Sampling the Hawaii Deep-set Longline Fishery and Point Estimators of Bycatch

Marti L. McCracken

U.S. DEPARTMENT OF COMMERCE
National Oceanic and Atmospheric Administrator
National Marine Fisheries Service
Pacific Islands Fisheries Science Center

NOAA Technical Memorandum NMFS-PIFSC-89
https://doi.org/10.25923/2psa-7s55

August 2019

Sampling the Hawaii Deep-set Longline Fishery and Point Estimators of Bycatch

Marti L. McCracken

Pacific Islands Fisheries Science Center
National Marine Fisheries Service
National Marine Fisheries Service
1845 Wasp Boulevard
Honolulu, HI 96818

NOAA Technical Memorandum NMFS-PIFSC-89

August 2019

U.S. Department of Commerce
Wilbur L. Ross, Jr., Secretary

National Oceanic and Atmospheric Administrator
RDML Tim Gallaudet, Ph.D., USN Ret., Acting NOAA Administrator

National Marine Fisheries Service
Chris Oliver, Assistnat Administrator for Fisheries

Recommended citation:

McCracken,ML. 2019. Sampling the Hawaii deep-set longline fishery and point estimators of by catch. U.S. Dept. of Commerce, NOAA Technical Memorandum NOAA-TM-PIFSC-89, 22 p. doi:10.25923/2psa-7s55

Copies of this report are available from:

Science Operations Division Pacific Islands Fisheries Science Center National Marine Fisheries Service National Oceanic and Atmospheric Administration 1845 Wasp Boulevard, Building #176 Honolulu, Hawaii 96818

Or online at:

https://repository.library.noaa.gov

Table of Contents

1 Introduction

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), Migratory Bird Treaty Act (MBTA), and their implementing regulations. With more than 100 species (some of which are listed as endangered or threatened) recorded as being caught in the Hawaii deep-set longline fishery, reliable estimates of each species' total bycatch need to be computed in a relatively quick manner on a yearly basis. A probability sample and corresponding design-based estimators provide the framework for producing such estimates. Characteristics of the NOAA Fisheries Pacific Islands Regional Observer Program (PIROP), established to monitor bycatch, and the Hawaii deep-set longline fishery require a sampling design that can adapt to fluctuating observer coverage. This document is the first of a series of documents that describe the methods used to derive the required bycatch estimates. In this section, I provide a brief introduction to the fishery and PIROP, define bycatch, and give an overview of the problem. This document was written in 2015 and covers years 2002-2015.

1.1 Hawaii Longline Fisheries and PIROP

The Hawaii-permitted longline fishery is a limited-entry fishery with a maximum of 164 permits and is split into two management components: a shallow-set fishery (targeting primarily swordfish) and a deep-set fishery (targeting primarily tunas, most commonly bigeye tuna). In this document, a Hawaii longline fishing trip is defined as any commercial fishing trip by a vessel that fishes using a Hawaii longline permit. The shallow-set longline (SSLL) fishery consists of trips that are declared to PIROP by the vessel owner or operator, prior to departure, as a shallow-set trip. The SSLL fishery has 100% observer coverage, so no sampling or estimates are needed. The deep-set longline (DSLL) fishery consists of all other trips and must comply with the regulations for this fishery, including the requirement of observer placement on a sample of vessels. A permitted vessel may participate in both fisheries. A trip is considered to end when the vessel comes into port and the catch is landed. Although a DSLL trip may only last a few days, they typically last from 2 to 4 weeks. After each trip, the appropriate representative from the vessel is required to submit completed logbooks to NOAA National Marine Fisheries Service (NMFS) describing certain characteristics of the trip. The entries in these logbooks are entered into a database, referred to as the Hawaii longline logbook database (Pacific Islands Fisheries Science Center, 2017). In 2013, the DSLL fishery completed 1,328 DSLL trips, 18,750 fishing operations (sets), and hauled 46,769,514 hooks.

Although bigeye tuna are classified as highly migratory species, they don't appear to have well-defined patterns of movement. Consequently, the specific area that the fishery will be utilizing at any given time is basically unknown.

The federally mandated PIROP was officially established in 1994. During the first 6 years of the program, observers were placed on approximately 3%–5% of fishing trips by the fleet. In 2000 the Hawaii longline fishery was split into the two management components: deep-set and shallow-set. Since this split, an observer must be aboard monitoring bycatch on at least 20% of a year's DSLL trips. In 2013 the PIROP observed 272 DSLL trips. The shallow-set component of this fishery was basically closed as a result of court-ordered area closures in 1999 and remained closed until late in 2004. When the SSLL fishery opened, it was restricted to an annual limit of 2,120 shallow sets north of the equator and would close if hard caps on the number of observed leatherback and loggerhead sea turtle takes were exceeded. Although the amount of effort permitted and the hard limits have changed in the subsequent years, 100% coverage has been maintained in this fishery. In 2013 there were 51 SSLL trips.

During an observed trip, observers record a suite of variables concerning the trip, fishing operation, catch, bycatch, and characteristics of bycaught protected species and marine mammals.

This information is entered into a database called the Longline Observer Data System (LODS) (Pacific Islands Regional Office, 2017). The observer manual (Pacific Islands Regional Observer Program, 2014) provides information on the program and the variables recorded.

When deciding on a sampling design for selecting DSLL trips for observer placement, fishing regulations and PIROP policies need to be considered. The following are relevant fishery regulations:

(1) The vessel owner or operator must notify the observer program at least 72 hours prior to their intended departure date and declare the intended trip type (shallow-set or deep-set). Once a trip type has been declared, the operator must make sets only of that type.
(2) Within 48 hours after the notification of an intended trip, the appropriate representative of a vessel selected for observer placement must be notified by the observer program that they will have an observer aboard.
(3) When selected for observer placement, the vessel is required to carry an observer and follow the observer guidelines (Pacific Islands Regional Observer Program, 2014).

The following are relevant policies of the PIROP:

(1) Observers must successfully complete a PIROP observer training course prior to being placed aboard a longline vessel. These courses are offered when it is expected that the number of employed observers will soon fall below the number needed to maintain minimal coverage, currently 15% (NOAA, 2012*a*).
(2) Because observers are not paid while waiting to be deployed, they need to be deployed with minimal delay. The alternative of paying them while they are waiting to be deployed will increase the cost of the observer program or reduce the level of coverage.
(3) Because 100% coverage in the SSLL fishery is a requirement, observer placement on a SSLL trip has priority over placement on a DSLL trip.
(4) The hiring and placement of observers is fulfilled by a contractor.

Since these policies make it impractical to instantaneously adjust the number of observers on staff to brief and sometimes unpredictable changes in the volume of fishing or to hire new observers to immediately replace observers who leave the program, the level of observer coverage is expected to fluctuate. Between training courses, the coverage level will decline. This decline is usually not monotonic as fluctuations in the volume of fishing or observers returning from leave may increase the level of observer coverage temporarily.

1.2 Definition of Bycatch

The definition of bycatch depends on the species being analyzed. Within the MSA the term “bycatch" is defined as fish which are harvested in a fishery, but which are not sold or kept for personal use, and includes economic and regulatory discards (SEC 3(2)). The term “fish" means finfish, mollusks, crustaceans, and all other forms of marine animal and plant life other than marine mammals and birds (SEC 3(12)). Although sea turtles fall under the definition of “fish" under the MSA, the species of sea turtles caught in the Hawaii longline fisheries are listed as endangered and will be referred to as protected species. Herein, the bycatch of fish refers to the total number of events where a fish is hooked or entangled and not kept. Fish are considered kept if at least part of it is retained by the fishermen for sale, personal consumption, or any other use. A remora or any other fish that is attached to a caught animal is not considered bycatch.

Except as permitted by regulations, the taking of species covered under the MMPA, MBTA, and ESA is prohibited. There is a process within each of these statues that authorizes “incidental takes" in a commercial fishery. The MMPA, MBTA, and ESA define “take” in slightly different ways, but basically, “take” means to catch, kill, or harm a marine mammal or protected species in any way. An “incidental take” means a take that is incidental to, and not the purpose of, the carrying out of an otherwise lawful activity. Herein, the “bycatch” of a protected species or marine mammal refers to the total number of take events in which an animal is hooked or entangled in the longline gear. Under this definition, bycatch is a component of the total incidental take in the Hawaii DSLL fishery.

A couple of practical constraints on the definition of bycatch is used herein. First, observers are instructed to record all observed hooked or entangled animals during the haul-back operation of the longline gear (Pacific Islands Regional Observer Program, 2014). Animals observed hooked or entangled that are freed before being landed on deck are included in this definition. However, hooked or entangled animals that are removed (e.g., by predators) or freed (e.g., by escape or drop-off) from the longline prior to it becoming visible on the haul-back operation will not be observable and therefore cannot be recorded, unless warranted by convincing circumstantial evidence of their capture. These "missed" animals are not included in the bycatch estimates as there is no practical way to quantify them. Nor do the estimates include animals that are not hooked or entangled but are in some other unobserved way caught, killed or harmed by the activity of DSLL fishing. Such events are not included because it is not feasible to monitor all aspects of a trip. If an observer witnesses an interaction with a marine mammal or sea turtle that does not fall into the category of hooked or entangled, the interaction is reviewed by NMFS sea turtle or marine mammal scientists to determine if the interaction should be considered a bycatch event. There have only been two observed takes between 2002 and 2014 that were not categorized as hooked or entangled. One incident involved a marine mammal that was briefly restrained by the gear and non-seriously injured. This interaction was determined to be a bycatch event. The second incident involved a marine mammal that was believed to have bit the line and was not injured. This interaction was determined not to be a bycatch event.

Second, the estimates of bycatch refer to the total number of bycatch events, which may exceed the number of individual animals that are caught. It is possible for an animal to be observed caught, then freed or released, and then caught again during the same year. For example, a loggerhead sea turtle was observed to be caught twice during a shallow-set trip in 2012. These two events are considered two bycatch events.

1.3 Overview of Problem

Because so many of the bycatch estimates are legally mandated and used for management purposes, a probability sampling design that is expected to provide the best average efficiency over the species of greatest concern, typically the protected species and marine mammals, seems preferable. In addition to bycatch estimation, the observer data is used by different people for multiple purposes, such as predicting future bycatch levels and exploring multivariate relationships with bycatch and potential explanatory variables. These studies have and may continue to disregard the sampling design that generated the observer data and the resulting data structure; therefore, a design that does not influence inference will mitigate the impact of a naive analysis. Taking into account these multiple objectives, it is natural to consider using the simple random sampling (SRS) design as it is the design that a naive analysis commonly assumes. This design is not practical or cost effective for sampling the DSLL fishery as it requires employing enough observers to cover the clumps of samples in a brief time period that can naturally occur when drawing a SRS or when the fleet becomes very active.

When deciding on a sampling design and estimators of bycatch for the DSLL fishery, there are three core problems to solve:
(1) Create a probability sampling design that can adapt to the fluctuations in observer coverage.
(2) Establish point estimators of bycatch that take into account the sampling design and resulting unequal sampling probabilities.
(3) Develop interval estimators that are appropriate for the sampling design and the variety of probability distributions of bycatch.

This document describes the novel sampling design created to sample the DSLL fleet. The properties of the sampling design, resulting data structure, and corresponding design-based point estimators of bycatch are presented in this document. The sample that is actually selected using this novel design can influence inference; consequently, an analysis that ignores the design features can result in misleading inference. Potential inferential disasters resulting from ignoring the sampling mechanism and data structure are described in further detail in Section 5. Interval estimators of bycatch are covered in McCracken (in prep.*b*) and domain estimators for the total number of cetacean bycatch events resulting in a dead or serious injury classification are covered in McCracken (in prep.*a*).

2 Sampling Design

When a trip is selected to be sampled, an observer is placed aboard the vessel for the duration of the trip and instructed to observe the complete haul-back of every fishery operation. Our problem is to determine how to select trips for observer placement, preferably using a probability sampling design. To begin addressing this question, a sampling frame needs to be chosen. The list of permitted vessels and the list of notifications are readily available for use as a sampling frame. Regarding the list of notifications, notifications are recorded sequentially in the order they are accepted. Since they are accepted prior to a trip's departure, they can be used as a sampling unit. Once a notification is selected, the trip linked to it is deemed selected for observation. A drawback to using this list is that it is not complete until the year's end. In contrast, the list of permitted vessels is available prior to the new year. There are a couple problems with using this list. First, a permitted vessel may not be active in the DSLL fishery throughout the year. Second, all trips by the selected vessels will need to be observed (likely to be perceived as unfair by the crews of the selected vessels), or a secondary random sample of trips from each selected vessel is required. Because a list of a vessel's DSLL trips does not exist prior to the new year, drawing a probability sample is difficult unless we use the list of notifications. After considering both lists, we decided to use the list of notifications as the sampling frame.

The next problem to resolve is how to randomly select the notifications. A systematic sample is a natural design to consider: it can be easily drawn using the list of notifications and is efficient when units near each other tend to be similar (Thompson, 1992, p. 123). The systematic sample is a probability sample; that is, all notifications have a probability of being sampled and this probability is positive and known. Regarding placing observers, two benefits of this design are that the selected notifications are spread out evenly among the notifications and we know when the next selected trip is approaching. Two characteristics of the systematic sample do not meet our needs. First, the systematic sample maintains a constant level of coverage and enough observers are needed to cover all selected trips during the periods of higher fishing activity. Second, the systematic sample will not accommodate the periods when there are more observers ready for deployment than required to cover the systematic sample. For example, when an observer training class is completed or the fishery is not very active. To maintain the systematic sample at the required 20% coverage level will increase the cost of the observer program because more observers will need to be on staff and paid when they are not deployed. These requirements typically cannot be met under the current level of funding.

Alternatively, a systematic sample can be drawn at a level that observer coverage will not exceed and the missing observations handled appropriately during estimation. How to appropriately handle the missing data will depend on if it can be considered missing completely at random (MCAR), missing at random (MAR), or missing not at random (MNAR). An informal definition of these three missing-data mechanisms in terms of our problem is as follows. If the probability a selected notification is not sampled is completely unrelated to its $bycatch$-value or any other observed variable, including those used in the sample design, the missing data are MCAR. If the probability a selected notification is not sampled does not depend on its $bycatch$-value but can be fully accounted for by observed variables where there is complete information (not missing for selected notifications), the data are MAR. If the probability a selected notification is not sampled depends on its *bycatch*-values and cannot be completely explained by observed values, the data are NMAR. As observer coverage fluctuates over time, the missingness will be related to when the notification is received; that is, the missing data are not MCAR but may be MAR. If the missing data are MAR and a model can explain the missing-data mechanism, the missing data can be ignored after the model accounts for it. In our situation, the challenge to inference based on the assumption that the data are MAR will be specifying the model. It will be impossible to verify statistically that the missing data are MAR and that we have successfully modeled the missingness. Modeling the missing-data mechanism may improve many of the estimates, but will rarely eliminate all missing-data bias as the model is very unlikely to describe the true state of affairs. Little and Rubin (2002) provide formal definitions of these data-missing mechanisms and information concerning statistical analysis with missing data.

Another alternative is to draw a systematic sample at a level that observer coverage will not go below and draw additional notifications, when needed, by another strategy. Since few trips selected by the systematic sample are likely to be missed and the missing data can likely be assumed MCAR, the analysis of the systematic sample is straightforward and well established. The challenge with this strategy is how best to draw additional notifications and include them in an estimator. Because this strategy provides a probability sample from a well established design for part of the complete sample, the decision was to develop it. To reach a balance between obtaining a probability sample and being cost effective, the sampling design used since mid-year 2002 has a two-stage sampling protocol. This two-stage design accommodates fluctuating coverage levels while utilizing observers efficiently.

2.1 The First Stage

The first stage of the sampling protocol is a systematic sample. The systematic sample is drawn at approximately 5% lower coverage than the targeted coverage level specified by the observer program. Drawing the systematic sample at this level seems to provide the maximal percent coverage by the systematic sample in which few selected trips are missed.

A detail to be resolved concerning the systematic sample is the number of starting points. Once a starting point is selected, every k^{th} unit thereafter is chosen to be sampled. When drawing a systematic sample for sampling the DSLL fishery, 5 starting points are selected from the integers 1 to k using simple random sampling without replacement (SRSWOR). Using 5 starting points provides the benefits of multiple starting points while preventing too many randomly selected trips being clumped together by random chance. The decision to use 5 starting points was based on practical considerations and not on any statistical inference regarding the number of starting points for maximal precision. For m denoting the number of starting points and C denoting the targeted percent coverage of the systematic sample, $k = 100m/C$. As an example, when $C = 15\%$ and $m = 5$, $k = 33.33$. Rounding k down to the integer 33 provides approximately 15.15% coverage by the systematic sample. This level of coverage has been the most commonly targeted level of coverage by the systematic sample. Hereafter, let M denote the number of clusters—the rounded value of k.

A systematic sample is a special case of a cluster sample with M clusters in the population. For example, suppose there are a total of 100 trips and a systematic sample at 20% coverage with 5 starting points is to be drawn. Using sets of notification numbers, the $(100 * 5)/20 = 25$ clusters that define the population are $\{1, 26, 51, 76\}, \{2, 27, 52, 77\}, \ldots, \{25, 50, 75, 100\}$. Similarly, for a total of N trips, the i^{th} cluster ($i = 1, \ldots, M$) is the set of numbers defined by the sequence of numbers starting at the value of i and increasing by the increment of M up to the sequence value less than or equal to N.

To draw a sample of 5 clusters, only 5 starting points between 1 and 25 need to be drawn to define the selected clusters. If only 1 starting point is drawn, then the sample contains 1 cluster; consequently, it is not possible to obtain an unbiased estimate of the variance of the estimated bycatch (Thompson, 1992, p. 119).

In summary, the systematic sample is a probability sample of the fleet. The systematic sample is a one-stage cluster sample where all elements in the selected clusters are sampled. Hereafter, the clusters defined by a systematic sample are called "systematic clusters." These systematic clusters are the primary sample units of the systematic sample, and the notifications are the primary elements of the systematic clusters. The primary sampling units are selected by SRSWOR.

2.2 The Second Stage

Now let's consider drawing the additional samples required to achieve the targeted coverage level. Only after all upcoming notifications selected by the systematic sample are assigned an observer and there are still observers ready to be deployed should additional samples be drawn. The method for drawing these samples needs to be straightforward as they are needed quickly and with little forewarning. Drawing the additional notifications using SRSWOR from the list of notifications still eligible for observer placement is straightforward and the method that the observer program is instructed to use.

Because the occasions when secondary samples are drawn are not randomly selected but determined by the need to deploy observers, the probability a notification is selected by the secondary sample is unknown and needs to be approximated. To approximate these probabilities, the contractor's list of notifications is used. Examination of this list reveals periods when coverage appears to have been greater or less than the full targeted coverage. Further details regarding approximating these probabilities are provided in Section 3. An outcome of the secondary sample is that notifications are selected with unequal probability. For example, notifications that are included in the sampling frame of the secondary sample will have a greater probability of being selected than those excluded.

2.3 Systematic-Plus Sample

Hereafter, this two-stage design is called a "systematic-plus" (SYSPLUS) design. The collection of samples selected by the secondary method is called the "plus sample." The term "day sample" refers to a sample that is drawn from all eligible notifications on a day when additional observers need to be deployed. The plus sample typically consists of several day samples. The SYSPLUS sample is not a traditional two-stage sample, as the term "two-stage design" commonly refers to a design in which after selecting a sample of primary units, a sample of secondary units is selected from each of the selected primary units. The second stage of the SYSPLUS sample selects notifications that were not selected by the first stage systematic sample. The SYSPLUS sampling design is a complex adaptive design: it adapts to the availability of observers.

2.4 Implementation of the Systematic-Plus Sample

Since 2013 a new systematic sample is drawn yearly at a level of coverage that can be maintained, usually 15% coverage. If the percent coverage of the systematic sample needs to be adjusted during the year, a new systematic sample is drawn. This strategy encourages maintaining at least 15% observer coverage and allows for a quick reaction to a shortage of observers. For example, as a consequence of having to delay an observer training course, there was a shortage of observers at the beginning of 2014, so a systematic sample was drawn at 10% coverage. After newly trained observers passed the required exam and were ready to be deployed, a new systematic sample was drawn at 17.25% coverage and maintained until the 2015 systematic sample began. When more than one systematic sample is drawn in a year, the year's sample is stratified with SYSPLUS sampling within stratum. Hereafter, a stratified sample with SYSPLUS sampling within stratum is called a “stratified SYSPLUS sample.”

This current strategy differs from the original protocol. When the SYSPLUS sample was first used in 2002, there was some interest in quarterly bycatch estimates, so a new systematic sample was drawn quarterly. During years 2005-2009, the first quarter systematic sample was drawn at 10%, instead of 15%, coverage to ensure observers were available to cover this sample and the SSLL fleet (the SSLL fleet is usually most active the first quarter). The lower coverage in the first quarter was offset by drawing day samples throughout the year so that the required annual 20% coverage was achieved. This strategy allowed for greater variability in the level of coverage throughout the year at the expense of precision. In an effort to increase the precision of the annual estimates with minimal additional cost, the observer program was encouraged to reduce the variability in observer coverage and maintain a systematic sample at 15% coverage throughout the year.

2.5 Accommodating Research Trips

In previous years, trips that participated in one of several NOAA research projects may have interfered with normal fishing operations. Except for a few small research projects, all trips that participated in a project had a NOAA observer aboard performing their normal responsibilities and tasks required by the project. Trips involved in projects with 100% observer coverage were excluded from the sampling frame and generally considered unrepresentative of unsampled trips. There is a field in the Hawaii longline logbook database and LODS that identifies research trips. Because these trips fished within the DSLL regulations and under the incidental take permit for the DSLL fishery (rather than under a research permit), their bycatch was part of the fishery’s bycatch. For the years where there were research projects with 100% coverage, the estimated bycatch was the sum of the bycatch on the research trips and the estimated bycatch for the remaining fleet’s effort.

Most projects without observers aboard did not interfere with the normal fishing operation and involved no more than a couple trips per year. For these projects, the participating trips were considered part of the sampling frame and treated as if they were not selected by the SYSPLUS sample. For the projects without observers aboard that did interfere with the normal fishing operation, there was a NOAA scientist aboard that recorded the bycatch for the trip. Trips with a NOAA scientist aboard recording bycatch were treated in the same manner as those with a NOAA observer aboard. Trips involved with research projects in the future will likely be handled in the same manner.

2.6 Exclusion Bias

On 27 August 2012, a change in the regulations of the Hawaii DSLL fishery imposed new limits on swordfish landed (NOAA, 2012*b*). The new limits are as follows: (1) With a NMFS observer aboard, there is no limit on the number of swordfish landed or possessed on a trip, regardless of the type of hook used; (2) If the vessels uses only circle hooks and does not have a NMFS observer aboard, the limit is 25 swordfish landed or possessed on a trip; (3) If the vessel uses any hooks other than circle hooks and does not have a NMFS observer aboard, the limit is 10 swordfish landed or possessed on a trip. In essence, this regulation created three components of the DSLL fishery defined by the number of swordfish a trip can keep. Since 27 February 2013 the Hawaii DSLL fishery has been required to use circle hooks as part of the final False Killer Whale Take Reduction Plan and regulatory measures (NOAA, 2012*a*), thus, the third component no longer exist. Regardless what sampling design is used to select notifications, the first component (no limit on swordfish kept) will have 100% coverage: the second component will have no observer coverage. Prior to the new regulations, all DSLL trips had a trip limit of 10 swordfish landed. The regulation limiting number of swordfish landed was put into place to discourage trips from targeting swordfish, which typically implies setting the gear shallow. The shallow setting of gear has historically resulted in different observed catch rates for the protected species.

The exclusion of some of the population—the second component—from the sample gives rise to the potential of exclusion bias and places limits on how much information our sample can provide about the population. Extrapolating from our sample to the population requires making assumptions about the population that cannot be confirmed from the sample. Advice was put forward that the regulations should be revised and observer presence should never have any bearing on what fishing practice is allowed. No revision has occurred.

Although a SYSPLUS sample is still generated, the new regulations changed what the sample represents. Prior to the new regulations, the sample was a random sample of all DSLL trips fishing under a uniform set of rules and requirements. Under the new regulations, what is being randomly drawn is a selection of trips that will have an observer aboard and no limits on swordfish landings.

2.7 Notation for the Stratified SYSPLUS Sample

Herein, the following notation is used when referring to the notifications and SYSPLUS sample. Let i represent a notification (trip) and selected information linked to the notification. This information includes (1) the landing year of the trip, (2) the trip's inclusion probability, (3) the trip's observed bycatch, and (4) measures of effort for the trip. The measures of effort include (1) a variable called *ntrip* that refers to a DSLL trip for which

$$ntrip = \begin{cases} 1 & \text{the gear is deployed at least once and the catch is landed in the year of interest} \\ 0 & \text{otherwise,} \end{cases}$$

(2) a variable called *nsets* that equals the total number of fishing operations during the trip ($nsets = 0$ if $ntrip = 0$), and (3) a variable called *nhooks* that equals the total number of hooks set during the trip ($nhooks = 0$ if $ntrip = 0$).

For a stratified SYSPLUS sample, let h denote a stratum and H denote the number of strata. For stratum h, let $\mathcal{U}_h$ denote the set of all notifications received in h, C_h denote the set of systematic clusters that are sampled, $\mathcal{P}_h$ denote the set of notifications in the plus sample, and $\mathcal{S}_h$ denote the SYSPLUS sample, $\mathcal{S}_h = (C_h, \mathcal{P}_h)$. Specific to the systematic sample, let M_h denote the number of clusters in h and m_h denote the number of clusters sampled in h.

3 Approximating Inclusion Probabilities

Design-based estimators are based on inclusion probabilities—the probability that a unit of the population is included in the sample. The fundamental idea behind design-based inference is that an individual with an inclusion probability of π_i represents $1/\pi_i$ individuals in the population. The value $w_i = 1/\pi_i$ is called the sampling weight. Being the reciprocal of the inclusion probability, notifications with higher inclusion probabilities have smaller sampling weights and vice versa. If a systematic sample generated the complete sample in strata h, then $\pi_{hi} = Pr(i \in C_h) = m_h/M_h, \forall i \in \mathcal{U}_h$.

For a stratified SYSPLUS sample, the inclusion probability for a notification in stratum h is

$$\pi_{hi} = Pr(i \in \mathcal{S}_h) = Pr(i \in C_h) + Pr(i \notin C_h)Pr(i \in \mathcal{P}_h \mid i \notin C_h)$$

where

$$Pr(i \notin C_h) = 1 - Pr(i \in C_h) = 1 - \frac{m_h}{M_h}.$$

The $Pr(i \in \mathcal{P}_h \mid i \notin C_h)$ is the probability that notification i is included in the plus sample given it is not in the systematic sample. Denote this conditional probability as π_i^+. There are challenges in determining π_i^+. Some of the challenges are as follows:

(1) When day samples are drawn is not predetermined or selected using a probability sampling design.
(2) Because the timing of the day sample determines what notifications are in the sampling frame, the number of notifications in the sampling frame is a random number.
(3) The sampling frames for the day samples are not necessarily mutually exclusive because day samples drawn less than 48 hours apart may include some of the same notifications.
(4) What systematic sample is drawn may influence the size and timing of the day samples; hence, the sample size of a day sample is a random number.

The consequence of adding the plus sample to the systematic sample is that the π-values become unknown. If a design-based estimator is to be employed, either we ignore the plus sample and only use the systematic sample or we approximate the π^+-values.

Let's consider ignoring the plus sample and using a design-based estimator for a the systematic sample. As the theory behind a systematic sample and corresponding design-based estimators are well established, the estimation of bycatch is straightforward. Along with the potential of losing precision, there is a practical concern to ignoring the plus sample. Several species whose bycatch is of concern, such as marine mammals and sea turtles, are rarely bycaught and few—if any—are observed bycaught within a year. Potentially, the only observed bycatch event could be in the plus sample.

Now let's consider approximating the π^+-values. A set of assumptions that provide a straightforward way to approximate the π^+-values is to assume that the sampling frame and sample size of the day sample are predetermined. The drawback to this approach is that w_i can constantly fluctuate between values from 1 to M/m ($M/m \approx 6.7$ for 15% coverage by the systematic sample). For example, two day samples with relative high coverage ($w_i \approx 1$) can be separated by only a few notifications where $w_{hi} = M/m$, yet all the notifications involved may have been received during the same week. It is difficult to justify the weighting in this example, especially if the year's only observed bycatch event occurred during this period.

The objective of the SYSPLUS sampling design is to allow and account for fluctuating levels of observer coverage. I proceed with the strategy of smoothing over daily fluctuations of coverage while preserving the general pattern of increased and decreased coverage levels. For example, if there is an interval of time where the day samples are drawn at approximately the same rate, the notifications accepted during this time interval are assumed to have equivalent inclusion probabilities. The systematic sample running through this interval is still treated as a systematic sample. It is the computation of $Pr(i \in \mathcal{P}_h \mid i \notin C_h)$ that is defined by this approach.

Following this strategy, the period that defines a systematic sample (stratum) is split into one or more time intervals where the π^+-values are assumed to be equivalent within an interval. These time intervals are mutually exclusive, and their union includes all notifications within the stratum. Henceforth, these intervals are called the "π^+-classes": notifications have been classified into classes (intervals) in which inclusion probabilities are equivalent within a class.

When computing the π^+-values within a π^+-class, several variables are assumed to be fixed numbers (i.e., not random numbers). The following are these variables:

(1) The number of π^+-classes within a stratum.
(2) The notifications that define the beginning and ending of an interval, implying the number of notifications within each group is a fixed number.
(3) The number of plus samples in each π^+-class.

(4) The number of notifications selected by the systematic sample that fall within each π^+-class.

Let G_h denote the number of π^+-classes in stratum h. For each π^+-classes $g = 1, \ldots, G_h$, the sampling weights for the plus sample are computed assuming that this sample is a SRSWOR of n^+_{hg} notifications from the N^+_{hg} notifications in g that are not part of the systematic sample. That is,

$$\pi^+_{hgi} = \frac{n^+_{hg}}{N^+_{hg}},$$

and

$$\pi_{hgi} = \frac{m_h}{M_h} + (1 - \frac{m_h}{M_h})(\frac{n^+_{hg}}{N^+_{hg}}) \quad \forall i \in \mathcal{U}_{hg} \tag{3.1}$$

where $\mathcal{U}_{hg}$ is the set of all notifications in g.

Statistical properties, such as unbiasedness, of design-based estimators assume the sampling weights are known. The sampling weights of a SYSPLUS sample are approximated; consequently, the accuracy of a design-based estimator of bycatch will depend on how well notification i represents the other notifications of its class. The frequency that notifications are sampled using the plus sample will likely influence the accuracy of the approximated weights. If only a couple notifications are selected by the plus sample, there will be little variability between the sampling weights and the weights will be equal or near the weight of the systematic sample, M/m. Whereas, if a large percentage of the SYSPLUS sample is composed of the plus sample and the day samples come in clumps and voids, the approximated weights will be more variable and potentially less accurate. Because this strategy will never exactly describe the true state of affairs, its implications should always be considered. Including the plus sample by approximating the sampling weights may improve many of the estimates but is not expected to eliminate all the bias introduced by including this sample.

3.1 Determining the π^+-classes

Because the first stage sample is a systematic sample, we know that a block of k consecutive notifications has at least m selected notifications. Using this fact, blocks of k consecutive notifications are identified where the first block contains notifications 1 to k, the second block contains notifications $k + 1$ to $2k$, the third block contains notifications $2k + 1$ to $3k$, and so on. Within each block, the notifications selected by the plus sample are identified.

The next step is to combine neighboring blocks into a collection of blocks whose notifications appear to have approximately equal probability of being selected by the plus sample. The level of observer coverage can change when the activity level of either fishery (shallow-set or deep-set) or the number of observers currently active (deployed or waiting to be deployed) changes. To aid in deciding what blocks to combine, the number of observers deployed and the level of observer coverage is computed on a daily basis and examined for changes in observer coverage.

Except for the final block in a stratum, each block contains M notifications, and the number of notifications in a π^+-class is a multiple of M. Thus, the number of systematic samples in the group is m times the number of blocks in the π^+-class (denote as b), and the number of notifications not selected by the systematic sample is $b(M - m)$. Regarding the final block, which usually has less than M notifications, it is assumed that the number of systematic and plus observations are fixed.

3.2 Adjustments to Inclusion Probabilities

There are situations when adjustments to some of the values involved in (3.1) are required. These situations are rare and typically involve missing observations or trips that did not materialize as plan. Descriptions of these situations and their corresponding adjustments complete this section on computing inclusion probabilities.

3.2.1 Missing Observations

Although every reasonable effort is made to sample selected trips, occasionally, a selected trip is not observed. When this happens, the reason is entered into the notification logbook or LODS. These reported reasons help in determining if the missing observations can be treated as MCAR. The two most common reasons are (1) an observer was not available for deployment and (2) the observer was unable to fulfill their duty because of illness or injury. Because day samples are drawn when an observer is ready for deployment, only the systematic sample is subject to observer unavailability. In rare cases, after debriefing an observer, the observer program determines that the data collected by the observer is unreliable. In this situation the selected trip is treated as a missing observation.

Since 2010 the yearly percentage of unobserved selected trips has been between 1% to 0%. As these missed trips were scattered throughout the year, they are assumed to be MCAR. Since observer coverage did not fall below the level of the systematic sample, the closest plus observation was substituted for the missed trip.

Prior to 2010, the coverage level occasionally fell below that of the systematic sample; consequently, a series of blocks would have less than m_h observed samples per block with no plus observations nearby (within approximately a week). During the estimation process, this series of blocks would be treated as a stratum with a SRSWOR of notifications. Because the current protocol is to draw a new systematic sample when there is a shortage of observers, this situation no longer occurs.

3.2.2 Notifications that do not Materialize as Planned

We now discuss situations where an observer has been assigned to a selected trip, but the trip did not materialize as expected. Because having an observer aboard is generally considered undesirable from the fishermen's viewpoint, policies are made to insure that selected trips do not avoid being observed.

First, consider when a selected trip delays its departure for many days or returns to port before setting any hooks and remains in port for many days. In either case, a new notification is required before the vessel can depart on a trip; this notification is automatically selected. When computing the inclusion probabilities, both notifications are considered part of the random sample. The second notification is labeled as a systematic observation if its number is part of the systematic sample; otherwise, it is labeled as a plus observation. The original selected notification is labeled as a “did not fish" (DNF) trip and assigned no effort or bycatch.

Originally, the realization of the trip was linked to the original notification, and the second notification was not considered part of the sampling frame. Upon reflection, I realized that the timing of the original and second notifications influence the selection of notifications nearby. With the original notification, the observer is quickly reassigned to another selected trip, and with the second notification, an available observer is assigned to the trip, reducing the number of observers available to sample nearby notifications. Moreover, delayed and unfulfilled trips are not always identified when their notifications are not selected, yet a notification with a revised departure date if often provided. When a design-based estimator is used, assigning no effort and bycatch to the original notification accounts for the notifications of unrealized trips that were not selected. Furthermore, if the departure data is related to bycatch, the realized trip's bycatch is expected to be more representative of its second notification's π^+-class. So far, there have been 1 to 5 DNF trips per year.

Next, let's consider when fish are landed after just a few fishing operations. To discourage the practice of doing short trips when an observer is aboard, when fish are landed after less than 5 fishing operations, an observer is assigned to the vessel's next trip. The notification linked to the next trip is considered part of the random sample and labeled as a systematic observation if its number belongs to the systematic sample; otherwise, it is labeled as a plus observation. Short trips rarely occur. There is one exception to this policy. On rare occasions, fishermen will depart from Oahu, fish a small number of sets, land their catch at a port on the Big Island, and very soon afterwards continue on a longer trip. These trips are usually treated as one trip and a second notification is not normally required. An observer will stay aboard for the full length of the trip if it has been selected. In rare instances, less than four a year, a captain or vessel owner contacts PIROP and wants to switch a DSLL notification to a SSLL notification, or vice versa. When a DSLL notification is switched to a SSLL notification, prior to computing sampling weights, the block where the DSLL notification occurred is treated as if one less DSLL notification was recorded. If the DSLL trip belongs to the systematic sample, a nearby plus observation replaces it. When a SSLL notification is switched to a DSLL notification, the notification of the change is typically recorded and given the appropriate notification number. Depending on the circumstances, the observer assigned to the trip prior to the change may stay assigned to the trip. In this situation, the notification of the change is identified as a systematic or plus observation depending on if its number belongs to the systematic sample. This policy will be reviewed if these switches become more frequent.

4 Point Estimation of Total Bycatch

4.1 Assigning a Trip's Bycatch to a Year

Because the dates of a trip's notification, departure, and landing can belong to different years, the year that a trip's bycatch is assigned to needs to be determined. The most straightforward analyses occur when the trip's bycatch is assigned to the trip's notification date. Unfortunately, this assignment is not acceptable from the perspective of fishery management. Since annual bycatch estimates of protected species are often needed soon after the completion of the year, a trip's bycatch is assigned to the trip's landing date. This assignment insures that all trips assigned to the year are completed by the end of December. The disadvantage of this assignment is that a stratum can have trips belonging to different years; consequently, estimation is more complicated and less precise.

4.2 Estimators of Bycatch

For an unequal probability sample without replacement, the Horvitz-Thompson estimator (HTE) and generalized ratio estimator (GRE) are appropriate estimators of population totals. The HTE is simply the sample sum of the observed values expanded by their sampling weights. Let τ denote a year's bycatch and y_i denote trip i's observed bycatch, where $y_i = 0$ if the trip did not land during the year. The HTE of τ for a stratified SYSPLUS sample can be expressed as

$$\hat{\tau}_\pi = \sum_{h=1}^{H} \sum_{g=1}^{G_h} \sum_{i \in \mathcal{S}_{hg}} \frac{y_{hgi}}{\pi_{hgi}} = \sum_{h=1}^{H} \sum_{g=1}^{G_h} w_{hg} \sum_{i \in \mathcal{S}_{hg}} y_{hgi}$$

where all strata and π^+-classes with trips landing during the year are incorporated, including those from the previous year's sample. When the inclusion probabilities are known, the HTE is an unbiased estimator.

Because bycatch is often perceived to be proportional to fishing effort (x), the GRE is of interest. For the GRE to apply, the two quantities y and x must be measured on each sample unit and the population total of the x-values exactly known. Let τ_x and τ_y denote the population totals of the x-values and y-values, respectively. The GRE is

$$\hat{\tau}_{gre} = \frac{\hat{\tau}_{\pi,y}}{\hat{\tau}_{\pi,x}}\tau_x \tag{4.1}$$

where the components $\hat{\tau}_{\pi,y}$ and $\hat{\tau}_{\pi,x}$ are the Horvitz-Thompson estimates of τ_y and τ_x. The GRE is not an unbiased estimator of the population total; although, for a large sample the bias of the GRE is typically sufficiently small enough to obviate concern (see Sarndal et al., 1992, sec. 7.3 or Gregoire and Valentine, 2008, p. 169 for more details).

With stratified sampling, we can first combine the strata to estimate $\hat{\tau}_{\pi,x}$ and $\hat{\tau}_{\pi,y}$, and then apply the GRE (4.1) to derive the combined GRE. Or, we can first apply the GRE to each stratum, then combine the estimates to derive the separate GRE. The separate GRE is expressed as

$$\hat{\tau}_{sgre,y} = \sum_{h=1}^{H} \hat{R}_h \tau_{x,h}$$

where $\hat{R}_h$ is the ratio of the stratum's Horvitz-Thompson estimates $\hat{\tau}_{\pi,y,h}$ and $\hat{\tau}_{\pi,x,h}$. The separate GRE can improve efficiency if the $\hat{R}_h$ vary from stratum to stratum; whereas, the combined GRE does not take advantage of the extra efficiency stratification provides. When some strata sample sizes are small, the separate GRE is not recommended because each ratio is biased and the bias can propagate through the strata (Lohr, 2010, p. 145).

Since the GRE is not unbiased, comparing its mean square error (MSE) relative to the MSE of other estimators provides a measure of its efficiency. The MSE of the GRE will be relatively small when the variance of the residuals $y_i - Rx_i$, where $R = \tau_y/\tau_x$, is much smaller than the variance of the y-values. The GRE is most appropriate when a straight line through the origin summarizes the relationship between the x-values and y-values and the variance of the y-values about the line is proportional to the x-values.

Next, let's consider the variance of the HTE. The variance of the HTE is relatively small when there is an approximate proportional relationship between the y-values and π-values; when there is no such relationship, the variance of this estimator can be very large. If the y-values are not proportional to either the π-values or any auxiliary variable, using the GRE with $x_i = 1$ for all units is recommended (Hajek, 1971 or Thompson, 1992, p. 69-70).

How does all this information relate to estimating bycatch using the stratified SYSPLUS sample of the DSLL fishery? Let's start with the expected relationships between the $bycatch$-values and the π-values or auxiliary variables $ntrip$, $nsets$, and $nhooks$. Because the π-values are related to the number of observers actively employed and the level of activity in the fishery, a proportional relationship between the $bycatch$-values and π-values is not expected. Therefore, we expect that the GRE with $x_i = 1$ will be more efficient than the HTE. Using the auxiliary variable $ntrip$ should have similar benefits to using the GRE with $x_i = 1$. Compared to using $x_i = 1$, the variables $ntrip$, $nsets$, and $nhooks$ have the advantage of naturally correcting for notifications where $y_i = 0$ because the trip landed in another year or was labeled DNF. Using the variable *nhooks* may introduce some bias because it is likely measured with some error. The variable *nsets* is probably pretty accurate; *ntrips* is likely very accurate.

Regarding the stratification, the strata are created to accommodate observer availability and not to maximize the difference between the values of R_h. Consequently, the best conditions for gaining efficiency by the separate GRE are unlikely to be met.

In practice, the HTE and the combined GRE with the auxiliary variables *ntrips*, *nsets*, and *nhooks* have all been considered. For each species, scatterplots of the *bycatch*-values and the variables *nsets* and *nhooks* have been examined for evidence of an approximate proportional relationship and the estimated efficiency of the estimates over several years have been evaluated (see McCracken (in prep.*b*) for estimators of variance and MSE). If an approximate proportional relationship between the *bycatch*-values and *nsets* or *nhooks* exist, it should be apparant year-to-year.

For most species, the *bycatch*-values do not have an approximate proportional relationship with *nsets* or *nhooks*. When *nhooks* exhibits this relationship, *nsets* also exhibits this relationship. Furthermore, the GRE with *nsets* results in similar or superior efficiency compared to the GRE with *nhooks*; therefore, using the less accurate metric of *nhooks* is not warranted. When comparing the GRE with *ntrips* to HTE, the HTE was less stable in the earlier years when observer coverage would go below 10% for a period of time. Since observer coverage has become less variable, the HTE has become more stable and the GRE with *ntrips* has less of an advantage.

When deciding on the estimator for the reported annual estimates of a species, the following guidelines are currently followed. If an estimator is consistently more efficient than the others, the reported annual estimates are derived using this estimator. When an estimator is not consistently more efficient, the HTE is used when domain estimators of τ for subgeographical areas are required (see McCracken (in prep.*a*) for an explanation); otherwise, the GRE with *ntrips* is used. As a year's bycatch is a count, the reported estimated total bycatch is rounded to the nearest integer.

5 Model-Assisted and Model-Dependent Inference

The availability of observers drives the sample design, and the frequent requirement to quickly estimate total bycatch for multiple species compels the use of the HTE and GRE. These are not the optimal conditions for statistical inference. As the practical constraints that instigated the creation of the SYSPLUS design are unlikely to vanish, there is limited opportunity to improve on the sample design. Other approaches to inference done on a per species basis may provide improved inference.

Two optional approaches that may be worth considering are (1) model-assisted inference and (2) model-dependent inference (also called model-based inference). With the model-assisted approach, the model motivates the form of the estimator, but inferences and properties of the estimators depend on the sample design. If the model is not a good approximating model, the model-assisted estimators may not increase precision, but the properties of the estimators will hold and inference will generally be correct. As this approach depends on the sample design, it is considered a sample-based approach. The GRE is a model-assisted estimator. Sarndal et al. (1992) provides a comprehensive description of the model-assisted approach.

The model-dependent approach uses an assumed model for the survey outcomes (y-values) to predict the outcomes of the units in the finite population that were not sampled. The main challenge with the model-dependent approach is specifying an appropriate model. As all models are approximations, they are subject to some level of misspecification. Serious model misspecification can lead to incorrect inferences. Before specifying a model, it is vital that the data collection mechanism is well understood. A model-dependent approach that ignores the design is not valid unless the design is ignorable (Rao, 2011; Gelman, 2007; Little, 2004; Sarndal et al., 1992; Gelman et al., 2004, chapt. 7; and Lohr, 2010, sec. 11.2). Chapter 7 in Gelman et al. (2004) provides a helpful discussion on ignorable and nonignorable designs and approaches for accounting for the data collection process. Erroneous model-dependent inference can be avoided by using a data collection and modeling strategy that provides robustness to model misspecification. One such strategy is to sample the finite population using a probability sampling design and then selecting models that properly incorporate the design features. The SYSPLUS design is noningnorable and its features need to be properly incorporated into the model for valid inference.

The two major model-dependent approaches to survey sampling inference are superpopulation modeling and Bayesian modeling. The superpopulation model-dependent approach assumes that the finite population y-values are a random sample from a superpopulation whose values are generated from a specified model and that the same model holds for the survey's observed values. If sampling weights, clustering, and stratification or other features of the design are related to the y-values, then the model needs to properly reflect these features. With the Bayesian approach, the method of data collection dictates the minimal level of modeling required (Rubin, 1983). The Bayesian model formulation generally involves classical Bayesian data analysis of the superpopulation model and modeling the data generating process. Gelman (2007) provides a good overview of poststratification and survey weighting as an approach for accounting for the data collection process in regression modeling.

For insightful discussions on the differences between the sample-based and model-dependent approaches see Rao (2011), Little (2004), Rao (1997), Hansen et al. (1983), and Sarndal (1978).

As the stratified SYSPLUS design involves the features of unequal sampling weights, clustering, and stratification; a sample-based or model-dependent approach that ignores the sampling mechanism is not valid unless the y-values are not related to these features. What can happen if the SYSPLUS design is ignored? If the π-values are related to the $bycatch$-values, then an analysis that does not account for the different π-values may lead to incorrect inference. Since the π-values vary over time, we need to consider if the $bycatch$-values also vary over time. For example, if we postulate that the population density of a bycaught species affects the chance of it being bycaught, then assuming that the $bycatch$-values do not vary over time seems questionable since the DSLL fishery follows the tuna populations whose movement may be different than the bycaught species. As we do not have a good understanding of all the variables that affect bycatch, assuming that $bycatch$-values may vary over time is recommended, unless there is sufficient evidence to suggest otherwise.

The SYSPLUS design involves cluster sampling at both stages: the systematic sample is a cluster sample of systematic clusters, and the plus sample is a cluster sample of trips. Regarding the systematic sample, if notifications can reasonably be conceived as being in random order, then the systematic sample is likely to produce a sample that behaves like a SRS of notifications. In this case, SRS procedures can be used for inference, and notifications within a systematic cluster can be treated as independent observations when modeling bycatch. However, if $bycatch$-values from nearby notifications tend to be more similar than $bycatch$-values from notifications that are farther apart, then using design-based estimators for SRS will likely overestimate the variance of the estimators of τ (Thompson, 1992, p. 119). Similarly, model-based predictors that assume independent $bycatch$-values for trips will likely produce incorrect standard errors and confidence intervals, even if the model parameters are approximately unbiased (Lohr, 2010, p. 435,453-455). As the systematic sample is drawn from time-ordered notifications, SRS inference is inappropriate if *bycatch*-values vary over time.

Let's now consider that a trip is a cluster of sets. If $bycatch$-values of sets from the same trip tend to be more similar than $bycatch$-values of sets selected at random from all DSLL sets, then the *bycatch*-values of sets within a trip are correlated. Since sets from the same trip tend to be more similar in location, fishing practices, skills of the crew, quality of bait, and other variables that are potentially related to bycatch, the *bycatch*-values of sets within a trip are likely correlated. If sets within a trip are correlated, then using design-based estimators that assume a SRS of sets or a model-based predictor that assumes independent *bycatch*-values of sets will likely lead to incorrect inference, similar to those resulting from ignoring systematic clusters.

6 Conclusions

Developing a sampling design for the Hawaii DSLL fishery that has known inclusion probabilities, achieves the 20% targeted coverage level, and accommodates the practical constraints is challenging. The SYSPLUS design is a comprise between drawing a sample with known inclusion probabilities and utilizing observers efficiently. The systematic sample drawn during the first stage of a SYSPLUS sample has known inclusion p robabilities. The problem with drawing a systematic sample at 20% coverage is that selected trips are likely to be missed in clumps when the actual coverage level falls below 20%; i.e., the missing data are not MCAR. The solution to this problem is not attainable under current funding because it entails employing more observers and paying them between deployments. Therefore, we can draw a systematic sample above 20% coverage and handle the missing observations during analysis or draw a systematic sample below 20% and draw the additional samples required to achieve 20% coverage when observers are available. With both designs, the challenge is specifying a model that explains observer availability. Drawing the systematic sample below 20% coverage preserves a sample with known inclusion probabilities for a sizable portion of the sample. A consequence of drawing the additional samples, the plus sample, is that the inclusion probabilities become unknown and need to be approximated. Because the statistical properties of design-based estimators assume the inclusion probabilities are known, approximating these probabilities introduces potential bias. To minimize this bias, the coverage of the systematic sample should be drawn at the highest level possible so that very few, if any, samples are missed. To estimate bycatch, the HTE or GRE can be used with the approximate inclusion probabilities. Regardless if a model-dependent or design-based approach is undertaken, the data collection method and resulting data structure needs to be taken into account to derive valid inference. As the systematic sample typically has very few missing samples and the missing samples are usually MCAR, it can be used to derive inference when it is undesirable to include the plus sample.

7 Acknowledgments

The author would like to thank PIROP for their constructive feedback when developing a sampling design and their dedication to quality data throughout the years and Josee Vincent for doing such a wonderful job of placing observers on the selected trips and keeping such good records.

8 References

Gelman A. 2007. "Struggles with Survey Weighting and Regression Modeling (with comment)," *Statistical Science*, 22(2), 153–164.

Gelman A, Carlin JB, Stern SS, and Rubin DB. 2004. *Bayesian Data Analysis*, 2nd edn, New York: Chapman and Hall.

Gregoire TG, and Valentine HT. 2008. *Sampling Strategies for Natural Resources and the Environment*, New York: Chapman and Hall.

Hajek J. 1971. “Discussion of an essay on the logical foundations of survey sampling, part one, by D. Basu,” in *Foundations of Statistical Inference*, eds. V. P. Godambe, and D. A. Sprott, Toronto: Holt, Rinehart, Winston.

Hansen MH, Madow WG, and Tepping BJ. 1983. “An evaluation of model-dependent and probability-sampling in sample surveys (with comment),” *Journal of the American Statistical Association*, 78(384), 776–793.

Little RJ. 2004. “To model or not to model? competing modes of inference for finite population sampling,” *Journal of the American Statistical Association*, 99(466), 546–556.

Little RJA, and Rubin DB. 2002. *Statistical Analysis with Missing Data*, 2nd edn, Haboken: Wiley.

Lohr SL. 2010. *Sampling: Design and Analysis*, 2nd edn, Boston: Brooks/Cole.

McCracken ML. (in prep. *a*). Domain estimators for the total number of cetacean bycatch events resulting in a dead or serious injury classification, Technical report, PIFSC, NOAA, U.S. Department of Commerce.

McCracken, ML. (in prep.*b*). Interval estimation of annual bycatch in the Hawaii deep-set longline fishery, Technical report, PIFSC, NOAA, U.S. Department of Commerce.

NOAA. 2012*a*. Taking of Marine Mammals Incidental to Commercial Fishing Operations; False Killer Whale Take Reduction Plan Western; Final Rule, Federal Register Vol. 77, FR 71259, U.S. Department of Commerce.

NOAA. 2012*b*. Western Pacific Pelagic Fisheries: Revised Swordfish Trip Limits in the Hawaii Deep-set Longline Fishery; Final Rule, Federal Register Vol. 77, FR 43721, U.S. Department of Commerce.

Pacific Islands Fisheries Science Center. 2017. *Hawaii Longline Logbook*.
URL: *https://inport.nmfs.noaa.gov/inport/item/2721*

Pacific Islands Regional Observer Program. 2014. *Hawaii Longline Observer Program Field Manual: LM.14.04*, Pacific Islands Regional Office, Honolulu, Hawaii.
URL: *www.fisheries.noaa.gov/pacific-islands/fisheries-observers/pacific-islands-region-observer-resources*

Pacific Islands Regional Office. 2017. *Longline Observer Data System*.
URL: *https://inport.nmfs.noaa.gov/inport/item/9027*

Rao JNK. 1997. “Developments in sample survey theory: an appraisal,” *The Canadian Journal of Statistics*, 25(1), 1–21.

Rao, JNK. 2011. “Impact of frequentist and bayesian methods on survey sampling practice: a selective appraisal,” *Statistical Science*, 96(2), 240–256.

Rubin DB. 1983. “An evaluation of model-dependent and probability-sampling in sample surveys: Comment,” *Journal of the American Statistical Association*, 78(384), 803–805.

Sarndal. CE. 1978. “Design-based and model-based inference in survey sampling (with discussion),” *Scandinavian Journal of Statistics*, 5(1), 27–52.

Sarndal CE, Swensoon B, and Wretman J. 1992. *Model Assisted Survey Sampling*, New York: Springer.

Thompson SK. 1992. *Sampling*, New York: Wiley.