

# Center for Independent Experts (CIE) Independent Peer Review of Sampling Design and Inferential Methods for Estimating Bycatch in the Hawaii Deep-Set Longline Fishery

Meeting Dates: August 24-28, Honolulu, Hawaii

Report to:

The Center for Independent Experts

Prepared by:

Mary C. Christman

MCC Statistical Consulting LLC

Gainesville, FL 32605

marycchristman@gmail.com

October 2015

# Table of Contents

Executive Summary .....	3
Background.....	4
Description of Reviewer’s Role in the Review Activities.....	4
Summary of Findings .....	5
ToR 1.....	6
Review the sampling design used to select trips for observer placement and determine if it is a preferred design for estimating bycatch considering constraints and reporting requirements. ...	6
ToR 2.....	7
Evaluate the point estimators and determine if they are good estimators given the sample design, observed frequency distribution of bycatch events, and constraints. ....	7
ToR 3.....	11
Evaluate the interval estimators and determine if they are good estimators given the sample design, observed frequency distribution of bycatch events, and constraints. ....	11
ToR 4.....	14
Evaluate estimators of total bycatch events resulting in a death or serious injury (DSI) classification and determine if they are good estimators given the sample design, observed frequency distribution of injury classifications (non-serious or DSI), and constraints. ....	14
ToR 5.....	15
Evaluate the subpopulation estimators being applied to estimate bycatch within a political geographical boundary and determine if they are good estimators given the sample design, reporting requirements under the MMPA, and constraints.....	15
ToR 6.....	15
Suggest future research priorities to improve methods for estimating bycatch with increased efficiency given the current data structure. Suggest future research priorities for improving the sampling design for the purposes of estimating bycatch, with efficient use of sampling resources as a consideration. ....	15
Conclusions and Recommendations.....	16
Appendix 1: Bibliography of materials provided for review .....	17
Appendix 2: A copy of the CIE Statement of Work.....	18
Annex 1: Format and Contents of CIE Independent Peer Review Report .....	23
Annex 2: Terms of Reference for the Peer Review .....	24
Annex 3: Tentative Agenda.....	25
Appendix 3: Panel Membership .....	26

## Executive Summary

Methodology for sampling and estimating bycatch of the Hawaii Deep-Set Longline Fishery was reviewed. This document reports the main findings of the methodology review and lists recommendations for future research. I note that the recommendations are just that and are not to be presumed to be criticism that the current approaches are incorrect. Instead, the recommendations provide possible approaches to increase the efficiency of the process of providing annual estimates of total species-specific bycatch.

The sampling design for observer assignment in the Hawaii DSLL fishery is composed of two parts in order to address the many constraints for observer placement. It is challenging to design a statistically rigorous sampling plan when the size of sampling frame is undetermined and the number of observers and their availability fluctuate. The two-stage SYSPLUS sample is a novel and highly adaptive approach that can effectively accommodate uncontrollable constraints. The design is independent of the large number of species for which bycatch estimates are required, many of which are often rarely observed. The amount of work and its complexity are impressive.

The review found that the main concerns reside in estimating total bycatch using confidence interval estimation. The point estimators are simple design-based estimators whereas the confidence interval estimators are either based on non-parametric bootstrapping or model-based approaches. Most of these approaches seem overly complicated. This report made several specific recommendations for modifications or future research:

- a) The large number of bycatch species should be divided into two or three groups according to their observed frequency. Different group of species should use appropriate estimation technique(s) which should be used consistently across years. Clear criteria for species categorization may be developed by examining historical data and results.
- b) Review the current sampling design for simplification or changes in the choices of allocation. For example, is 5 the optimal number of systematic starts during the year? Could a two-stage cluster sampling design with observers assigned at random to the set of selected SSUs be used in place of the current SYSPLUS design? Such an alternative would simplify the estimation procedures, yet still provide similar efficiency as the current design.
- c) The current method of estimating the inclusion probabilities for the PLUS sample appears to be overly complicated, yet at the same time averages the true inclusion probabilities over several adjacent time periods. These average probabilities are not likely to be close to the true probabilities, and so may be introducing bias into the estimation procedure. This method should be reviewed and possibly modified based on simulation studies.
- d) The current methods of interval estimation should be reviewed for simplification, and further there should be a concerted effort to associate point estimation methodology with the appropriate confidence interval estimator. For example, one should not use a design-based estimator based on unequal probability sampling with a confidence interval estimator based on a model which assumes data generating probability distribution and iid observations.
- e) Care should be taken when doing domain estimation as the subsetting of the sample data into smaller sets for estimation can lead to problems in estimating total bycatch by domain

using design-based approaches. This is especially problematic for species that are rare. In such cases, a model based approach that borrows strength from other areas should be considered.

## Background

The Hawaii deep-set longline (DSL) fishery catches over a hundred bycatch species, including marine mammals, seabirds, sea turtles, fishes and others. A combination of systematic and adaptive sampling strategies have been used to collect bycatch data for approximately the last 10 years. The adaptive component is in response to the uneven availability of observers during the year.

Based on the observer data, fleet-wide estimates of the total bycatch by taxonomic group are calculated using design-based estimators. Marti McCracken, NMFS Pacific Islands Fisheries Science Center, is required to provide annual point estimates of total bycatch by species or taxon, but also provides estimates of standard error (SE) and interval estimators. This document covers a review of the sampling design and the methods used to estimate bycatch and the uncertainty associated with total bycatch estimation.

The Center for Independent Experts (CIE) invited three independent reviewers for the DSL fishery. Background information regarding the Hawaii longline fishery and observer program was presented during the meeting. The project team provided detailed explanations about the methodology. Extensive discussions were exchanged between the team and the review panel.

## Description of Reviewer's Role in the Review Activities

Two weeks before the review meeting, I received three documents from the NMFS Project Contact:

1. "pifsc.bycatch.document1.pdf": Sampling The Hawaii Deep-Set Longline Fishery and Point Estimators of Bycatch;
2. "pifsc.bycatch.document2.pdf": Interval Estimation of Annual Bycatch In the Hawaii Deep-Set Longline Fishery; and
3. "pifsc.bycatch.document3.pdf": Domain Estimators for the Total Number of Cetacean Bycatch Events Resulting In a Dead or Serious Injury Classification.

These documents indicate that the review involves five components: (1) sampling design; (2) point estimation of total bycatch for each species; (3) uncertainty evaluation; (4) estimation of the total number of bycatch events classified as "dead or seriously injury" (DSI) among the total bycatch for each marine mammal species; and (5) estimation of DSI within sub-geographical areas.

### **Prior to the Meeting**

I received the 3 documents approximately 2 weeks before the in-person review meeting. All three were reviewed in anticipation of the panel meeting.

### **In-Person Meeting**

The review was held at the University of Hawaii – Manoa in Honolulu, HI from August 24 to 28, 2015. A description of the activities during the meeting follows.

### **Monday, August 24, 2015**

Chris Boggs, NMFS Pacific Islands Fisheries Science Center, gave an introduction and provided background information about the Hawaii Longline Fishery. He also provided a summary of the Terms of Reference for the review.

Joe Arceneaux, NMFS PIRO Observer Program, presented background information on the Hawaiian Longline Observer Program, including training and coverage rates. The program targets 20% observer coverage each year for the deep-set longline fleet (DSLL) and observes all trips in the shallow set longline industry (SSLL).

Marti McCracken, then gave a detailed presentation on sampling design. Extensive discussion, questions and answers took place during and after the presentation. Some additional materials, including examples of sample data in spreadsheet, were provided to the review panel.

### **Tuesday, August 25, 2015**

The second day focused on presentation and review of the point estimators and several alternative approaches to interval estimation. Again, there were extensive discussion, questions and answers between the presenter and the review panel. Graphic examples were provided for some species.

### **Wednesday, August 26, 2015**

Presentation and review of bootstrapping methods for variance and interval estimators continued. This was followed by presentation on domain estimators for cetacean bycatch. Some examples in spreadsheet were provided to the panel. As in the previous two days, the discussion, questions and answers lasted a full day.

### **Thursday, August 27, 2015**

Today was set for panel discussion and writing. As a group, we discussed the presentation materials and our first impressions of the approaches. We discussed pros and cons of the methodology used for the Hawaii DSLL fishery bycatch.

### **Friday, August 28, 2015**

The panel individually presented their impressions of the sampling and inference methods to the NMFS Project team. Recommendations for future work were also briefly discussed. The panel had the last opportunity to ask questions and clarify some issues in the documents and in the presentations over the last several days.

Overall, the review meeting was well organized and ran smoothly. The project scientist (Marti McCracken) was very helpful in clarifying questions during the meeting. The amount of work and its complexity are quite impressive.

### **After the In-Person Meeting**

All notes were reviewed again in order to develop this report. In addition, some simulations were used to verify approaches in the reports or to compare alternatives to the present methodology.

## Summary of Findings

The main findings of the methodology used for sampling and inference in the Observer program are presented in this section. Each of the terms of reference (ToR) is addressed separately.

## ToR 1.

Review the sampling design used to select trips for observer placement and determine if it is a preferred design for estimating bycatch considering constraints and reporting requirements.

The key objective of the sampling design is to estimate the total annual bycatch for each species in a relatively short time maintaining as much as possible a 20% coverage rate. Coverage rate refers to the proportion of total trips in a calendar year in the DSLL fleet that have observers placed on board. This makes designing a sampling strategy challenging because the total number of fishing trips and their distribution over time are unknown prior to selecting trips for observation and, in addition, the availability of observers is uneven throughout the year (often due to influxes immediately after training or around major holidays).

Currently, captains/owners provide notification of planned trips and whether they are shallow set or deep set at least 72 hours before the intended date of leaving. As a result, the sampling frame of trips evolves as the calendar year progresses. It is this growing list of notifications from which trips are selected to have an observer on-board.

The current sampling design can be described as two-stage design in the sense that a multi-start systematic sample of notifications is selected first. Once this set of trips has been selected, a second stage adaptive component is used to select additional trips for observer placement. The number of additional trips selections depends strongly on observer availability and trip notifications. Although not directly controlled by NOAA/NMFS, the selection of trips for the additional observers is purportedly done randomly from among the “recent trip notifications”, resulting in a random sample from those recent trip notifications not pre-assigned to the systematic sample.

The systematic component is currently set to provide approximately 15% of the annual coverage and the adaptive component covers the remaining 5%. On occasion, the design has been modified to create a stratified scheme for the systematic component when the coverage rate must be temporarily modified. This is accomplished by drawing a new set of systematic samples at the changed coverage rate for a particular time frame within the calendar year. The design is creative and adaptive given the constraints in observer availability and the lack of prior knowledge of trip availability.

Several requirements that need to be met are not under review here but do impact the design and the inferences derived from the data collected. They include:

- a) The coverage rate of 20% is mandated;
- b) The program does not focus on specific species but covers all possible bycatch species. This results in a growing list to be reported in the sense that species not seen previously may be observed and so are now reported and other species which have been observed in the past are still being reported. As a result, “rare<sup>1</sup> species” are being added to the list in some years; and,
- c) The uneven availability of observers throughout the year.

Overall, the sampling design meets its requirements and the needs of the program. There are a few issues that could be addressed. First is the *ad hoc* stratification of the systematic samples. This could be retooled by looking at the historical behaviour of observer availability. For example, if early in the

---

<sup>1</sup> I use “rare” here in the sense that the number of observations is low, either because the species was not previously observed or because the species is no longer being observed in high numbers. These are in addition to the truly rare species, such as some of the protected species like marine mammals and sea turtles, which are only rarely observed.

year (say the first 2 months, as an example) generally has insufficient observers to cover 15% under the systematic sample, then perchance the design should be re-tooled to regularly reflect that, rather than being reactive to the current status of observers. Should additional observers become on occasion available, then the adaptive component can absorb those extra observers. Second is the decision to use 5 systematic starts – no information was given as to whether this is optimal for estimating variability of the estimator.

Third, is whether a classical two-stage cluster sampling design could be used instead of the current design. Here, more systematic samples are identified at the start of the year but not every trip notification selected by that design necessarily would get an observer. Instead, the assignment of an observer to one of the available systemically chosen trip notifications would depend on observer availability. Observer-covered trips would have to be treated as randomly chosen secondary sampling units (SSUs) within the primary systematic unit (PSU). The adequacy of this approach relies on the assumptions that: 1) the observers are more or less available in a pattern that would still provide reasonable coverage throughout the year, and 2) observers are assigned at random to one of any of the available SSUs at the time they present for work. The advantage of this approach is that the inclusion probabilities used in the estimators are easily calculated (under the assumptions of simple random sampling (SRS) at each stage) and there exists much theory on the estimators for 2-stage cluster sampling (Cochran, 1987. *Sampling*. Wiley & Sons; Thompson, 1992, *Sampling*, 2<sup>nd</sup> ed. Wiley & Sons).

Recommendations:

- a) Review the historical data on observer assignments with respect to finding general trends in observer availability throughout the year. This would be used to determine if fixed temporal strata where the coverage of the systematic sample would differ could reduce the need for 5% coverage by the adaptive component and still ensure maximal use of observers.
- b) Review whether the choice of 5 systematic starts is optimal with respect to estimating variability of the chosen estimators as well as capturing the observer variability.
- c) Consider changing the sampling design to a two-stage cluster sampling approach in which  $n$  out of  $N$  systematic starts are selected at the start of the year, and observers are randomly assigned to one of the systematically chose trip notifications available at the time the observer presents for work.

ToR 2.

Evaluate the point estimators and determine if they are good estimators given the sample design, observed frequency distribution of bycatch events, and constraints.

There are two points to be covered under this term of reference. The first is the choice of estimator, and the second is calculation of the inclusion probabilities used in the chosen estimator. We start by reviewing the estimators and then discuss the calculation of the inclusion probabilities.

#### *Point Estimators*

The methodology described in the first document that was provided included two general approaches to estimation of the total bycatch of a species: the Horvitz-Thompson estimator (HTE)

and the generalized ratio estimator (GRE) (Thompson, 1992). In addition, it also suggests that the GRE could be based on one of three alternative metrics for effort: number of trips, number of sets, and number of hooks. Hence, there are a total of four different estimators for the species-specific total bycatch. These simple estimators can be calculated quickly using canned software (see for example the sampling, sampling book, and survey packages available in R). The HTE is appropriate for the sampling design given the unequal probabilities of selection and inclusion in the sample; the GREs may or may not be appropriate – the inclusion of an effort variable depends strongly on the correlation of that metric with the count data. An additional minor issue could arise when choosing the appropriate estimator since the GREs are biased in general, but the bias decreases as the sample size increases. I suggest that the current sample size for the DSL observer program is sufficiently large that the bias is minimal and so this is not an issue.

The disadvantage of calculating four estimates of the total bycatch for each bycatch species is that only one estimate is required for any given species and so several estimates should not be calculated. Instead, a single estimator should be chosen for a species (or better yet, group of species<sup>2</sup>), and this estimator should be the one regularly reported. If all estimators are calculated every year and the analyst then picks the “appropriate” one to be reported, the program runs the risk of introducing subjectivity to the reported findings. It does not appear that this is an issue based on discussions at the panel review but the provided documents do not make it clear that such subjective decision-making does not occur.

The recommendation that a single estimator be chosen for each species or group of species begs a new question. How should the estimator be chosen for an individual species? As was discussed at the meeting it appears that the analyst uses past history of which estimator has the lowest standard error or least bias or something similar. For the GREs, it makes more sense to regress the bycatch counts on the measures of effort to see which, if any, of the metrics are actually informative. If one of the metrics is clearly linearly correlated with the counts, then that is the metric of effort that should be used. If none are clearly informative, then the HTE is the appropriate estimator. This can be done for each species or group of species using one or several years of data, depending on whether sufficient data are available in a particular year. Further, once an estimator is chosen it should be used consistently unless the relationship between counts and effort changes in time, such as might occur if the regulations change (e.g. changes to hook size or shape).

The current methods use a design-based estimator for reporting annual total bycatch. An obvious alternative would be to consider a model-based approach. This could take a simple form such as assuming a particular probability distribution for the data generating process so that the counts are observations from that distribution. In this instance, the estimator, its variability, and confidence interval estimation are available from theoretical work (see for example, the Poisson distribution in Johnson, Kemp, and Kotz, 2005. *Univariate Distributions*, 3<sup>rd</sup> ed. Wiley-Interscience). Offsets for informative effort metrics can be included in these models as well. A more complicated model could also be considered, for example a Bayesian hierarchical approach where the parameters of the data generating distribution are themselves random, i.e. probabilistically determined.

The disadvantage of the model-based approach is that preliminary work is required to determine its adequacy. For example, the choice of the data generating distribution would need to be carefully vetted and likely based on several years of data; the choice of prior distributions in the Bayesian

---

<sup>2</sup> Species could be grouped by their statistical properties such as the shape of the frequency distribution of sample counts, their mean count or standard deviation, modality in the frequency distribution, their correlation with effort metrics, etc.



approach would require expert knowledge to populate the parameters of the prior distributions in addition to choosing the appropriate prior probability distributions.

If a model-based approach were to be considered, it would require that data that were collected in an unequal probability scheme be treated as though they were randomly sampled or the model would have to include either random effects to account for the sampling design or the use of a weighted modelling approach. Assuming a random sample is probably not unreasonable for the data collected in the systematic samples, but it is unclear whether it would be appropriate for the adaptively collected data in the current design. If the sampling design were changed to a two-stage cluster sampling design as mentioned under ToR 1, the model-based approach assuming iid sampling is probably not unreasonable.

One final note: if model-based approaches are considered, the measures of variability of the estimators are not equivalent in interpretation to the measures associated with design-based estimators (*cf.* Gregoire, 1998. Can. J. For. Res. 28: 1429–1447). Hence, decisions on optimal choices for estimators in these cases must rely on other information such as adequacy of the model assumptions and not on comparisons of the standard errors or similar measures.

Recommendations:

- a) A single estimator should be regularly used for each species. As is currently done, the choice of estimator should be based on historical information regarding the validity of the underlying assumptions of the estimator, if any. There should not be an annual re-assessment of the appropriate estimator as is implied in pifsc.bycatch.document1.pdf (pg. 15).
- b) If several estimators are valid, then the choice should be based on the associated standard errors or confidence interval widths in the case of design-based estimators and on model diagnostics for model-based estimators.

### *Inclusion Probabilities*

Systematic Component

For the current sampling strategy, the description in the reading material are correct for the inclusion probability of a single observation. An alternative description of the method used to derive these probabilities can be given more simply though. We first start by switching emphases from the individual trips (SSUs) to the 1-stage systematic samples (PSUs). For this, let the total observed count in the  $i^{th}$  PSU be denoted  $y_i = \sum_{j=1}^{M_i} y_{ij}$  where  $M_i$  is the number of SSUs in the  $i^{th}$  PSU and  $y_{ij}$  is the observed count for the  $j^{th}$  SSU in the  $i^{th}$  PSU. This is possible because the design is a 1-stage cluster sampling design with  $m$  clusters sampled out of  $M$  available clusters. Since the PSUs are selected by simple random sampling without replacement (SRSWOR), the inclusion probabilities ( $\pi_i$  in the documentation provided) for any PSU is the same for all PSUs and is given by

$$\pi_i = \frac{m}{M}.$$

The joint inclusion probability for every pair of PSUs, i.e. the probability that PSUs  $i$  and  $j$  are both included in the sample, is:

$$\pi_{ij} = \frac{m(m-1)}{M(M-1)}.$$

Appendix B of pifsc.bycatch.document2.pdf lists the pairwise inclusion probabilities for different combinations of systematic and adaptive component trip selections. The first one listed states that it is for the case when both notifications belong to the same systematic cluster. It appears to be incorrect given that the secondary sample (the PLUS component) does not influence the selection of systematic samples and so the summands that involve the secondary sample should not be included here.

#### Adaptive Component

The current method of selecting trip notifications to receive an observer in the adaptive component is described as “Contractors select the notifications. Instructions are to select the notifications using SRSWOR from the **recent notifications** [my emphasis] still available for observer placement.” (slide 16 of cie.day1.m08d23y15.pptx). Given that the trip notification must occur at least 72 hours before the planned departure and that the vessel representative must be notified within 24 hrs of departure that an observer will be placed on the trip, this implies that only a small set of trips are available to be randomly selected for observation. Essentially, it is those trips that have notified the contractor and are planning on going out more than 24 hrs from the time an observer becomes available. Let this set be denoted with the subscript  $g$  in the following.

As a result, the probability that a trip is selected is a function of the timing of the trip notification and planned departure and availability of an observer. So, given that  $n_g$  observers are available at a particular point in time and that the exact set of trips to which they could be assigned is of size  $N_g$  and that the systematic sample has already been chosen, the conditional probability that a trip in the set of  $N_g$  available trips is included in a sample of  $n_g$  observations given it is not in the systematic sample is:

$$\pi_{i,g} = n_g/N_g.$$

This is not the probability that a trip is selected to be in the annual adaptive sample (with 5% coverage) given that it was not included in the systematic sample nor is it the unconditional probability including the systematic sample. It represents the inclusion probabilities for individual small random samples taken independently throughout the year, each time from a different set of sampling units available and given that the systematic sample was taken. That is, each observer placement is a single observation from its own sampling frame (except in cases where more than 1 observer is placed on the same day). The difficulty with this approach is that the joint inclusion probabilities can be somewhat difficult to calculate since these sampling frames can overlap when observers become available close in time (such as adjacent days), but calculation is not impossible given that when the data analysis occurs, the ordering and timing of every trip notification during the calendar year are known.

The current method of approximating the inclusion probability for an adaptively sampled trip notification is to combine the small sets of sampling units (the  $N_g$  available to be selected and for which a sampling frame actually exists) into larger sets of units by clustering contiguous groups of trips and redefining both the sampling frame from which trips can be selected and the number of observers available to be placed on those trips. The method by which this is done is complex and involves several steps, including combining trips between systematic sampling units into “blocks”, then clustering those blocks into strata. This has the effect of potentially changing the true inclusion probabilities. For example, consider the first set of values used to demonstrate the method in the excel file “sample notification logs m08d23y15.xlsx”:

block	strata	# day samples
1	1	2
2	1	4
3	2	0
4	3	4
5	4	0
6	4	0
7	5	2
8	5	2
9	5	6
10	5	2

Each block constitutes 50 trip notifications, so stratum 1 contains 100 notifications. If the four sampled trips in block 2 are combined with the 2 trips in block 1 denoted as stratum 1 and treated as though they are part of a SRSWOR from 100 notifications, then the inclusion probability for these 6 samples is approximated as  $\pi_g = 6/100 = 0.06$  where  $g$  refers to the set of 100 notifications. Compare this to a more likely event where for each “day sample” there are a variable number of trips available for sampling, say they range from 7 to 25. In that case, the day samples could have inclusion probabilities ranging from  $1/25$  (only 1 selection from 25 available) to  $6/7$  (6 selections all from the same set of 7 sampling units) depending on the timing of the notifications and observers. As a result, the current method assigns incorrect weights to the samples in the adaptive component. This has the effect of using wrong expansion factors for the counts when estimating total bycatch; this in turn likely leads to biased estimates of the total annual bycatch and its variability.

Recommendations:

- A simulations study should be run to determine the effect of clustering of adjacent “blocks” on both the bias and variance of the estimators.
- If the precision and accuracy are impacted significantly by the current method, the approach described above that uses knowledge of the actual trip notifications available for selection should be considered as an alternative.
- Rather than use the SYSPLUS sampling design, the simpler approach that is based on a two-stage cluster sample, and the assumption that the observer availabilities are random, provides much easier calculations for the inclusion probabilities. In fact, since estimators based on 2-stage cluster sampling have been developed (Thompson, 2002), there is no need to calculate inclusion probabilities.

ToR 3.

Evaluate the interval estimators and determine if they are good estimators given the sample design, observed frequency distribution of bycatch events, and constraints.

Once again, the analyst provided several approaches to obtaining variance estimates for the HTE and for the interval estimators. I recommend that a single estimator be used consistently and it should not be the Horvitz-Thompson estimator of variance as that estimator has a positive probability of providing a negative variance estimate for some cases.

Four alternative methods for interval estimation have been used in the DSLL bycatch estimation project: (1) a method based on the finite population central limit theorem; (2) nonparametric bootstrapping; (3) a model-based method; and (4) a Bayesian approach. Similar to issues with the point estimators, the major problem is the lack of clear rules on which method(s) are to be applied to which species for inclusion in the annual reports. It is again recommended that a single approach be based on historical information and that it be used in the future unless some significant change occurs in either the fishery or in the abundance of the species.

An overall issue with the interval estimator approaches is the lack of consistency between the chosen estimation procedure and the chosen interval estimation procedure. If a model-based method is desirable for interval estimation, then the estimator of total bycatch and the estimator of its variance should also be model-based. The current inconsistency could lead to situations in which the estimate of total bycatch is not even included in the interval estimator of total bycatch or the design-based estimator of variance is very different than the one based on the model-based approach. As a result, it is recommended that a single approach be chosen for each species, and that it be used for all inferences associated with that species.

Following are a few comments about each of the alternative approaches that were given in pifsc.bycatch.document2.pdf and the presentation cie.day2.m08d25y15.pptx.

#### *Large Sample Normal Approximation*

This approach is valid if it has been determined that the estimator of total bycatch is approximately normally distributed. This can probably only be determined by either combining several years of data or bootstrapping to show that the bootstrapped distribution of the estimator is approximately normal.

#### *Nonparametric bootstrap sampling approximate CI*

The current approach to bootstrapping and interval estimation is much more complicated than it need be. I do not believe that there is a need for the super-population approach of Davison and Hinkley (1977. *Bootstrap Methods and their Application*, Cambridge University Press). An alternative that has been shown to be better (Presnell & Booth, 1994. *Resampling Methods for Sample Surveys*, Tech Rpt 470, Univ. Florida) is the method of Booth, Butler & Hall (1994. *Bootstrap methods for finite populations*. JASA 89, 1282-1289).

One of the reasons for the current complexity in the bootstrapping is that of the method in which the adaptive component data are used. For the systematic samples, one could argue that the observed samples represent the unobserved data and so replicating them by their weight creates a representative pseudo-population which does not require recourse to the adaptively sampled dataset. Since the systematic sample typically covers approximately 15% of the total population, this would not be an unreasonable approach. Conversely, the 5% of the population covered by the adaptive component is completely ignored and so if the adaptive sample is somehow different<sup>3</sup> from the systematic samples, the estimated values and the associated variances from the bootstrapping would not be appropriate.

---

<sup>3</sup> Different here does not refer to random variability, but instead to some systematic differences such as where the trips fish or the gear used or something that makes the adaptive samples distinct in their representation of a part of the fishery as a whole that is lacking in the systematic sample.

To determine how to include the adaptive component into the pseudo-population, the current method appears to replicate the systematic samples and then replace some subset of the SSUs with adaptive sample values. The size of the subset appears to be a random variable that is generated with each “super-population”. The effect of this approach on the bootstrapping results is unclear. Further, the replacement appears to occur before the systematic samples are selected and so none of the bootstrapped systematic samples are the same as the original systematic samples. This is probably not an issue based on the current method, but does contribute to the complexity of the approach.

Another issue with the bootstrapping method is the choice of confidence interval estimator currently used (Hall’s or Efron’s percentile method). These methods are valid for bootstrapped distributions of estimators that are well-behaved, more-or-less unimodal and approximately symmetric. For many species where the data are highly skewed or the variance of the data is related to the mean (as is very common for count data in fisheries), a better alternative for interval estimation would be to use the double or nested bootstrap approach (cf. Gregoire & Shabenberger, 1999. Sampling-Skewed Biological Populations: Behavior of Confidence Intervals for the Population Total. *Ecology*, 80(3), 1056-1065; Scholz, 2007. The Bootstrap Small Sample Properties. Tech Rpt. bcstech-93-051, Boeing Computer Services, Research and Technology).

A possible alternative construction of the pseudo-population that would be simpler to implement would be to create a “population” using all the observed data as follows. It starts by assuming that there are temporal variations that must be captured in the pseudo-population (the analyst mentions in several places that the systematic samples are used partly to ensure that the entire annual temporal variation is captured in the samples). It also assumes that there is no intra-class correlation among observations within a systematic cluster. If these assumptions are appropriate, then first divide the year into intervals within which any temporal variation is low (such as if there is little change in time in observed bycatch counts). Then, construct the pseudo-population within that time period by replicating all SSU observations (SYS and PLUS samples) according to their weights and assigning them at random to the trip notifications within the same time period. This creates a pseudo-population for bootstrapping. Bootstrap sampling from this pseudo-population should then follow the actual sampling design where the number of trip notifications within a particular time period, the number of available observers, etc., are kept in the same pattern as was actually observed in the year being analysed. In other words, the ordering and timing of the SSUs and observers are maintained but the values (bycatch and effort, if appropriate) associated with each SSU are varied to create the pseudo-population.

Recommendations:

- a) Redesign the approach used to construct the pseudo-population for bootstrapping.
- b) Use the double or nested bootstrapping approach for interval estimation for those species where the mean and variance of the data are related or where the data are highly skewed.

#### *Model-based interval estimates*

If a model-based approach is desired for estimation, such as might be appropriate for species with small total bycatch, then model-based estimation of variance and of interval estimators should also be used. As was mentioned earlier, diagnostics should be performed to verify that the chosen distribution is reasonable for the data. Further, if the sampling weights vary among sampling units, then these need to be addressed in the estimation of the model parameters.

The particular model-based approach described for rare bycatch species is an uncommon choice for estimation and seems unnecessarily complicated. An alternative that is also based on the assumption that each observed trip's bycatch is an observation from a Poisson distribution with parameter  $\lambda$  ( $Y \sim Poi(\lambda)$ ). Total annual bycatch then is  $\tau = N\lambda$  where  $N$  is the total number of trip notifications during the year and is known. To estimate  $\tau$ , the MVUE estimator for  $\lambda$  and the associated confidence interval can be easily calculated (John, Kemp, & Kotz, 2005) and then multiplied by  $N$ . This does not require assuming that  $\tau$ , the true total bycatch in a given year, is itself randomly distributed as is done in the report.

Recommendations:

- a) Consider use of a simpler approach to estimating the total annual bycatch by assuming that the total is a fixed but unobserved value which can be estimated from the sample rather than assuming that it is also random.
- b) Consider alternative models for the distributions of the trip-level counts (see the recommendations under ToR 6 for examples).

#### *Bayesian interval estimates*

This section contains several alternative approaches for Bayesian modelling in order to obtain confidence intervals for total annual bycatch. This may be acceptable if in fact they are for different groups of species that demonstrate distinct behaviours with respect to the distribution of the data, but I am not sure the Bayesian approach is required.

Recommendation:

- a) Bayesian models should be considered only for those species for which non-zero data are extremely rare. In that case, consider approaches in which the priors are based on information from several previous years under the assumption that the fishery and species have not significantly changed between years.

ToR 4.

Evaluate estimators of total bycatch events resulting in a death or serious injury (DSI) classification and determine if they are good estimators given the sample design, observed frequency distribution of injury classifications (non-serious or DSI), and constraints.

Estimation of total bycatch events resulting in DSI is required for marine mammals. Two methods have been used:

- (1) Simply expanding the observed DSI events to all trips that have an observer aboard, which is similar to estimating  $\tau$  in document #1.
- (2) First computing the probability that a bycatch event results in a DSI using 2002-2010 data and then applying this probability to the estimated total bycatch.

Both approaches have advantages and disadvantages as discussed in [pifsc.bycatch.document3.pdf](#). There is no major issue with these estimators except that (1) does not alleviate the issues associated with interval estimation of rare events. The problems with both of these approaches are the same as those mentioned earlier – confidence interval estimation should follow the approach taken for the

estimator. In other words, if the estimators are design-based, (1) above, then the interval estimator should also be design-based, which essentially requires bootstrapping. Hence (1) should be used when DSI is considered to not be rare.

Recommendations:

- a) Use method (1) only when the number of DSI are likely to be non-rare.
- b) For rare DSI events, the model-based approaches mentioned earlier should be used. In fact, the random nature of a DSI occurring given that an animal has been bycaught implies that a hierarchical Bayesian model should be considered.

ToR 5.

Evaluate the subpopulation estimators being applied to estimate bycatch within a political geographical boundary and determine if they are good estimators given the sample design, reporting requirements under the MMPA, and constraints.

The design-based HTE and GRE described in document #1 are used for subpopulation estimators within each geographical areas. The problem here is that the observed bycatch events are already rare for many species in the whole DSLL fishery. Breaking estimation of rare events down into multiple areas further reduces the frequency of these events at the domain level. Therefore, model-based approaches that treat region as a covariate and possibly use several years of data might be appropriate.

Recommendation:

- a) If the bycatch is rare, model-based methods should be used in which the mean (or total) within a region is a function of the region, i.e. use a generalized linear type of model.
- b) If the bycatch is not rare, then design-based post-stratification methods can be applied to obtain domain estimates.

ToR 6.

Suggest future research priorities to improve methods for estimating bycatch with increased efficiency given the current data structure. Suggest future research priorities for improving the sampling design for the purposes of estimating bycatch, with efficient use of sampling resources as a consideration.

- a) Since there are several years of data available now, simulations should be used to determine which estimation approaches are in fact optimal and efficient. This would start by modelling several years of bycatch data with predictors such as gear type (to account for changes in the regulations of the fishery) or year using a generalized linear mixed model. The model coefficients could then be used to simulate Monte Carlo (MC) populations whose characteristics mimic the samples taken each year. These MC pseudo-populations could then be used to compare estimators more explicitly rather than rely on a single year's data. Since the data from a single year is a random quantity, the conclusions one draws for one

year may not be appropriate for other years. Hence, if confidence in an approach to be used consistently from year to year is desired, comparisons of methods should also be based on multiple years of data. This has been done to some extent by comparing the estimation results among years, but is still somewhat ad hoc and requires that all methods be applied every year. Instead a single intensive study should be undertaken using a subset of species that are representative of the range of issues encountered during estimation.

- b) Consider a Bernoulli distribution for very rare species which are observed so rarely that generally a trip has observed 0 or 1 animal. In that case, a binary distribution would be appropriate. Then, the total bycatch in a given year could be estimated as the estimated probability of observing an animal in a trip times the total number of trips in a year. Other similar models could provide for more than one outcome in a trip such as a multinomial distribution for species counts that range over a small number of discrete values, for example, 0, 1, 2, or 3 times. One could then estimate the probabilities of each outcome and use those to estimate total bycatch. This removes the requirement that the data behave according to more defined distributional shapes such as the Poisson or Negative Binomial. I recommend that these also be considered in place of the more complex models described in [pifsc.bycatch.document2.pdf](#).
- c) If reducing variability is desired, constructing a more sophisticated modelling approach that includes say the spatial-temporal distribution of trips (or more likely sets) might provide insight into where and when particular species tend to get bycaught. Such information could be used to generate estimates of total bycatch based on predicting bycatch of trips which did not have an observer but where and when the trips occurred is known.

## Conclusions and Recommendations

Overall, the amount of effort expended to provide precise and accurate estimates of species-specific bycatch is prodigious. The level of effort could be reduced somewhat by using only a few approaches, one for each group of species exhibiting particular statistical properties, rather than many approaches for many species. Several recommendations for refining the current work are provided. The intent of the recommendations is to increase the efficiency of the estimation procedures and to motivate research into better approaches with limited resources.



## Appendix 1: Bibliography of materials provided for review

Documents provided two weeks prior to the review meeting:

1. McCracken, M.L. 2015. Sampling the Hawaii deep-set longline fishery and point estimators of bycatch. Pacific Islands Fisheries Science Center. 17p.
2. McCracken, M.L. 2015. Interval estimation of annual bycatch in the Hawaii deep-set longline fishery. Pacific Islands Fisheries Science Center. 18p.
3. McCracken, M.L. 2015. Domain estimators for the total number of cetacean bycatch events resulting in a dead or serious injury classification. Pacific Islands Fisheries Science Center. 11p.

Materials provided during the review meeting:

4. Longline logbook protected species marine mammals and turtles.
5. NMFS Western Pacific Longline Fishing Log.
6. Regulation Summary: Hawaii Pelagic Longline Fishing.
7. Hawaii Longline Observer Program Field Manual, version LM.14.04.
8. Sample notification logs m08d23y15.xlsx.
9. Multiple figures as examples for bycatch frequency.
10. Examples of marine mammal bycatch in spreadsheets.

## Appendix 2: A copy of the CIE Statement of Work

### **External Independent Peer Review by the Center for Independent Experts**

#### **Methodology for Sampling and Estimating Bycatch of the Hawaii Deep-Set Longline Fishery**

##### **Scope of Work and CIE Process:**

The National Marine Fisheries Service's (NMFS) Office of Science and Technology coordinates and manages a contract providing external expertise through the Center for Independent Experts (CIE) to conduct independent peer reviews of NMFS scientific projects. The Statement of Work (SoW) described herein was established by the NMFS Project Contact and Contracting Officer's Technical Representative (COTR), and reviewed by CIE for compliance with their policy for providing independent expertise that can provide impartial and independent peer review without conflicts of interest. CIE reviewers are selected by the CIE Steering Committee and CIE Coordination Team to conduct the independent peer review of NMFS science in compliance the predetermined Terms of Reference (ToRs) of the peer review. Each CIE reviewer is contracted to deliver an independent peer review report to be approved by the CIE Steering Committee and the report is to be formatted with content requirements as specified in **Annex 1**. This SoW describes the work tasks and deliverables of the CIE reviewer for conducting an independent peer review of the following NMFS project. Further information on the CIE process can be obtained from [www.ciereviews.org](http://www.ciereviews.org).

##### **Project Description:**

Quantifying bycatch in the Hawaii deep-set longline fishery is required by the Magnuson-Stevens Fishery Conservation and Management Act (MSA), Endangered Species Act (ESA), Marine Mammal Protection Act (MMPA), and Migratory Bird Treaty Act (MBTA) and their implementing regulations. As over a hundred species, some of them listed as endangered or threatened, have been recorded as being caught in the Hawaii deep-set longline fishery, reliable bycatch estimates need to be computed in a relatively quick manner on a yearly basis. Since mid-year 2002, a unique complex sampling design has been used to select deep-set longline trips for observer placement. While aboard a selected longline trip, NMFS trained observers collect information on bycatch and ancillary variables for each longline fishing operation. Based on the sampling design, bycatch estimates are computed for all marine mammals, protected species, sharks, and fish that have been observed at least once in the fishery or are of special interest. What estimators are used depends on the observed frequency distribution of bycatch events for the species of interest. Interval estimators have been developed for commonly, seldom, and very rarely bycaught species. Methods for estimating bycatch within political geographical areas within the fishing grounds and the total number of marine mammal bycatch events resulting in a death or serious injury (DSI) have also been developed as the MMPA requires estimates of DSI within and outside the Economic Exclusive Zones (EEZ) of the United States.

These annual bycatch estimates of sea turtles, seabirds, and marine mammals are used to monitor takes within the deep-set longline fishery. These estimates have a large potential impact on endangered species and the valuable longline commercial fishery in Hawaii. Additionally, bycatch estimates of all species are provided for inclusion in the National Bycatch Report, seabird and sea turtle estimates are submitted annually to the IATTC (Inter-American Tropical Tuna Commission) per Resolution C-11-02 and C-04-05, and marine mammal, seabird, and sea turtle

estimates are provided for inclusion in the annual WCPFC (Western and Central Pacific Fisheries Commission) National report. The methods to be reviewed have not undergone independent peer review and there is a need to evaluate the methods to improve the scientific basis for management. The Terms of Reference (ToRs) of the peer review are attached in **Annex 2**. The tentative agenda of the panel review meeting is attached in **Annex 3**.

**Requirements for CIE Reviewers:**

Three CIE reviewers shall conduct an impartial and independent peer review in accordance with the SoW and ToRs herein. Reviewers shall have working knowledge and recent experience in the application of statistical inference for finite populations. Reviewers should be statisticians with comprehensive knowledge of both theoretical and applied sampling design and analysis. Furthermore, reviewers should have some knowledge of analyzing rare events, bootstrap techniques for finite population sampling, and frequentist and Bayesian inference for finite populations. Experience in statistics related to natural resources is beneficial.

Each CIE reviewer's duties shall not exceed a maximum of 14 days to complete all work tasks of the peer review described herein.

**Location of Peer Review:** Each CIE reviewer shall conduct an independent peer review during the panel review meeting scheduled in Honolulu, HI during August 24-28, 2015.

**Statement of Tasks:** Each CIE reviewers shall complete the following tasks in accordance with the SoW and Schedule of Milestones and Deliverables herein.

Prior to the Peer Review: Upon completion of the CIE reviewer selection by the CIE Steering Committee, the CIE shall provide the CIE reviewer information (full name, title, affiliation, country, address, email) to the COTR, who forwards this information to the NMFS Project Contact no later the date specified in the Schedule of Milestones and Deliverables. The CIE is responsible for providing the SoW and ToRs to the CIE reviewers. The NMFS Project Contact is responsible for providing the CIE reviewers with the background documents, reports, foreign national security clearance, and other information concerning pertinent meeting arrangements. The NMFS Project Contact is also responsible for providing the Chair a copy of the SoW in advance of the panel review meeting. Any changes to the SoW or ToRs must be made through the COTR prior to the commencement of the peer review.

Foreign National Security Clearance: When CIE reviewers participate during a panel review meeting at a government facility, the NMFS Project Contact is responsible for obtaining the Foreign National Security Clearance approval for CIE reviewers who are non-US citizens. For this reason, the CIE reviewers shall provide requested information (e.g., first and last name, contact information, gender, birth date, passport number, country of passport, travel dates, country of citizenship, country of current residence, and home country) to the NMFS Project Contact for the purpose of their security clearance, and this information shall be submitted at least 30 days before the peer review in accordance with the NOAA Deemed Export Technology Control Program NAO 207-12 regulations available at the Deemed Exports NAO website: <http://deemedexports.noaa.gov/>

[http://deemedexports.noaa.gov/compliance\\_access\\_control\\_procedures/noaa-foreign-national-registration-system.html](http://deemedexports.noaa.gov/compliance_access_control_procedures/noaa-foreign-national-registration-system.html)

Pre-review Background Documents: Two weeks before the peer review, the NMFS Project Contact will send (by electronic mail or make available at an FTP site) to the CIE reviewers the necessary background information and reports for the peer review. In the case where the

documents need to be mailed, the NMFS Project Contact will consult with the CIE Lead Coordinator on where to send documents. CIE reviewers are responsible only for the pre-review documents that are delivered to the reviewer in accordance to the SoW scheduled deadlines specified herein. The CIE reviewers shall read all documents in preparation for the peer review.

Documents will describe:

- The stratified systematic-plus design and approximation of inclusion probabilities.
- Point estimators of total bycatch.
- Interval estimators of total bycatch, including estimators for very rarely bycaught species.
- Estimators for subpopulation totals, specifically estimators of bycatch within geographical areas of the fishing grounds.
- Estimators of total number of marine mammal bycatch events resulting in a classification of dead or serious injury (DSI).

Panel Review Meeting: Each CIE reviewer shall conduct the independent peer review in accordance with the SoW and ToRs, and shall not serve in any other role unless specified herein. **Modifications to the SoW and ToRs can not be made during the peer review, and any SoW or ToRs modifications prior to the peer review shall be approved by the COTR and CIE Lead Coordinator.** Each CIE reviewer shall actively participate in a professional and respectful manner as a member of the meeting review panel, and their peer review tasks shall be focused on the ToRs as specified herein. The NMFS Project Contact is responsible for any facility arrangements (e.g., conference room for panel review meetings or teleconference arrangements). The NMFS Project Contact is responsible for ensuring that the reviewers understand the contractual role of the CIE reviewers as specified herein. The CIE Lead Coordinator can contact the Project Contact to confirm any peer review arrangements, including the meeting facility arrangements.

Contract Deliverables - Independent CIE Peer Review Reports: Each CIE reviewer shall complete an independent peer review report in accordance with the SoW. Each CIE reviewer shall complete the independent peer review according to required format and content as described in Annex 1. Each CIE reviewer shall complete the independent peer review addressing each ToR as described in Annex 2.

### **Specific Tasks for CIE Reviewers:**

The following chronological list of tasks shall be completed by each CIE reviewer in a timely manner as specified in the **Schedule of Milestones and Deliverables**.

- 1) Conduct necessary pre-review preparations, including the review of background material and reports provided by the NMFS Project Contact in advance of the peer review.
- 2) Participate during the panel review meeting in Honolulu, HI, from August 24-28, 2015.
- 3) Conduct an independent peer review in accordance with the ToRs (**Annex 2**).
- 4) No later than September 14, 2015, each CIE reviewer shall submit an independent peer review report addressed to the "Center for Independent Experts," and sent to Dr. Manoj Shivlani, CIE Lead Coordinator, via email to [mshivlani@ntvifederal.net](mailto:mshivlani@ntvifederal.net), and Dr. David Die, CIE Regional Coordinator, via email. [ddie@rsmas.miami.edu](mailto:ddie@rsmas.miami.edu). Each CIE report shall be written using the format and content requirements specified in Annex 1, and address each ToR in **Annex 2**.

**Schedule of Milestones and Deliverables:** CIE shall complete the tasks and deliverables described in this SoW in accordance with the following schedule.

<i>July 20, 2015</i>	CIE sends reviewer contact information to the COTR, who then sends this to the NMFS Project Contact
<i>August 10, 2015</i>	NMFS Project Contact sends the CIE Reviewers the pre-review documents
<i>August 24-28, 2015</i>	Each reviewer participates and conducts an independent peer review during the panel review meeting
<i>September 14, 2015</i>	CIE reviewers submit draft CIE independent peer review reports to the CIE Lead Coordinator and CIE Regional Coordinator
<i>October 2, 2015</i>	CIE submits CIE independent peer review reports to the COTR
<i>October 5, 2015</i>	The COTR distributes the final CIE reports to the NMFS Project Contact and regional Center Director

**Modifications to the Statement of Work:** This ‘Time and Materials’ task order may require an update or modification due to possible changes to the terms of reference or schedule of milestones resulting from the fishery management decision process of the NOAA Leadership, Fishery Management Council, and Council’s SSC advisory committee. A request to modify this SoW must be approved by the Contracting Officer at least 15 working days prior to making any permanent changes. The Contracting Officer will notify the COTR within 10 working days after receipt of all required information of the decision on changes. The COTR can approve changes to the milestone dates, list of pre-review documents, and ToRs within the SoW as long as the role and ability of the CIE reviewers to complete the deliverable in accordance with the SoW is not adversely impacted. The SoW and ToRs shall not be changed once the peer review has begun.

**Acceptance of Deliverables:** Upon review and acceptance of the CIE independent peer review reports by the CIE Lead Coordinator, Regional Coordinator, and Steering Committee, these reports shall be sent to the COTR for final approval as contract deliverables based on compliance with the SoW and ToRs. As specified in the Schedule of Milestones and Deliverables, the CIE shall send via e-mail the contract deliverables (CIE independent peer review reports) to the COTR (William Michaels, via William.Michaels@noaa.gov).

**Applicable Performance Standards:** The contract is successfully completed when the COTR provides final approval of the contract deliverables. The acceptance of the contract deliverables shall be based on three performance standards:

- (1) The CIE report shall be completed with the format and content in accordance with **Annex 1**,
- (2) The CIE report shall address each ToR as specified in **Annex 2**,
- (3) The CIE reports shall be delivered in a timely manner as specified in the schedule of milestones and deliverables.

**Distribution of Approved Deliverables:** Upon acceptance by the COTR, the CIE Lead Coordinator shall send via e-mail the final CIE reports in \*.PDF format to the COTR. The COTR will distribute the CIE reports to the NMFS Project Contact and Center Director.

**Support Personnel:**

Allen Shimada  
NMFS Office of Science and Technology  
1315 East West Hwy, SSMC3, F/ST4, Silver Spring, MD 20910  
Allen Shimada@noaa.gov Phone: 301-427-8174

Manoj Shivilani, CIE Lead Coordinator  
NTVI Communications, Inc.  
10600 SW 131st Court, Miami, FL 33186  
mshivilani@ntvifederal.com Phone: 305-968-7136

**Key Personnel:**

NMFS Project Contact:

Marti McCracken  
Mathematical Statistician  
Fisheries Research and Monitoring Division  
National Marine Fisheries Service  
Pacific Islands Fisheries Science Center  
1845 Wasp Boulevard, Bldg. #176  
Honolulu, Hawaii 96818  
marti.mccracken@noaa.gov Phone: (808) 725-5736

Annie Yau  
Annie.yau@noaa.gov  
Phone: (808) 725-5350

## Annex 1: Format and Contents of CIE Independent Peer Review Report

1. Each CIE independent peer review report shall be prefaced with an Executive Summary providing a concise summary of the findings and recommendations following Annex 2 Terms of Reference.
2. The main body of the reviewer report shall consist of a Background, Description of the Individual Reviewer's Role in the Review Activities, Summary of Findings for each ToR in which the weaknesses and strengths are described, and Conclusions and Recommendations in accordance with the ToRs.
  - a. Reviewers should describe in their own words the review activities completed during the panel review meeting, including providing a brief summary of findings, of the science, conclusions, and recommendations.
  - b. Reviewers should discuss their independent views on each ToR even if these were consistent with those of other panelists, and especially where there were divergent views.
  - c. Reviewers shall provide a critique of the NMFS review process, including suggestions for improvements of both process and products.
  - d. The CIE independent report shall be a stand-alone document for others to understand the weaknesses and strengths of the science reviewed. The CIE independent report shall be an independent peer review of each ToRs.
3. The reviewer report shall include the following appendices:

Appendix 1: Bibliography of materials provided for review

Appendix 2: A copy of the CIE Statement of Work

Appendix 3: Panel Membership or other pertinent information from the panel review meeting.

**Methodology for Sampling and Estimating Bycatch  
of the Hawaii Deep-Set Longline Fishery**

1. Review the sampling design used to select trips for observer placement and determine if it is a preferred design for estimating bycatch considering constraints and reporting requirements.
2. Evaluate the point estimators and determine if they are good estimators given the sample design, observed frequency distribution of bycatch events, and constraints.
3. Evaluate the interval estimators and determine if they are good estimators given the sample design, observed frequency distribution of bycatch events, and constraints.
4. Evaluate estimators of total bycatch events resulting in a death or serious injury (DSI) classification and determine if they are good estimators given the sample design, observed frequency distribution of injury classifications (non-serious or DSI), and constraints.
5. Evaluate the subpopulation estimators being applied to estimate bycatch within a political geographical boundary and determine if they are good estimators given the sample design, reporting requirements under the MMPA, and constraints.
6. Suggest future research priorities to improve methods for estimating bycatch with increased efficiency given the current data structure. Suggest future research priorities for improving the sampling design for the purposes of estimating bycatch, with efficient use of sampling resources as a consideration.

*Note – CIE reviewers typically address scientific subjects, hence ToRs usually do not involve CIE reviewers with regulatory and management issues unless this expertise is specifically requested in the SoW.*



Annex 3: Tentative Agenda

**Methodology for Sampling and Estimating Bycatch of the Hawaii Deep-Set Longline Fishery**

24-27 August: Honolulu Service Center, NOAA Fisheries Pier 38, Honolulu Harbor, 1139 N. Nimitz Hwy, Suite 220, Honolulu, HI 96817

28 August: NOAA Daniel K Inouye Regional Center, 1845 Wasp Boulevard, Building 176, Conference Room 2545, Honolulu, HI 96818

8:30am-5:00pm, 24-28 August 2015

Monday, August 24

1. Introduction
2. Background information - Objectives and Terms of Reference
3. Observer Program and Longline Fishery  
Observer program (presented by Pacific Islands Observer Program)  
Deep-Set Longline Fishery
4. Review of Sampling Design
5. Review of Approximation of Inclusion Probabilities

Tuesday, August 25

6. Review of Point Estimators of Bycatch
7. Review of Interval Estimators

Wednesday, August 26

8. Review of Estimators of DSI (marine mammals)
9. Review of Estimators of Subpopulation Totals. ,

Thursday, August 27

10. Panel discussions (Closed)

Friday, August 28

11. Panel discussions
12. Adjourn

## Appendix 3: Panel Membership

Mary C. Christman, MCC Statistical Consulting, USA

Yan Jiao, Virginia Polytechnic Institute and State University, USA

Shijie Zhou, CSIRO, Australia

Other participants:

Joe Arceneaux, NMFS PIRO Observer Program

Chris Boggs, NMFS Pacific Islands Fisheries Science Center

Asuka Ishizaki, Western Pacific Fishery Management Council

Jarad Makiau, NMFS PIRO Observer Program

Marti McCracken, NMFS Pacific Islands Fisheries Science Center

Ben Richards, NMFS Pacific Islands Fisheries Science Center