensure that bycatch levels do not exceed 78 animals, or that bycatch levels are less than the arbitrary value of 10 animals. We can therefore run the simulations with coverage levels of 10 to 360 hauls in steps of 10 (that is from 2.7% to 99% coverage), under assumptions that the total bycatch might be between 0 and 120 to determine what levels of coverage would be consistent with us obtaining a bycatch estimate with an upper confidence limit that is less than either of our two limits.

The results of this trial are shown in Figures 2 through 5. Implementation details are also given in the Technical Annex.

Figure 2 suggests that if the real bycatch were to be, for example, about 40 animals, and our limit is 77 animals we could be 90% sure that the 90% upper confidence limit on an estimate of bycatch would be less than 77 if we sampled about 90 tows. If the true bycatch was 10, then we need only sample around 30 tows to be 90% sure of the 90% confidence limit being less than 77 animals. If, conversely, the true take is more than our limit of 77, and we wanted to be sure that it really was greater, then the closer the true total is to 77 the more sampling we will need to do to ensure that our upper confidence limit is greater than 77 animals. If the true total were, for example, 120 animals, 90% of our UCLs would be expected to exceed 77 if we sampled about 180 tows (the 0.9 line is not drawn in this figure but is inferred from the colours). If the true total were to be around 80 animals, however, we would need to sample almost all of the fishery to ensure that our confidence limits was more than the take limit of 77.

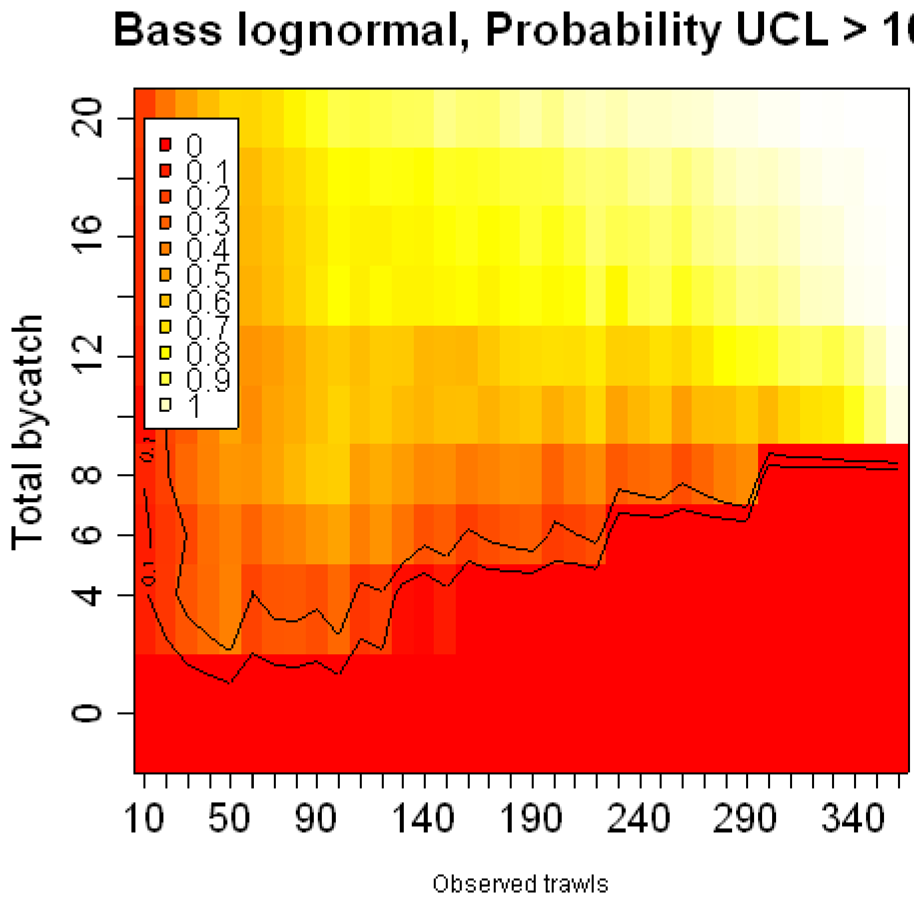
Figure 2: Bass fishery simulation: take limit set at 77 animals. ZIG distribution of bycatches. Lognormal one-sided upper confidence limit calculated for 1000 iterations at each step, 10-360 hauls, actual bycatches of 0 to 120 animals. A q value of 0.24 was used.



A noteworthy feature of Figure 2 is that for very low levels of coverage confidence intervals become narrower. This is because as one samples fewer and fewer tows the probability of seeing no animals at all increases, and so a take rate of zero with confidence limits of zero in this case also ensue. Wherever lognormal confidence intervals are being used (or where catches are demonstrably non-binomial and frequent) it is therefore important to try to minimise the chance of observing no takes.

In the bass fishery we have good estimates of the true bycatch rate. If the past season's performance with mitigation measures can be used to extrapolate forward to next season, we might expect only around 8-10 animals to be taken next season. If we expected around 10 animals and wanted to be sure that the 77 limit is not exceeded, then 30 tows would be sufficient. However, if we set a limit of just 10 animals the sampling requirements change considerably.

Figure 3: Bass fishery with a take limit of 10 animals, lognormal UCL; q=0.24



In this case if the true total bycatch were 8 animals we would need to sample the entire fishery to be 90% sure that our upper confidence interval is less than the limit of 10 animals. At low sampling levels (closer to 50 tows) the probability of seeing no catches increases dramatically, which greatly affects the lognormal UCL calculations.

We might also consider that at such low catch rates bycatches might occur with a *q* value closely approaching 1 rather than the current 0.25, so that bycatch events consists of just one animal, which would allow us to use a binomial distribution. We simulated the same sampling regime and total kill rates using a binomial distribution and the results are shown in Figure 4.

Comparing Figures 3 and 4, although both indicate that the entire fishery would need to be sampled if total catches really are close to the take limit of 10, the binomial model also indicates that for lower total bycatches of 2 to 6 animals higher sampling rates are required to be sure of obtaining UCLs of less than the take limit if 10 than are predicted by the ZIG distribution with log-normal confidence intervals. This is because of the occurrence of significant numbers of samples where the recorded bycatch is zero, biasing the UCL downwards in the ZIG model. The conclusion here is that although it may be possible to tell that a mitigation measure is reducing bycatch rates (this is something we didn't examine explicitly here), it will be very difficult to establish that the true bycatch is less than 10, unless it is substantially less than 10 or we do a lot of sampling.

Figure 4. Bass fishery with a take limit of 10, Binomial distribution

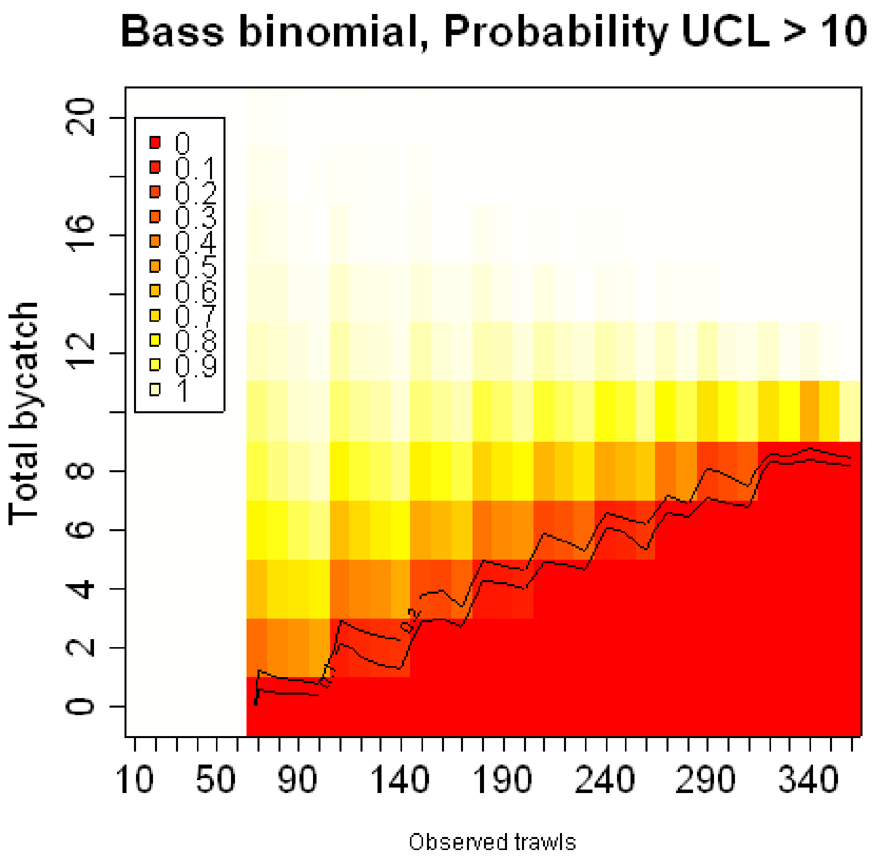
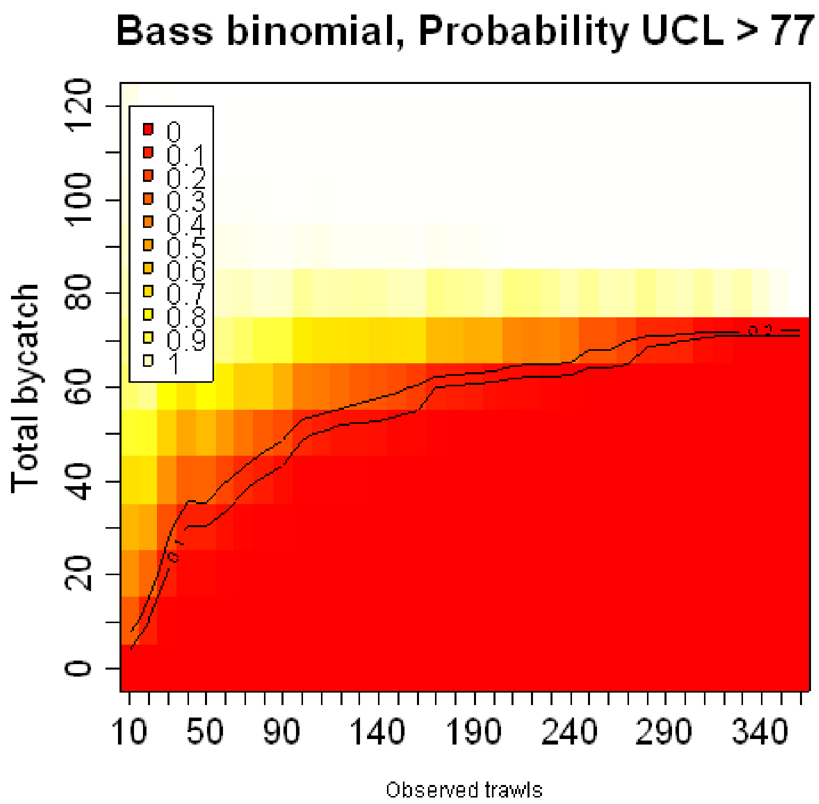


Figure 5. Bass fishery with a take limit of 77, Binomial Distribution



For completeness we can also examine the case where we might assume a binomial distribution of catches and where we set a take limit of 77 animals. The results of these simulations are shown in Figure 5.

As with the lognormal confidence intervals, this suggests that if the true bycatch rate is around 10 animals, then sampling about 20 hauls should be enough to obtain a UCL that is less than 77. This of course assumes animals are only taken singly.

The bass fishery is unusual because we have a lot of information about it and we can define likely future bycatch rates and set a take limit (albeit somewhat arbitrarily). This has allowed us to explore various aspects of the overall question in more detail than may be possible elsewhere. Two important general points emerge from this. Firstly, where the absolute kill rate is low and the take limit is low, then a large proportion of the fishery needs to be sampled.

For example, for bottlenose dolphin populations that typically number in the tens or low hundreds, take limits and real bycatch rates are almost always likely to be in single figures, which makes monitoring fisheries an impractical option as it will require most fishing operations to be monitored in order just to determine whether the true take rate is more or less than the take limit.

The second and more general point that follows is that sampling requirements are largely defined by the difference between the real take rate and the take limit. If the true take rate is low compared to the take limit then relatively limited sampling can show that the true catch rate is likely to be lower than the take limit. As the true take rate approaches the take limit, so the amount of sampling required to be sure that the take rate is less than the limit increases towards 100%. Once the true catch rate exceeds the limit, the amount of sampling required to be sure that it exceeds the limit declines again and becomes very small as the catch rate becomes much larger than the take limit.

From this it is clear that the single parameter which we can control and which has the greatest impact on these calculations is the take limit. Setting a high take limit relative to the expected total catch will lower sampling requirements, while setting a low take limit will increase the amount of sampling required. Unfortunately, take limits are hard to define in the absence of abundance estimates.

Other parameters that we can set, but which may have a lesser influence will be the level of the confidence limit of our estimate of bycatch (here we have used 90%), and the certainty with which we wish to attain this level (again we use 90%). Lower values of either of these parameters would result in lower sampling requirements. By choosing to use 1-sided confidence limits we have also lowered sample size requirements without losing information as the lower confidence limits are irrelevant to our fundamental questions.

<u>Other pelagic trawl fisheries.</u>

For the pelagic trawl fisheries in general, we have arbitrarily set a take limit of 100 animals for each of the five fisheries (excluding bass for which we have different bycatch parameter estimates). We have used the predicted number of tows for 2002 from Table 1 to provide estimates of total effort in future years. We do not have any good estimate of the total takes in these fisheries, but we have observed 114 hauls with 0 bycatches recorded among all of these other fisheries, providing us with an estimate for the parameter *p* (proportion of hauls with zero bycatches) of 0. Confidence limits can also be calculated using a standard method based on the *F* distribution (see Technical Annex), which yields a one-sided 90% UCL for *p* based on these observations of 0.9741 (this is the same method used in the binomial simulations for the bass fishery above). We can therefore say that based on what we have already observed, we are 90% sure that 97.4% of all hauls in other pelagic fisheries as a group will have zero bycatches.

We have no information on the likely number of animals taken in any non-zero tows (1/q). For this we use the 'worst case' scenario data from the Irish pelagic trawl fishery for albacore, where q= 0.22361, equivalent to the assumption that he expected (mean) take in a non-zero bycatch haul is 4.47 animals.

Based on these assumptions about the maximum plausible bycatch rates in the other pelagic trawl fisheries we can then calculate maximum possible takes in each of the five fisheries in question. These are listed below in Table 4. The calculations are also detailed in the Technical Annex.

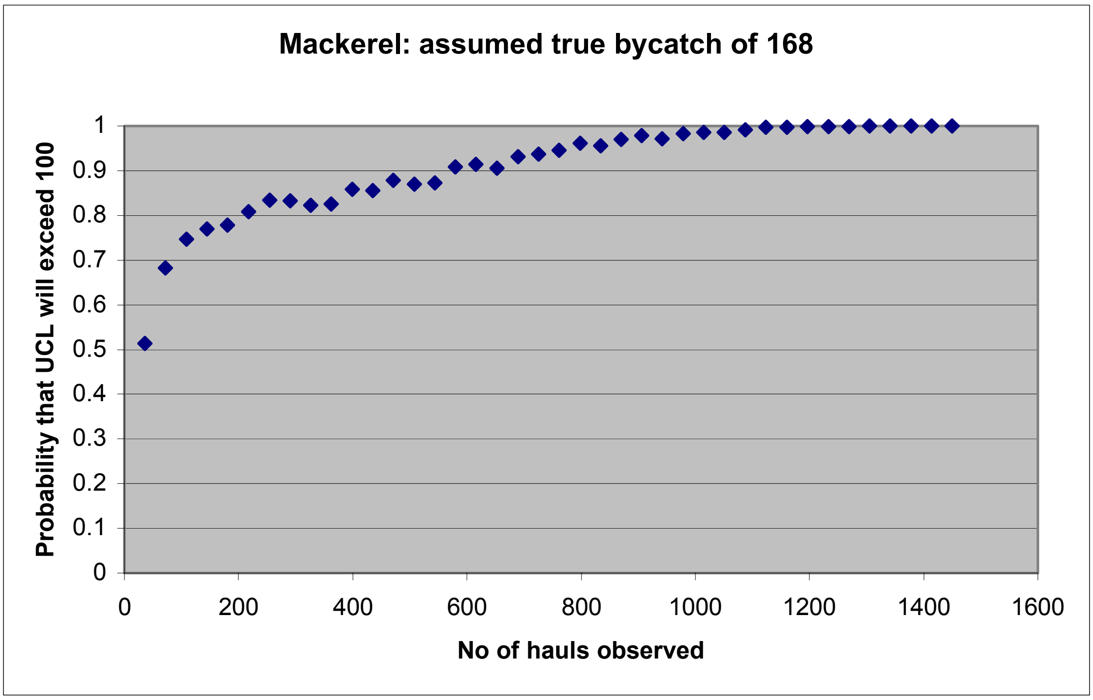
Table 4: Predicted effort levels in other pelagic fisheries and maximum likely dolphin takes based on estimated p and q values from observations in these fisheries and the Irish albacore trawl fishery.

Fishery	Estimated effort (tows)	Maximum likely bycatch
Mackerel	1450	168
Herring	1367	158
Sprat	747	87
Horse mackerel	189	22
Pilchard	188	22

Using these values as predicted true kill rates and a universal take limit of 100 dolphins per fishery, we have then run 1000 simulated observation schemes for each of 40 different levels of observer coverage ranging from 2.5% to 100%. The results are expressed in terms of the probability of the 90% UCL for the resulting bycatch estimate being more than 100 animals. The results are shown below in Figures 6-10. Note that in two cases the postulated total bycatch exceeds the arbitrary take limit, and in these cases we determine how much sampling we need to be sure that the take exceeds the limit, rather than how much sampling we need to do to be sure it doesn't exceed the limit.

In the mackerel fishery the amount of fishing effort is enough to ensure that at the postulated bycatch rate the total bycatch will exceed the arbitrary take limit of 100 animals. Because the assumed true rate is close to the take limit, it is necessary to sample about 770 tows or 53% of all fishing operations to be 90% sure that the estimated take rate really is above the 100 animal limit. However, one could be about 70% sure of this after sampling only about 80 hauls or 5.5% of total annual effort.

Figure 6: Mackerel trawl fishery - assumed true bycatch rate exceeds the take limit by 68 animals.



Similarly in the herring fishery where the highest plausible take rate also exceeds but is close to the arbitrary take limit, high levels of sampling would be required to be sure that the estimated catch really is above that limit. In this case only around 650 hauls would be required to be 90% sure of obtaining an UCL in excess of the take limit. We could be 70% sure that the limit was exceeded with about 70 observed tows.

Figure 7: Herring trawl fishery - assumed true bycatch rates exceeds the take limit by 58 animals

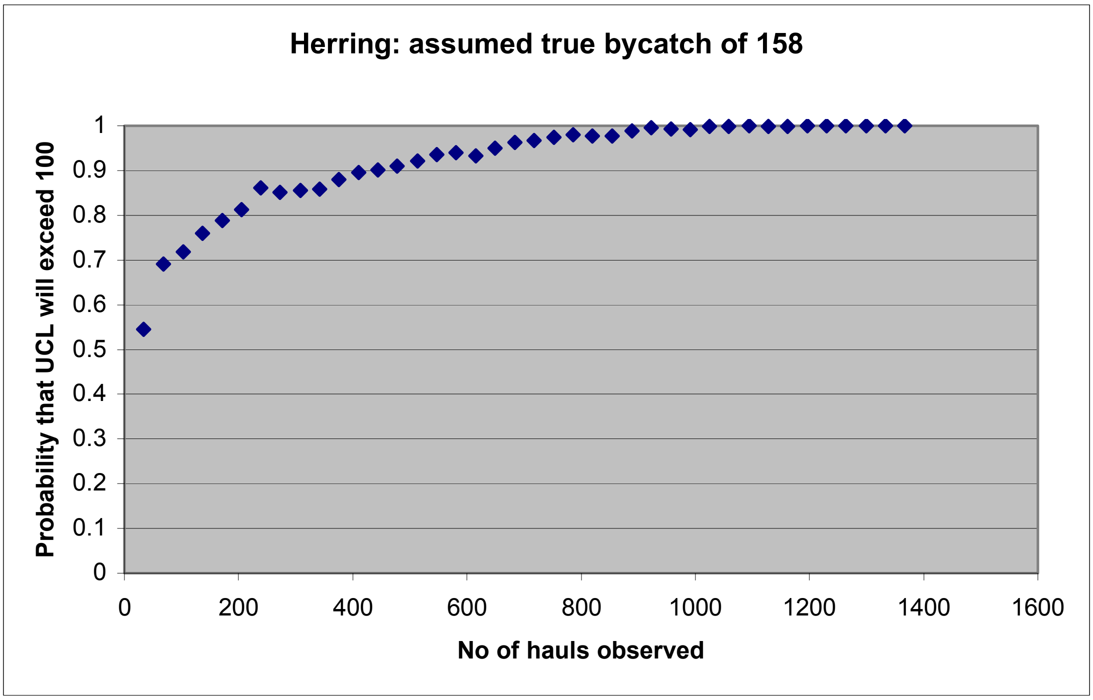


Figure 8: Sprat trawl fishery: assumed bycatch rates less than take limit by 13 animals

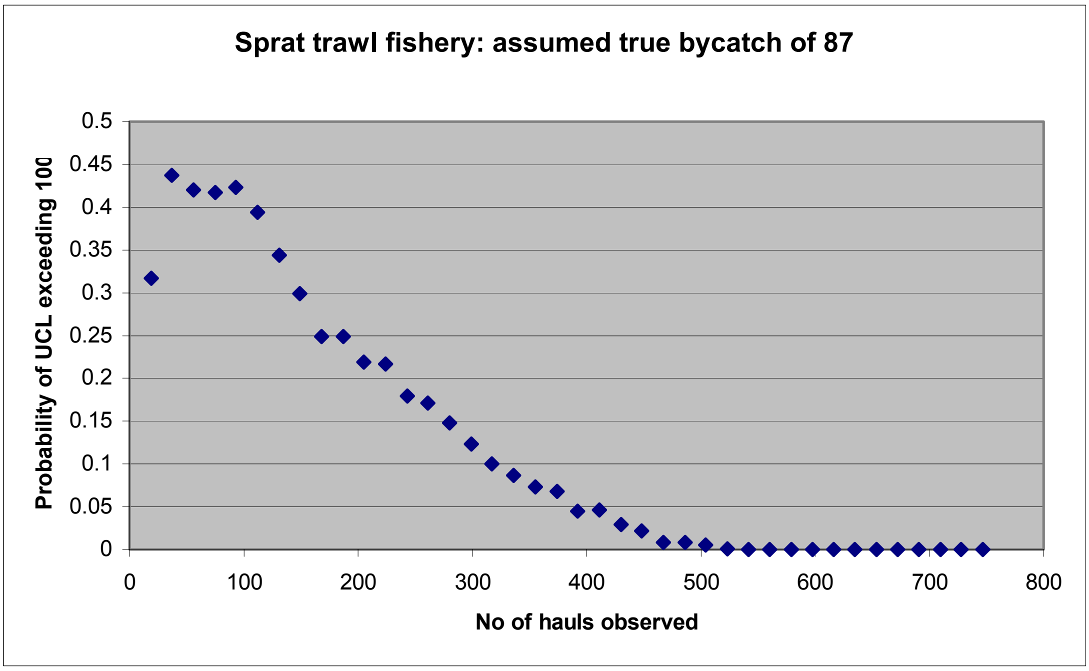


Figure 9: Horse mackerel trawl fishery: assumed bycatch rates less than take limit by 78 animals

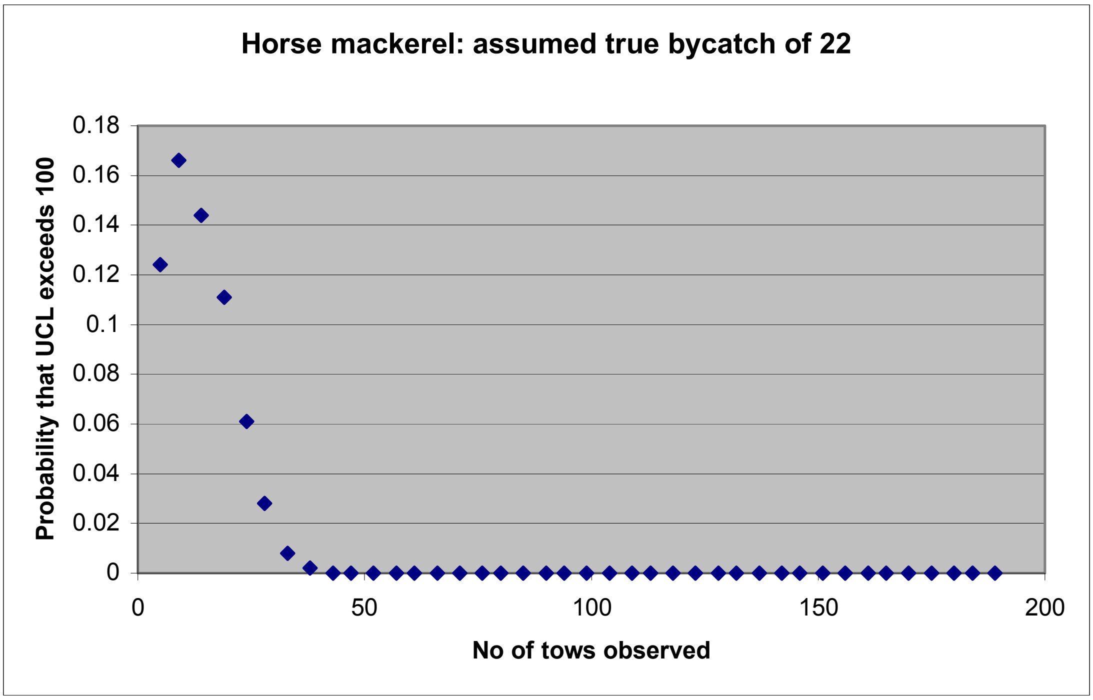
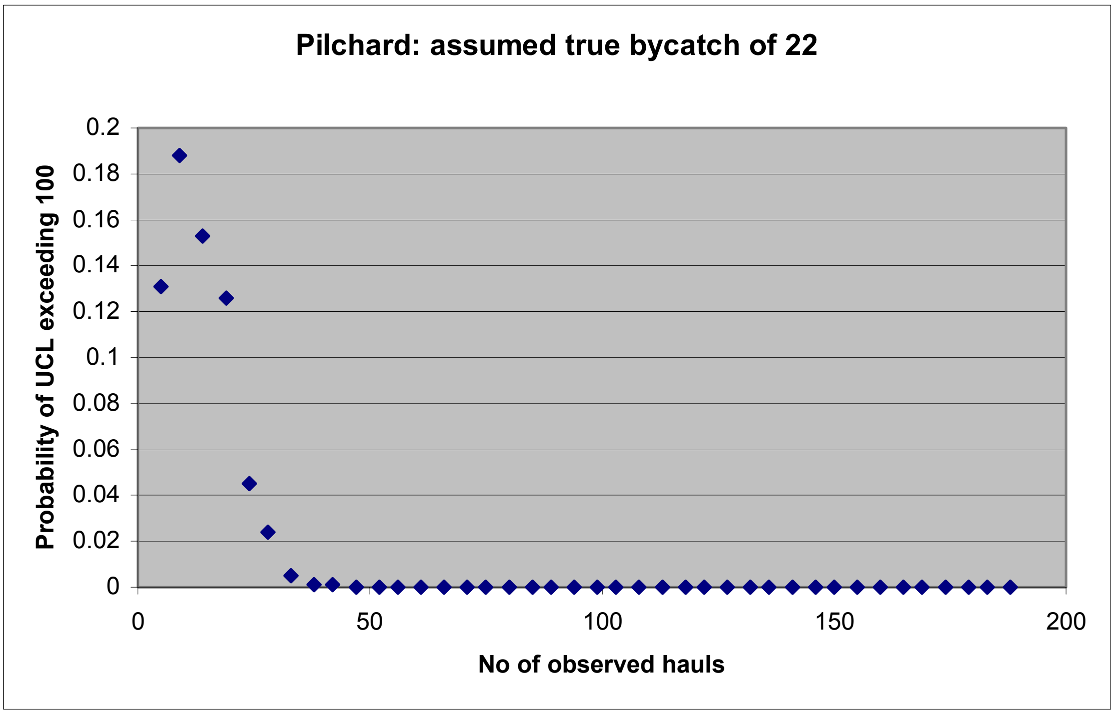


Figure 10: Pilchard trawl fishery: assumed bycatch rates less than take limit by 78 animals



In Figures 8 through 10 we show examples where the expected true bycatch rate is lower (87) and much lower (22) than the take limit. In the case of the sprat fishery it would still be necessary to monitor about half of the 750 tows to be 90% sure that the true bycatch rate s less than 100. For the horse-mackerel and pilchard fisheries however, 90% certainty can be achieved with as few as 30 tows.

Once again, it is clear that at low sampling levels for all of these fisheries, many sets of observations will yield no records of bycatch, which in each case results in a lognormal UCL of zero, which biases the results for the lower sampling regimes. Clearly, if lognormal confidence intervals are to be used, it is important to ensure that, for a given take limit, there is enough sampling to ensure a very low probability of observing zero bycatches

Gillnet fisheries

For gillnet fisheries we have divided all the fisheries into geographical sectors, and estimated the number of hauls during the year 2002. To do this we have assumed an overall mean of 5.4 hauls per day at sea based on our observations, and used the official landings records for England and Wales and for Scotland to obtain estimates of the numbers of days at sea in each of 5 regions (Table 5).

We have used observed bycatch data from Table 2 to set bounds on the likely total catch levels in each area. For a lower bound we have used the lowest measured bycatch rate of 0.0065 porpoises per haul from the North Sea cod fishery, while we have used the highest observed value of 0.0503 from the skate fishery to set an upper limit on likely total bycatches. Where available we have also drawn on published estimates of porpoise abundance to generate 1.7% take limits. Those abundance figures in italics (Irish Sea, West Scotland) are simply guesses. For the Channel and the deep-water areas, we assume that porpoise abundance is zero.

Table 5: Gillnet fisheries scenario.

	Total	Celtic Sea	Channel	Deep water Atlantic	Irish Sea	North Sea	West Scotland
Hauls:							
Assuming 5.4 per day	212549	40565	75460	24948	1750	57035	3159
High rate of bycatch	10691	2040	3796	1255	88	2869	159
Low rate of bycatch	1382	264	490	162	11	371	21
Supposed porpoise abundance		36000	0	0	*15000*	268000	*30000*
Nominal 1.7% Limit		612	0	0	255	4556	510
UK Proportion of catch		0.3	0.5	1	0.6	0.15	1
Nominal UK Take Limit		184	0	0	153	683	510

We have then run the usual simulations for each of the four areas where Take Limits have been defined, this time using two possible 'true' bycatch values. Note that in three of the 8 scenarios the supposed bycatch level already exceeds the take limit, so that in these cases we already expect the bycatch to exceed the take limit, and the exercise therefore tells us how much sampling we need to be sure that it exceeds the limit, rather than how much sampling we need to do to be sure it doesn't exceed the limit.

In all cases we have assumed a binomial distribution and calculated the binomial upper confidence limits. The reasons for this are discussed in the Technical Annex.. Figures 11-14 show the results of these simulations.

For the North Sea, if the true bycatch rate was really as high as 2869 then we could be almost certain of obtaining a UCL greater than 683 with fewer than 200 hauls observed, which is equivalent to less than 40 days at sea or less than 0.5% of effort. If, however, the overall true bycatch total was as low as 264 per year, then it would be necessary to monitor somewhere around 2500 hauls to be 90% certain that the UCL would be less than 683.

For the West of Scotland, whether true bycatches are as high as 159 or as low as 21, we would be 90% sure that the UCL of any bycatch estimate would be less than 510 animals with between 20 and 60 hauls observed.

For the Irish Sea if the bycatch is as low as 11 animals, then we should be 90% sure that about 45 hauls should be enough to generate an estimate with a UCL that is less than the proposed take limit of 153. However, if true bycatches are as high as 88 per year, about 300 hauls would be needed to be sure of producing an UCL that is less than 153. This would amount to about 17% of fishing effort, or 55 days at sea.

Figure 11: Gillnet fisheries in the North Sea – sampling requirements under two assumptions of true bycatch with a take limit of 683 porpoises

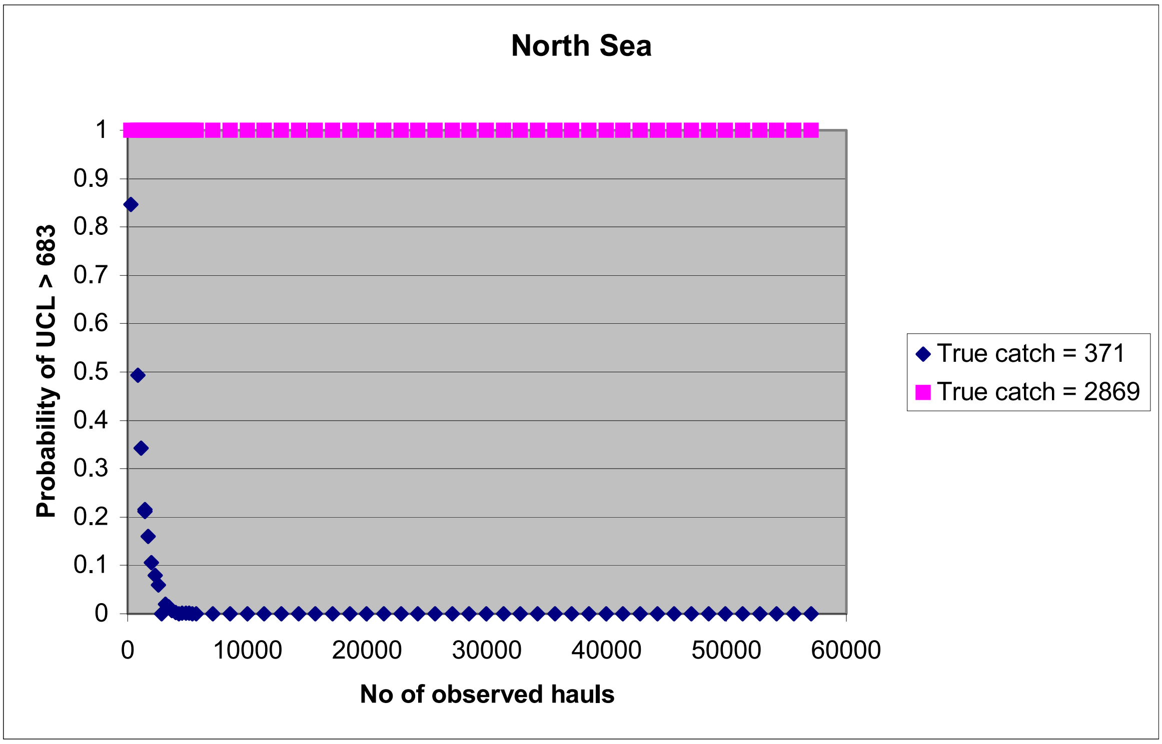


Figure 12: Gillnet fisheries to the West of Scotland – sampling requirements under two assumptions of true bycatch with a take limit of 510 porpoises

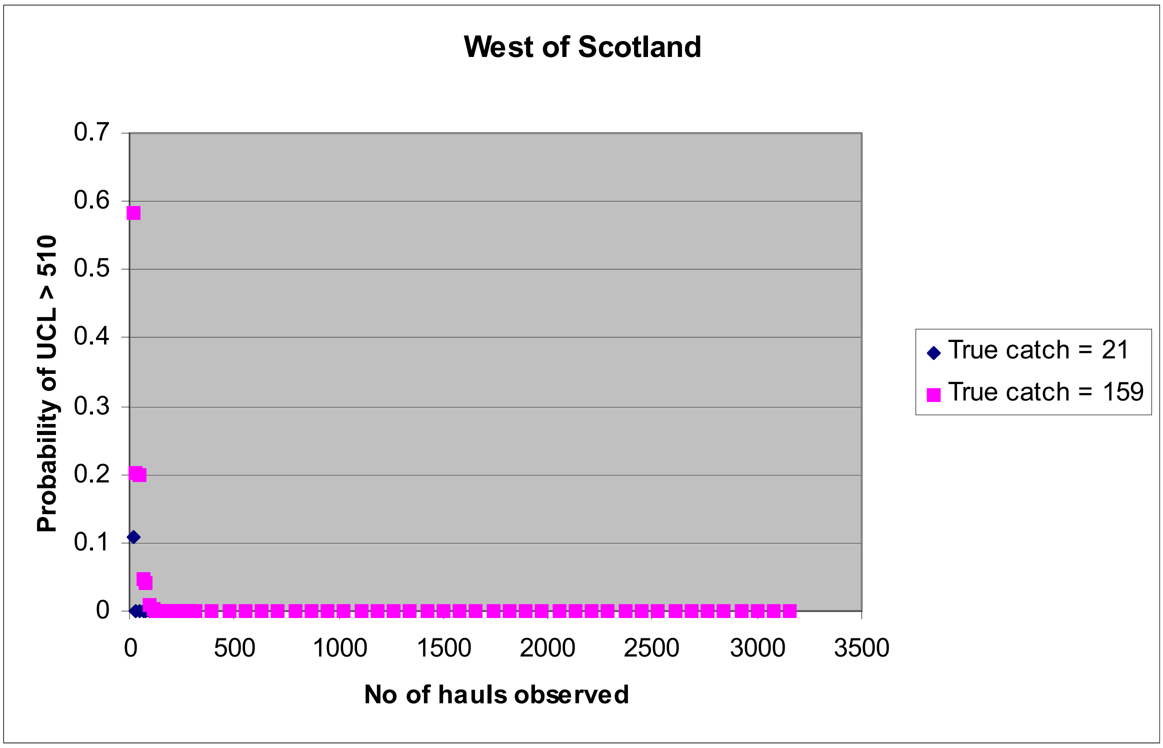


Figure 13: Gillnet fisheries in the Irish Sea – sampling requirements under two assumptions of true bycatch with a take limit of 153 porpoises

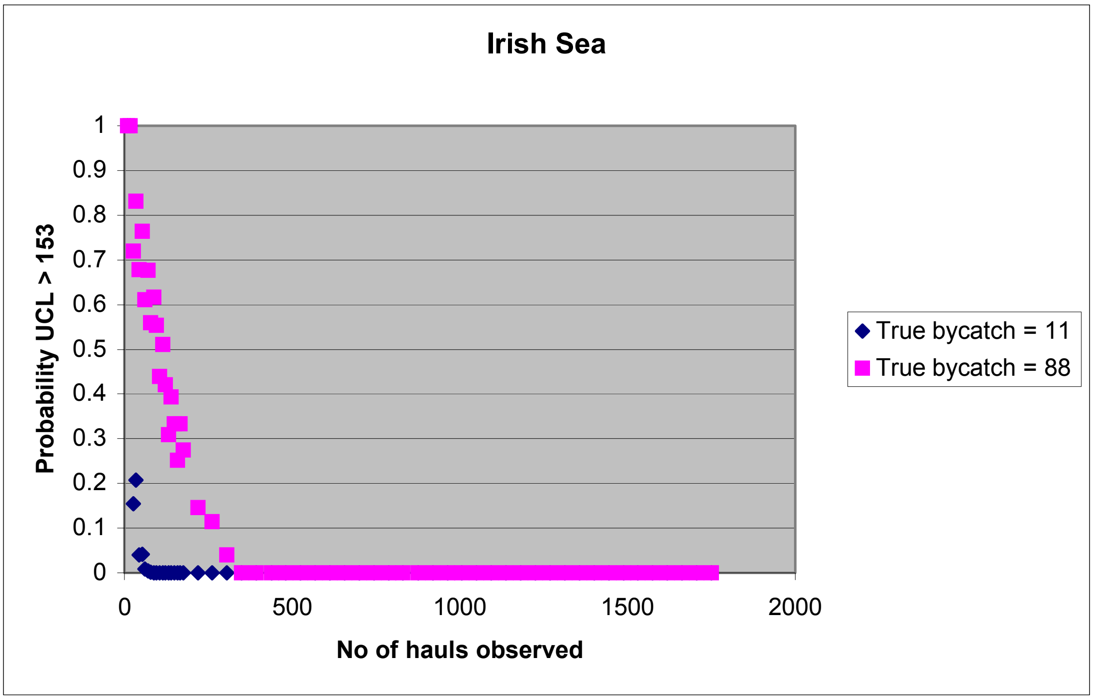
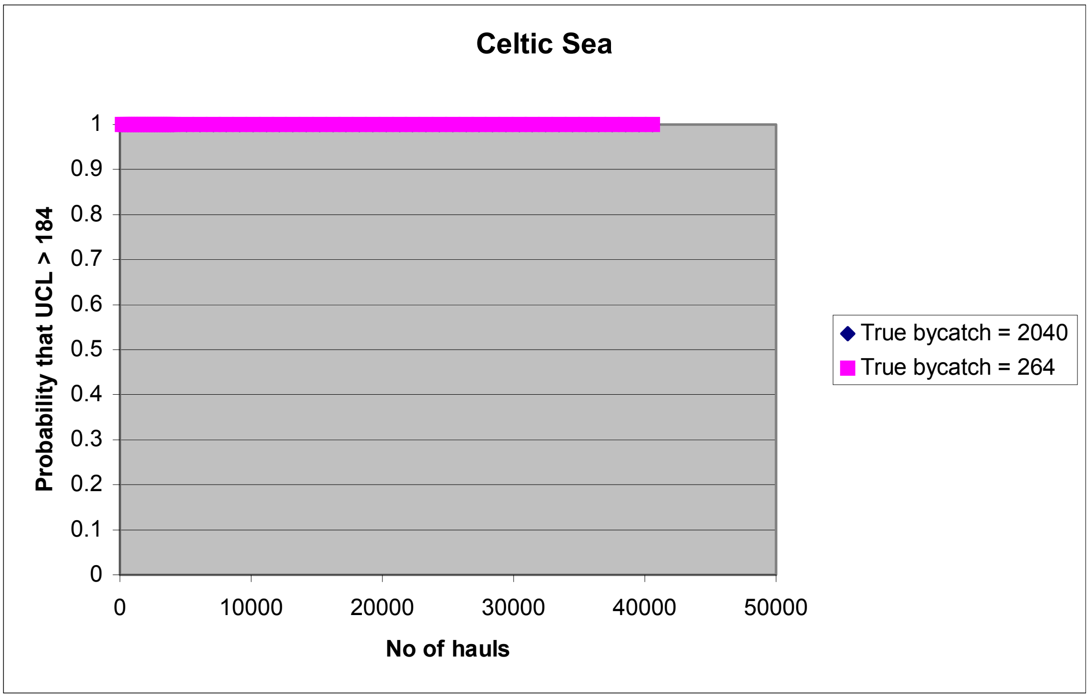


Figure 14: Gillnet fisheries in the Celtic Sea– sampling requirements under two assumptions of true bycatch with a take limit of 184 porpoises



For the Celtic Sea, both the low and high proposed bycatch totals exceed the take limit of 184 by a large margin, which means that very little sampling would be needed to generate an estimate with a UCL higher than 184, and we would be 90% sure of obtaining such an estimate with fewer than 200 hauls

Discussion.

It is clear from the foregoing that there is not a simple answer to the question of how much sampling is required to meet the requirements of the Habitats Directive. The critical uncertainty is the Take Limit, against which any monitored bycatch should be measured. In almost all cases this is poorly defined, either because there are inadequate abundance estimates, or because there is no recognised rationale for allocating parts of a Take Limit based on 1% or 1.7% of an abundance estimate among the fisheries concerned. A second issue is that to estimate the number of observations needed to achieve some level of precision, one needs to have some idea of what the true bycatch rate might be. We have such measures for many of the fisheries concerned, but not all. We can and have estimated a maximum likely number based on the number of zero tows or hauls that we have observed, but this may be far more than the true value.

There are several areas that could be followed up from this work. We have assumed that sampling needs to be done all in one year, whereas the UCL will improve over years as more observations are made, on the assumption that there is an underlying (more or less) constant bycatch rate for each fishery. Bycatch rates may (and do) fluctuate from year to year, however, so sampling over several years would be advisable to determine the extent of any such fluctuation. This is an issue that deserves more attention, as it may be possible to optimise sampling levels through several years, for example by ensuring enough sampling in the first year to be just 70% sure of obtaining an appropriate UCL, with sampling extended over several more years, perhaps at lower levels to improve the accuracy of the estimate, and at the same time look for changes bycatch patterns.

There are several statistical issues that also should be addressed. Perhaps the most important is the possibility of auto-correlation in bycatch events. Thus if bycatch events are clumped in space or time, the assumptions of independence inherent in our modelling approach will be violated. An alternative modelling approach could be devised to deal with this eventuality, but the results would inevitably mean that sampling requirements would need to be increased.

We also need to consider how best to avoid sampling and obtaining observations of no bycatches when we are using lognormal confidence limits, as such zero results lead us to assume we need less monitoring than we really do. The binomial approach describe above is one way of dealing with this.

In our work we have split pelagic trawl fisheries by their target species and gillnet fisheries by areas of operation. It would be preferable to split them all by target species and area of operation, though by so doing we would be spreading existing estimates of bycatch so thinly that we may end up using unrealistic estimates of likely true bycatch rates.

The work described above raises a number of management questions. These include such issues as what Take Limits to set when there is insufficient information, and how best to decide on likely True Bycatch levels. It may be sensible to address these issues in a suitable forum such as the Inter-Agency Marine Mammal Working Group before refining any of the estimates of sampling requirements proposed here.

References

Anonymous, 1996. Report of the Scientific Committee, Annex H. Rep. int. Whal. Commn **46**: 168.

ASCOBANS, 2000. Proceedings of the Third Meeting of Parties to ASCOBANS. ASCOBANS, M.O.P., Bristol, United Kingdom.

Barlow, J., 1989. Estimating sample size required to monitor marine mammal mortality in California gillnet fisheries. La Jolla, Southwest Fisheries Science Centre**:** 8.

Blaker, H., 2000. Confidence curves and improved exact confidence intervals for discrete distributions. Canadian Journal of Statistics (28):783-798

Goujon, M., 1996. Captures accidentelles du filet maillant derivant et dynamique des populations de dauphins au large du Golfe de Gascogne. Laboratoire Halieutique, Ecole nationale Superieure Agronomique de Rennes**:** 239.

Goujon, M., L. Antoine, Collet, A., and Fifas, S., 1993. Approche de l'impact ecologique de la pecheries thoniere au filet maillant derivant en Atlantique nord-est., IFREMER**:** 47.

Hammond, P. S., P. Berggren, Benke, H, Borchers, D. L., Collet, A., Heide-Jorgensen, M. P., Heimlich, S., Hiby, A. R., Leopold, M. F., Oien, N., 2002. Abundance of harbour porpoise and other cetaceans in the North Sea and adjacent waters. Journal of Applied Ecology **39**(2): 361-376.

Razzaghi, M. 2002. On the estimation of binomial success probability with zero occurrence in sample. Journal of Modern Applied Statistical Methods (1):326-332

Ridout, M.S., C.B.G. Demetrio and J.P. Hinde, 1998. Models for count data with many zeros. Proceedings of the XIXth International Biometric Society Conference, IBC98, Cape Town, December 1998:179-192.

SEC, 2002. Subgroup on fishery and environment (SGFEN). Scientific, technical and economic committee for fisheries (STECF). Incidental catches of small cetaceans. Brussels, Commission of the European Communities**:** 83.

Smith, T., 1991. Overview of the assessment of harbor porpoise status. SAW/13/SARC/4. Woods Hole MA.

Wade, P., 1999. Planning observer coverage by calculating the expected number of observed mortalities. Development of a process for the long-term monitoring of MMPA Category I and II commercial fisheries. Proceedings of a Workshop, Silver Spring, Maryland.

Wilson, B., P. Hammond, P. Thompson, 1999. Estimating size and assessing trends in a coastal bottlenose dolphin population. Ecological Applications **9**(1): 288-300.

Zar, J.H., 1996. Biostatistical Analysis. 3rd Edition. Prentice Hall, Upper Saddle River, New Jersey.

Technical annex

A Binomial probability of observing zero bycatch

This section repeats the method presented by Wade (1999) for determining the probability of observing zero bycatch (with slightly different notation). We assume that: bycatch is a binomial random variable – i.e., that there is a fixed probability, p, of observing no bycaught animals in a haul; each haul is independent and identically distributed; the maximum number of animals caught per haul is 1.

The probability mass function (pmf) for a binomial distribution gives us the probability of observing K bycaught animals in H_o observed hauls as

$$P(K|H_o, p) = \binom{H_o}{K}(1-p)^K p^{H_o-K} \tag{1}$$

So the probability of observing zero bycatch ($K = 0$) is

$$P(0|H_o, p) = p^{H_o} \tag{2}$$

B Distribution of bycatch data

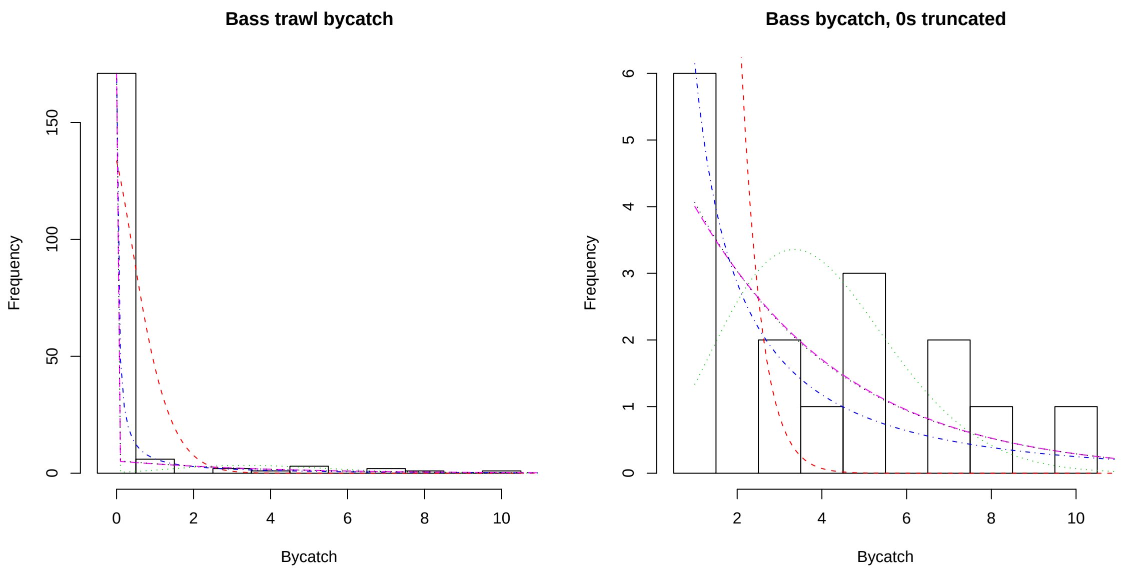
B.1 Bass trawl data

The data consist of 187 tows observed on UK bass pair trawlers working in the Channel between 2001 and 203. Of these 171 had no bycatch and 16 recorded bycatch of common dolphins. The numbers caught were: 1, 1, 1, 1, 1, 1, 3, 3, 4, 5, 5, 5, 7, 7, 8, 10 (Figure 1). These data share two features with many other bycatch datasets: (1) a large proportion of zeros, and (2) a wide range of numbers in the non-zero observations, with a few quite large schools caught. From these data, the mean catch rate is 0.337 and the variance in catch rate is 1.912.

Count data is often modelled as coming from a Poisson distribution. In that distribution there is one parameter, λ, the mean catch rate. This is also the variance of the distribution, so given that the variance of the data are 5 times the mean this is clearly going to fit badly. Overdispersed count data are often modelled using the negative binomial distribution. This distribution has two parameters; there are various parameterizations, and one way to describe the distribution is as the number of 'failures' that occur before r successes are observed, when the probability of a success is q. Under this parameterization, the mean is $r(1-q)/q$ and the variance $r(1-q)/q^2$.

Despite the additional flexibility given by the extra parameter, data with so many zero values may still not be modelled well with the negative binomial distribution. Hence we also investigated three 'zero inflated' distributions: zero inflated Poisson (ZIP), zero inflated negative binomial (ZINB) and zero inflated gamma (ZIG). These are mixture distributions – mixing zero counts that occur with probability p with Poisson, negative binomial or gamma distributed counts that occur with probability $1-p$. If $p > 0$ then you get more zeros than would be predicted under the Poisson, negative binomial or gamma distribution alone. (Note that in the parameterization used here, no zeros would

Figure 1: *Bass bycatch data with 5 fitted distributions: Poisson (red dashed), negative binomial (blue dot-dashed), ZIP (green dotted), ZINB (pink dashed) and ZIG (black dotted, almost identical to ZINB). For clarity, the right figure shows the data and distributions with zero counts removed.*



be expected under the gamma distribution alone – see below.) The ZIP distribution is used quite commonly for data with many zeros (see review by Ridout et al. 1998), and we are aware of at least one other application of the ZINB distribution for ecological data (B.R. Gray, pers. comm.) but not of the ZIG distribution.

The ZIP distribution, has two parameters, p and λ (note that the mean catch rate is now no longer simply λ, but is $(1-p)\lambda$). The pmf is

$$f(x|p,\lambda) = \begin{cases} p + (1-p)e^{-\lambda} & : \quad x = 0 \\ (1-p)e^{-\lambda}\lambda^x/x! & : \quad x > 0 \end{cases} \qquad (3)$$

where p is bounded $(0, 1)$ and λ is bounded $(0, +\infty)$. The ZINB distribution has three parameters, p, q and r, and the pmf is:

$$f(x|p,r,q) = \begin{cases} p + (1-p)q^r & : \quad x = 0 \\ (1-p)\binom{r+x-1}{x}q^r(1-q)^x & : \quad x > 0 \end{cases} \qquad (4)$$

where p and q are bounded $(0, 1)$ and r is bounded $(0, +\infty)$.

The geometric distribution can be thought of as a special case of the negative binomial when r is fixed at 1. In addition, we here use the definition that it is the number of *trials* before obtaining 1 success, rather than the number of *failures*. This means the distribution of counts is in the range $1, 2, \ldots, +\infty$, and does not include 0. The mean of the geometric distribution is then $1/q$. Following along the same lines as for the two previous zero inflated distributions, the ZIG therefore has two parameters, p and q and has pmf:

Table 1: *Parameter estimates, log-likelihood and AIC for 5 distributions fitted to bass bycatch data*

Distribution	Parameter estimates	LnL	AIC	ΔAIC
ZIG	$p = 0.9144(0.0204)$, $q = 0.2539(0.0584)$	-90.33	184.66	0
negative binomial	$r = 0.0402(0.0132)$, $q = 0.1067(0.0482)$	-91.19	186.38	1.72
ZINB	$p = 0.8868(0.0443)$, $r = 1.0494(1.1575)$, $q = 0.2606(0.1639)$	-90.33	186.66	2.00
ZIP	$p = 0.9126(0.0209)$, $\lambda = 3.854(0.5062)$	-94.84	193.68	9.02
Poisson	$\lambda = 0.3367(0.0424)$	-195.42	392.84	208.18

$$f(x|p,q) = \begin{cases} p + (1-p)q & : \quad x = 0 \\ (1-p)q(1-q)^x & : \quad x > 0 \end{cases} \tag{5}$$

where p and q are both bounded $(0, 1)$. The parameters of the ZIG have intuitive meaning: p is the probability of getting a zero bycatch, and $1/q$ is the mean number caught in hauls containing bycatch. (As an aside, the mean bycatch rate is $(1-p)/q$ and the variance is $(1-p)(1+p-q)/q^2$).

We fit these distributions to the bass trawl data by numerically maximizing the log likelihood (i.e., the log of the pmf). All calculations reported in this annex were performed using the free statistical sofware R, and code is available on request. The fitted distributions are plotted together with the data in Figure 1, and parameter estimates and model selection statistics are shown in Table 1.

AIC values indicate that the ZIG distribution is the preferred model (lowest AIC), with the negative binomial and ZINB second best with very similar AICs. The likelihood of the ZIG and ZINB distributions are almost the same, and from the figure it can be seen that the fits are almost identical. The ZIG distribution is therefore preferred because it uses one less parameter (which explains why the AIC is 2 lower). We also note that the estimate of r from the ZINB is close to 1, which provides support to the idea of fixing it at 1 as in the ZIG distribution. The ZIP distribution comes a poor fourth place and the Poisson distribution does not fit the data at all well, as expected.

The above addresses the question of relative fit of different models, but it is important also to test absolute fit. A χ^2 goodness-of-fit test of the ZIG distribution, pooling group sizes of 5 and larger, yields $\chi^2 = 20$ on 16 d.f., making $p = 0.2202$.

B.2 Irish tuna pair trawl data

These data were kindly provided by Dominic Rihan, BIM. They are of 520 observed tows on tuna fishing vessels. Of these, 484 had no bycatch, the remaining 36 observations were: 6 1's, 11 2's, 4 3's, 4 4's, 5 5's, 1 7, 2 9's, 1 10, 1 15 and 1 30 (Figure 2).

We fit the same distributions to these data, and the results are shown in Figure 2 and Table 2.

Again, the ZIG distribution is favoured, with the ZINB second. Although the estimate of r from the ZINB is 0.6381, far from 1, it's standard error is large (0.4307) and so the estimate is not significantly different from 1. Looking at absolute fit of the ZIG distri-

Figure 2: *Irish tuna bycatch data with 5 fitted distributions: Poisson (red dashed), negative binomial (blue dot-dashed), ZIP (green dotted), ZINB (pink dashed) and ZIG (black dotted, almost identical to ZINB). For clarity, the right figure shows the data and distributions with zero counts removed.*

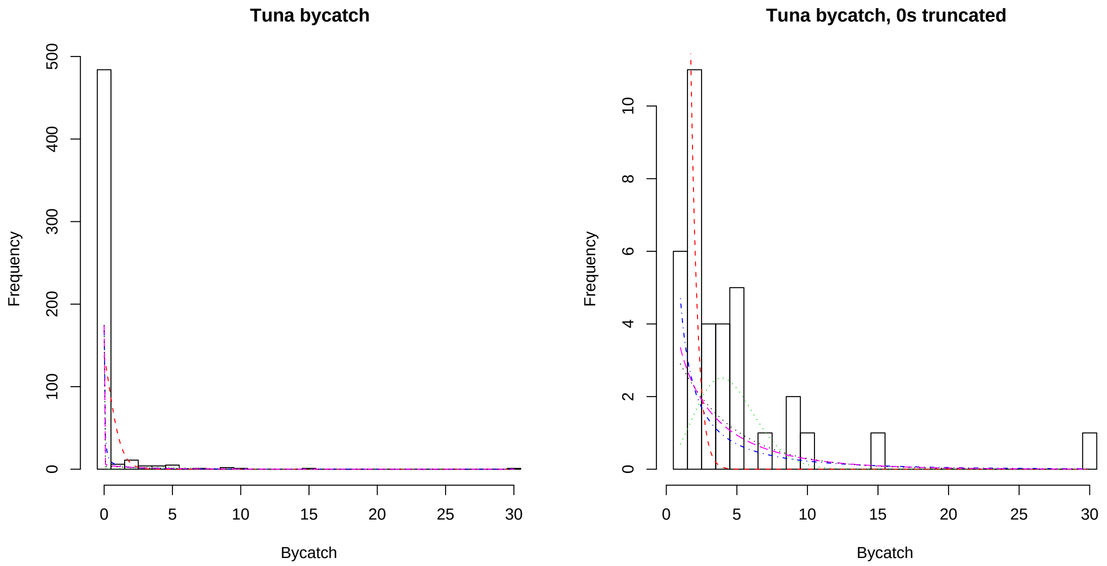
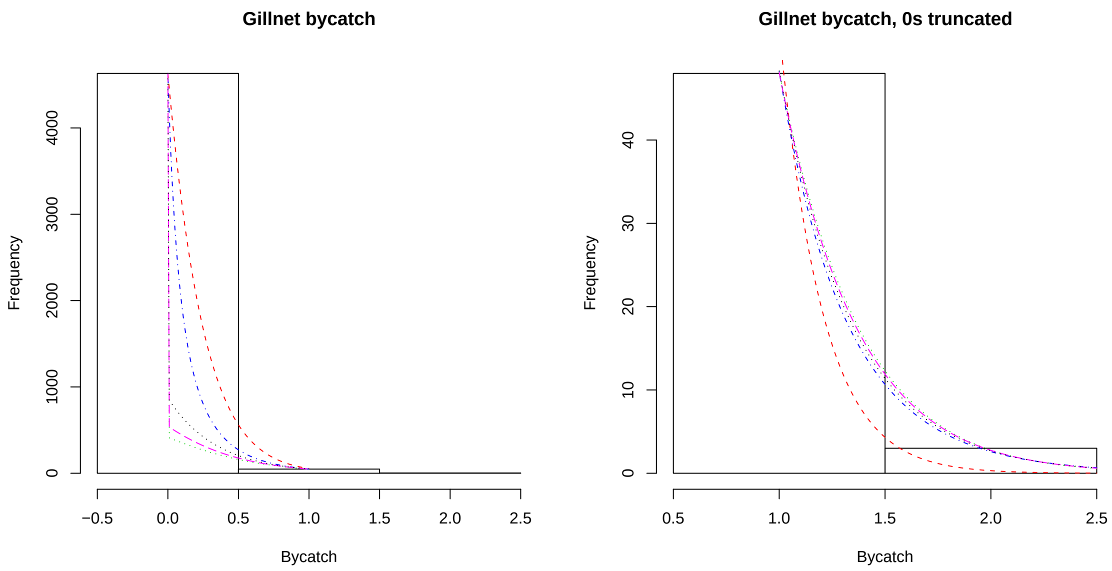


Table 2: *Parameter estimates, log-likelihood and AIC for 5 distributions fitted to Irish tuna bycatch data*

Distribution	Parameter estimates	LnL	AIC	ΔAIC
ZIG	$p = 0.9307(0.0111)$, $q = 0.2236(0.0328)$	-216.42	436.83	0
ZINB	$p = 0.8967(0.0322)$, $r = 0.6381(0.4307)$, $q = 0.1754(0.0677)$	-216.14	438.27	1.44
negative binomial	$r = 0.0296(0.0063)$, $q = 0.0872(0.0271)$	-217.77	439.56	2.73
ZIP	$p = 0.9299(0.0112)$, $\lambda = 4.418(0.3578)$	-253.50	511.01	74.18
Poisson	$\lambda = 0.3096(0.0244)$	-552.99	1107.98	671.15

Figure 3: *Gillnet bycatch data with 5 fitted distributions: Poisson (red dashed), negative binomial (blue dot-dashed), ZIP (green dotted), ZINB (pink dashed) and ZIG (black dotted, almost identical to ZINB). For clarity, the right figure shows the data and distributions with zero counts removed.*



bution, a χ^2 goodness-of-fit test of the ZIG distribution, pooling group sizes of 10 and larger, yields $\chi^2 = 77$ on 70 d.f., making $p = 0.2647$.

B.3 Gillnet data

These data come from the Sea Mammal Research Unit observer programme between 1995 and 2000. The data consist of 4632 hauls with 4581 empty, 48 with 1 animal and 3 with 2 animals. These data are not going to be terribly informative about which distribution to fit, since they are 'mostly binomial' – in other words even a perfectly fitting distribution will do about as well as assuming that the 3 hauls with 2 animals were in fact 6 hauls with 1 animal and then proceeding as if the data were binomial.

In any case, we can fit the various distributions to the data (Figure 3, Table 3). From the AIC values, it is clear that any of the 2 parameter distributions provide an almost equally good fit to the data, with the 3-parameter ZINB being penalized for the extra parameter so having an AIC 2 higher, and the Poisson distribution fitting much less well. Again, the r parameter in the ZINB distribution is not significantly different from 1, although it has a very high variance which highlights the fact that fitting this distribution is asking too much of the data.

Table 3: *Parameter estimates, log-likelihood and AIC for 5 distributions fitted to gillnet bycatch data*

Distribution	Parameter estimates	LnL	AIC	ΔAIC
ZIP	$p = 0.9000(0.0555)$, $\lambda = 0.1154(0.0659)$	-292.76	589.52	0
ZIG	$p = 0.9891(0.0015)$, $q = 0.9444(0.0311)$	-292.82	589.64	0.12
negative binomial	$r = 0.1072(0.0710)$, $q = 0.9030(0.0593)$	-292.87	589.73	0.21
ZINB	$p = 0.8727(0.1185)$, $r = 3.3837(10.0316)$, $q = 0.9739(0.0578)$	-292.79	591.58	2.06
Poisson	$\lambda = 0.0115(0.0016)$	-297.07	596.13	6.61

C Estimation of upper confidence limits

As outlined in the body of the report, we suggest basing management decisions on one-sided upper confidence limits (UCLs) on bycatch. The calculation of these confidence limits is outlined in this section.

There are several potential methods for obtaining confidence limits on bycatch given some data.

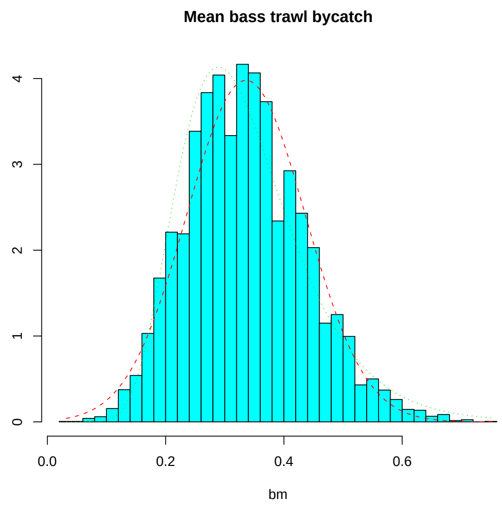
1. If we are unsure of the distribution of the bycatch data, we could use a nonparametric bootstrap. This method has the advantage that it makes no distributional assumptions. It has two disadvantages. Firstly, it is computer intensive, and so is not really feasible for use as part of a simulation exercise (as in section D). Secondly, and more importantly, when no bycatch is observed, the upper confidence limit will be estimated to be zero, regardless of the size of the dataset. Clearly when we observe no bycatch from, say, only 10 hauls, we are not certain that the true bycatch in the entire fishery is zero!

2. If the data are thought to follow a ZIG distribution, we could use asymptotic theory for maximum likelihood estimates which states that the estimates are asymptotically normally distributed. The variance of the mean of a ZIG distribution is $(1-p)(1+p-q)/(q^2 H_o)$ where H_o is the number of observed hauls, so by substituting in estimates for p and q we can obtain the standard error and so z-based confidence intervals. There are two potential problems with this approach. Firstly, the mean bycatch may not be normally distributed with smaller sample sizes, and secondly when we observe no bycatch we have no information with which to estimate the q parameter, so the variance becomes undefined.

3. An alternative to assuming the mean bycatch is normally distributed would be to assume it is lognormally distributed. This makes sense because mean bycatch cannot be lower than zero, so we might expect a normal assumption to be invalid especially at lower bycatches. In common with previous methods, when we observe no bycatch the estimated UCL will be zero.

4. If we could assume that bycatch follows a Poisson distribution then we can calculate exact confidence limits (computer intensive) or use approximations based on a χ^2 distribution (Zar 1996, p572). An advantage of this approach is that the UCL is not

Figure 4: *Histogram of the means of 9999 bootstrap resamples from the bass bycatch data, with a normal (red line) and lognormal (green line) distribution fitted to the means.*



zero when we observe zero bycatch. A disadvantage is that the Poisson distribution does not appear to fit the datasets we analyzed well (see previous section).

5. If the chance of catching more than 1 animal in a net is very small (such as with the gillnet data of the previous section) we could assume that bycatch follows a binomial distribution and calculate exact confidence limits (again, computer intensive) or use approximations based on the F-distribution (Zar 1996, p524). Again, the UCL will not be zero when we observe zero bycatch.

We examined options 2 and 3 by generating 9999 bootstrap resamples of the the bass bycatch data, taking the mean of each, and fitting both normal and lognormal distributions to the means (Figure 4). Both fit the data reasonably well, although neither are broad enough at the peak or fit very well in the tails of the distribution. This is confirmed by qq-plots (not shown). Kolmogorov-Smirnov tests favour the lognormal distribution ($D = 0.6998$ for the normal distribution, $D = 0.4916$ for the lognormal; p-values are irrelevant as the bootstrap data are an arbitrary sample size). If we consider bootstrap confidence limits (based on the percentile method) as the 'gold standard' then we can compare the normal and lognormal confidence limits with the bootstrap ones (Table 4). In terms of the upper limit (which we are more interested in), the Normal CIs appear slightly closer.

To come to a firm conclusion about the better analytic confidence interval from ZIG data, it would be necessary to perform a simulation study at a range of sample sizes and ZIG parameter values – something outside the scope of this report. A priori we prefer the lognormal intervals because they cannot go below zero, so in the bycatch simulations of section D we use lognormal confidence intervals when we simulate from a ZIG distribution. It would be nice to come up with a better parametric approximation to the ZIG confidence limits (like the F-based limits for the binomial distribution, or χ^2 limits for

Table 4: *Comparison of bootstrap, normal and lognormal confidence limits estimated from bass bycatch data*

Assumption	95% Confidence limits	Difference from bootstrap CIs
Bootstrap	0.1604,0.5561	0,0
Normal	0.1394,0.5323	-0.0211,-0.0239
Lognormal	0.1721,0.5961	0.0116,0.0399

the Poisson) but this is again outside the scope of this report.

In section D, we perform some simulations using the binomial distribution, and for these we use approximate binomial confidence limits. The calculation of these and the lognormal limits are described in the following sub-sections. In calculating both limits, we follow a suggestion of Mark Bravinton (pers. comm.) to incorporate finite population correction into the calculation. This has the effect of substantially reducing the width of the confidence interval as the observed hauls becomes a reasonably large proportion of the total hauls.

C.1 Lognormal confidence limits on bycatch

Let H be the total number of hauls of which H_o are observed and H_u unobserved. On each observed haul, k_h animals are observed to be caught ($h = 1, \ldots, H_o$), making a total of K_o animals observed killed. The estimated mean bycatch, $\hat{k}_. = K_o/H_o$ has sampling error

$$\hat{\text{SE}}(\hat{k}_.) = \frac{\sum_1^{H_o}(k_h - \hat{k}_.)^2}{H_o(H_o - 1)} \tag{6}$$

and the estimate of total bycatch over H hauls is $\hat{K} = K_o + H_u k_.$ with standard error $\hat{\text{SE}}(\hat{K}) = H_u \hat{\text{SE}}(\hat{k}_.)$ and coefficient of variation

$$\hat{\text{CV}}(\hat{K}) = \frac{H_u \hat{\text{SE}}(\hat{k}_.)}{\hat{K}} \tag{7}$$

The upper $(1 - \alpha)$-level one-sided lognormal confidence interval can then be calculated as

$$\hat{\text{UCL}}_{alpha}(\hat{K}) = K_o + H_u \exp\left[\mu + \Phi_\alpha \sqrt{\sigma^2}\right] \tag{8}$$

where $\mu = 1/2 \ln\left(\hat{k}_.^{\,2} / \left[1 + \hat{\text{CV}}(\hat{K})\right]\right)$, $\sigma^2 = \ln(1 + \hat{\text{CV}}(\hat{K}))$, and Φ_α is the upper α-level quantile from the standard normal distribution.

C.2 Binomial confidence limits on bycatch

There are at least 4 ways to calculate an α-level one-sided upper binomial confidence limit on bycatch, given estimated bycatch rate p. The exact method involves finding the p such that the integral of the pmf from p to 1 conatins $(1 - \alpha)$ of the integral from 0 to 1.

Table 5: *Comparison of four different binomial upper 95% one-sided confidence limits based on zero bycatch observed in 114 trawls.*

Method	UCL
Exact	0.025716
F-based	0.025936
Profile Lnl	0.016
Blaker	0.000

This is computer-intensive since it must be done numerically. A much faster alternative is the standard F-based approximation (Zar 1996, p524):

$$\hat{\text{UCL}}_\alpha(\hat{p}) = \frac{(K_o + 1)F_{\alpha,\nu_1,\nu_2}}{H_o - K_o + (K_o + 1)F_{\alpha,\nu_1,\nu_2}} \tag{9}$$

where $\nu_1 = 2(K_o + 1)$ and $\nu_2 = 2(H_o - K_o)$. This has quite poor coverage at low H_o or K_o and extreme levels of α. Another method is the profile likelihood, which is based on the asymptotic distribution of the log-likelihood surface so is again approximate at small sample sizes. There is also a method by Blaker (2000) which apparently has better coverage properties. We compared these over a range of situations (not reported here) and came to the conclusion that the approximate F-based method was quite adequate for our purpose. For example, estimated upper confidence limits for the pelagic trawl data are given in Table 5. Here, no bycatch was observed in 114 trawls. The F-based UCL is nearly identical to the exact limit.

Given the estimated UCL on bycatch rate, we can estimate UCL on total bycatch in the same was as in the previous subsection as:

$$\hat{\text{UCL}}_\alpha(\hat{K}) = K_o + H_u \hat{\text{UCL}}_\alpha(\hat{p}) \tag{10}$$

As an aside, the problem of estimating the binomial probability of a rare event given limited data showing all zeros is addressed in an interesting paper by Razzaghi (2002). Using the usual frequentist methods, when we observe no bycatch, the estimated bycatch rate is 0.0, whatever the sample size. Razzaghi describes and compares several methods (including Bayesian methods) where the estimate will be greater than 0 for small samples, better reflecting our belief that the true bycatch rate is not zero. An alternative approach is to manage based on UCLs as we have suggested in this report.

D Simulation studies

The aim of the simulation studies was to predict the proportion of estimated UCLs that exceed a given take limit, given the total number of hauls, H, the number of observed hauls, H_o, the true number of cetaceans caught, K, and the confidence level of the UCL, α. We did two types of simulation, as detailed below.

In the first, we assumed that bycatch follows a ZIG distribution, and that the UCLs would be estimated using the lognormal method given in section C.1. For these simulations, in addition to the true bycatch, K, we were required to specify the q parameter from the ZIG distribution. Given this q, we could calculate the expected frequency of bycatches of $k = 1, 2, 3, \ldots$ by multiplying the geometric density by K and rounding:

$$E(\text{freq}(k = X)) = \text{round}\left[Kq(1-q)^{X-1}\right] \tag{11}$$

Remaining hauls were assigned a bycatch of 0. This gave us the 'true' set of bycatches for all H hauls, which we could then sample from to create observed bycatch datasets. For each simulation, we sampled H_o hauls at random from the 'true' set and calculated the lognormal UCL. We repeated this 1000 times, and calculated the proportion of the 1000 lognormal UCLs that exceeded the take limit.

In the second set of simulations, we assumed that bycatch follows a binomial distribution, and that the UCLs would be estimated using the approximate F-based method given in section C.2. In this case, the 'true' bycatches are simply K catches of one animal and $H - K$ zeros. We sampled from this to create observed bycatch datasets as in the previous simulation, and for each dataset calculated the binomial UCL. We repeated this 1000 times, and calculated the proportion of the 1000 binomial UCLs that exceeded the take limit.

Each of these simulations was repeated over a range of values of H, H_o, K and, for the ZIG simulations q, as detailed in the main body of the report.

One refinement of the above would be to repeat the second simulations assuming that bycatch follows a Poisson distribution. We expect that the results would be almost identical to those we presented in the report, so long as assumed bycatch rates were low. For high bycatch rates, the ZIG-based simulations are more appropriate(Section B).