National Marine
Fisheries Service
NOAA

Fishery Bulletin
established in 1881

Spencer F. Baird
First U.S. Commissioner
of Fisheries and founder
of *Fishery Bulletin*

Development and implementation of a fully randomized sampling design for a fishery monitoring program

Jennifer Cahalan (contact author)[1]
Craig Faunce[2]

Email address for contact author: jcahalan@psmfc.org

[1] Pacific States Marine Fisheries Commission
7600 Sand Point Way NE
Seattle, Washington 98115

[2] Fisheries Monitoring and Analysis Division
Alaska Fisheries Science Center
National Marine Fisheries Service, NOAA
7600 Sand Point Way NE
Seattle, Washington 98115

Abstract—Fishery observer programs in the United States are mandated to collect fishery-dependent data. These monitoring programs must have clear objectives and be statistically sound despite the constraints imposed by commercial fishing operations. In addition, successful programs continuously self evaluate and improve. Decisions regarding sampling design must integrate the concerns and needs of industry and fisheries managers while ensuring the scientific integrity of the data. Here, we use examples from a large observer program to illustrate how the elements of a statistical design can be effectively implemented. With the rich economic resources of the fishing fleet in Alaska and a long history of observer coverage, the North Pacific Observer Program currently has high overall coverage rates and employs a statistically rigorous sampling design. The North Pacific Fishery Management Council receives 2 annual reports on the objectives and performance of fishery monitoring in the region. As fisheries management has become more sophisticated, data needs have increased. Balancing the amount and types of data collected with observer workloads is a critical component of designing sampling methods.

Manuscript submitted 29 May 2019.
Manuscript accepted 5 March 2020.
Fish. Bull. 118:87–99 (2020).
Online publication date: 23 March 2020.
doi: 10.7755/FB.118.1.8

Monitoring catch and bycatch is an important component of successful fisheries management (Sutinen, 1999; Hilborn et al., 2005). Fisheries, meanwhile, benefit from an increased focus on accountability, sustainability, and compliance with seafood traceability standards that foster a positive public perception (Lewis and Boyle, 2017). Because fishery monitoring has had positive effects, numerous studies have been devoted to how fisheries can and should be monitored (e.g., van Helvoort, 1986; Davies and Reynolds, 2003; Zollett et al., 2015).

Few studies however have examined how scientifically and statistically robust monitoring programs are enacted (e.g., Cotter and Pilling, 2007; ICES[1]). If fisheries monitoring programs were well funded and provided with stable, noncompetitive objectives, establishing a successful program should be simple. However, the political, social, and economic landscape affecting such programs can change rapidly, often resulting in changing monitoring objectives. Consequently, there are no monitoring programs for many fisheries and bycatch species (e.g., Lewison et al., 2004; Roberson et al., 2018), and already established programs can be ill-equipped to address new mandates (Borges, 2015).

We have drawn on our experience with a large fisheries monitoring program to show how similar programs can successfully address multiple objectives while still adhering to a scientifically defensible sampling design. In particular, we focused on how design elements address the logistical constraints imposed on data collection from commercial fishing operations.

Materials and methods

Keys to a successful sampling design

Studies have described the effective design and components of fishery monitoring programs that gather useful

[1] ICES (International Council for the Exploration of the Sea). 2014. Report of the third workshop on practical implementation of statistical sound catch sampling programmes (WKPICS3); Copenhagen, Denmark, 19–22 November 2013. ICES CM 2013/ACOM:54, 109 p. [Available from website.]

data (e.g., Cotter and Pilling, 2007; Vølstad et al., 2014; ICES[1]). These design elements include randomized data collections over spatial and temporal scales (a probability sample), the collection of sufficient data, and the use of stratification and prespecification of sampling intensity to control precision of estimates. In order to meet the data needs of fisheries managers and scientific researchers, a monitoring program design must also make efficient use of available funding, use the time and energy of observers wisely, and maintain observer safety standards.

Monitoring objectives

Data collected by fishery monitoring programs have many uses. These data include collections in support of stock assessments (NRC, 1998), ecosystem-based fisheries management (Gilman et al., 2017), avoidance of bycatch *hot spots* by industry and research partners (Gilman et al., 2006; O'Keefe et al., 2014), and detailed catch accounting (Karp et al., 2019). Because the fisheries data collected by these monitoring programs are used in analyses that meet a wide variety of management and scientific needs, it is often difficult to specify a single objective for these programs. Programs designed to meet the data needs for only one objective could potentially look very different depending on the specifics of both the objective and the target populations. For example, because of differences in the prevalence and spatial distribution of different bycatch species within groundfish fisheries, a monitoring program designed to collect biological samples and data (e.g., otoliths and length measurements) from specimens of a single targeted species might be very different than one designed to monitor bycatch of a less common species. Instead, the objective of these programs should be to collect representative data from the relevant fisheries such that characteristics of catch (e.g., species prevalence and at-sea discard rate) are accurately reflected in the monitoring data.

Probability sampling

A sampling design links the sample data and the sampled population. Unlike other types of sampling, such as proportional-to-size or adaptive sampling designs, a probability sampling design that makes the fewest assumptions about the characteristics of the sampled population (e.g., the catch or length distribution for specific species) is a necessity for most monitoring programs (Cotter et al., 2002).

The basic elements of a probability sample are well known: the target population is identified, and any differences between the target population and the sampling frame (sampled population) are identified. The sampled population is then divided into unambiguously defined sampling units so that each element of the population is in one and only one sampling unit, a method for randomly selecting units is prespecified, and data from the sampling unit are collected (e.g., Cochran, 1977; Thompson, 2012). In fisheries monitoring, it is often not possible to completely list all the sampling units in a population or it is infeasible to randomly select individual units without first selecting a larger grouping (i.e., random sample of fishing events, or hauls, without first selecting fishing trips). Regardless, these design principles can be incorporated into monitoring programs. Sampling of fishing effort can be naturally organized into a hierarchical structure (i.e., trips, then hauls, then samples), and leveraging this structure enables implementation of a probability design and yields efficiencies in sampling. Multistage cluster designs are used to sample each level of the target population independently. This design not only preserves the probability sample by allowing selection probabilities to be specified for each sampling unit but also may decrease logistical difficulties and costs.

Stratification and allocation

Stratification of the target population is used to decrease the variability in estimates or to increase cost effectiveness and logistical efficiency of sampling. Specification of sampling strata, sample allocation among strata, and minimal sample size targets are design components that can be used to develop sampling recommendations, and the methods for these components are well established. However, balancing the sampling needs for different species across different fisheries requires using different approaches to development of sampling recommendations, and, in particular, recommendations of sampling intensity.

When a single objective is clearly defined (e.g., estimation of single-species catch), minimization of variance for a given cost can be used to determine sampling rates, appropriate stratification, and other elements of the sample design (Cochran, 1977). However, for sampling of multiple species, the rate of variance minimization and recommended sample sizes will be different for each species, depending on the variance structure of the species within the catch. For example, Babcock et al.[2] demonstrated that, for species that are less common in the catch, higher sample rates are required to achieve the same precision as that for discards of more common species (i.e., discards of rarer species typically have higher variance between sampling units). Therefore, monitoring programs with multiple objectives need to balance the competing sample size recommendations for different objectives (i.e., species) and accommodate the logistical constraints inherent in sampling commercial catches. One way to address these issues is through the use of a probability sample with a sample size large enough to reasonably assure that all aspects of the monitored fisheries are represented in the collected data, minimizing the number and types of gaps in the data set.

Evaluation of implementation

Evaluating whether a sampling design has been successfully implemented is critical to ensuring that a monitoring program is efficient and that collected data are suitable

[2] Babcock, E. A., E. K. Pikitch, and C. G. Hudson. 2003. How much observer coverage is enough to adequately estimate bycatch?, 29 p. Oceana, Washington, DC. [Available from website.]

Table 1

Summary of the hierarchical structure and randomization of the sampling design used and data collected by the North Pacific Observer Program in 2018. This program of the National Marine Fisheries Service collects data that are representative of fishing activities throughout the federally managed groundfish fisheries off Alaska.

Hierarchy level	Sampling unit definition	Randomization method	Data collected	Constraints	How addressed
Primary	Trip	Binomial selection	Vessel and gear types, port of departure and return, delivery information	Cost	Fee, eliminating individual cost burden (works like insurance)
Secondary	Hauls	Constrained random	Fishing effort (date, location, depth, duration of trawl tows, number of hooks or pots fished)	Fatigue	Random sample of hauls within trips
Tertiary	Portions of catch (gear or weight)	Situation dependent; systematic random preferred	Species composition (proportions), percentage discarded by species	Space and physical ability	Maximization of sample size within constraints, use of multiple samples
Quaternary	Individual catch items (fish)	Systematic random preferred, simple random acceptable, varies with research needs	Length measurements, samples from specimens (e.g., otoliths, stomachs)	Time and availability of species	Annual collaboration with stock assessors to focus sampling on high-priority species

for use in a range of analytic activities. That data originating from observer programs are randomly collected and representative of fishing activities are basic underlying assumptions of many analyses, such as bycatch estimation and stock assessments. Evaluations of these basic assumptions should include assessment of spatial and temporal coverage of fishing activities, adequacy of sampling frames and definitions of sampling units, and effectiveness of stratification definitions.

Implementing these basic design elements can be difficult to achieve in commercial fisheries because of logistical and funding constraints that limit deployments of observers and because of the chaotic environment on deck where observers sample catches. However, in spite of these challenges, designing and implementing a probability-based sampling program is possible. From what we have learned in implementing the successful North Pacific Observer Program of the National Marine Fisheries Service (NMFS), we describe how all these elements and considerations can come together to create a working science-based sampling program.

Results

Case study: North Pacific Observer Program

Fisheries in the 3.77 million km^2 that make up the U.S. exclusive economic zone off Alaska are managed by the NMFS in consultation with the North Pacific Fishery Management Council (NPFMC). The waters off Alaska account for half of all annual U.S. seafood landings. In 2018, 1084 vessels fished with trawl, longline, pot, and jig gear in the U.S. exclusive economic zone off Alaska, catching over 2.3 million metric tons (t) of groundfish worth more than $2.5 billion (Ganz et al., 2019; Fissel et al., 2019). In-season, near real-time accounting of retained and discarded catch for federal fishery quota monitoring is supported in Alaska with data collected by the North Pacific Observer Program (DiCosimo et al., 2010).

The North Pacific Observer Program is the largest in the country; in 2018 over 400 observers and electronic monitoring (camera) systems were deployed for over 40,000 sea days (AFSC[3]). The goal of the North Pacific Observer Program is to collect data that are representative of fishing activities throughout the federally managed groundfish fisheries off Alaska. The North Pacific Observer Program employs a stratified multistage design in which the primary sampling units are individual trips taken by vessels. Within the fishing trip, the secondary sampling units are individual fishing events (hauls, or sets), tertiary sampling units are portions of the catch, and quaternary sampling units are individual fish; these sampling units are collected on the vessel once the observer is deployed (Table 1). Sampling occurs throughout the federally managed groundfish fisheries off Alaska; in 2018 over 58,000 fishing events (hauls) occurred on sampled trips.

Through annual deployment plans (ADPs), the NMFS defines how monitoring tools will be deployed for the

[3] AFSC (Alaska Fisheries Science Center) and Alaska Regional Office. 2019. North Pacific Observer Program 2018 annual report. AFSC Processed Rep. 2019-04, 148 p. [Available from website.]

following year. At the NPFMC October and December meetings, the NMFS prepares a draft and final ADP for NPFMC review. In the draft ADP, multiple competing deployment options, including alternative stratum definitions and sample allocations, are compared. After receiving input from the NPFMC, stakeholders, and the public, the NMFS presents in the final ADP only the preferred design, predicted total and observed fishing effort, and associated selection rates for the coming year (see Ganz and Faunce, 2019).

Sampling design

The North Pacific Observer Program divides the fleet of vessels into 2 broad categories: those that are observed for every trip (full coverage) and those that are not (partial coverage). The first category consists of factory vessels (catcher-processors and motherships, with limited exceptions), catcher vessels participating in management programs that have transferable allocations of prohibited species catch as part of their catch-share program, catcher vessels using trawl gear that have requested full coverage for all fishing within the Bering Sea and Aleutian Islands, and inshore processors receiving or processing walleye pollock (*Gadus chalcogrammus*) caught in the Bering Sea under cooperative agreements (American…1998). These vessels are required to carry one or more observers for all of their trips and to pay for that observer coverage through companies certified by the NMFS to provide observers. In the second category, vessels without factories (catcher vessels) participating in open-access fisheries or in individual fishing quota fisheries carry an observer only on a randomly selected subset of their trips. Observers in partial coverage are provided through a contract between the NMFS and a single provider and are supported through ex-vessel fee revenues from the prior year.

In 2018, electronic monitoring was added as an alternative to observers for fishery monitoring of vessels in the partial coverage category (NMFS[4]). Vessels in the electronic monitoring sampling strata do not carry observers on any fishing trips but are equipped with video camera systems that record all fishing activities. Video files are later reviewed, all catch items are counted and identified to species, and the disposition (retained or discarded) of each item is recorded. Vessels that use non-trawl gear volunteer to participate in electronic monitoring for the upcoming year, and if selected by the NMFS to participate, have electronic monitoring enabled on a random subset of their trips. Vessels also pay the ex-vessel fee to support deployment of electronic monitoring—just as they would if the coverage was by observers. Data from these electronically monitored trips are combined with data from the trips monitored by observers to estimate the catch of each species.

With slight variations between years, the North Pacific Observer Program has stratified the partial coverage fleet for deployment of observers and electronic monitoring systems on the basis of vessel size and gear properties following an initial regression analysis by the NMFS that indicated these were the main drivers of differences in landed catch (Gasper et al., 2019). Vessels under 12 m (40 ft) and those using jig gear operating in the North Pacific Ocean are currently not required to be observed (zero coverage stratum) because they catch few fish and cannot easily and safely accommodate an observer. Stratum definitions must be based on trip qualities known before a trip begins so that selection rates can be properly assigned.

Allocation strategy

There is a strong impetus to maximize efficiency and utility of fishery monitoring data, especially at the trip level, because of stakeholder concerns about costs. The NMFS has focused sample allocation on discarded groundfish and minimization of data gaps, while the NPFMC has expressed a desire for the NMFS to prioritize monitoring of bycatch of Pacific halibut (*Hippoglossus stenolepis*) and Chinook salmon (*Oncorhynchus tshawytscha*), which are prohibited from being retained and sold by certain gear groups. To address these multiple objectives, a 2-step sample allocation process for observer deployment was developed.

First, baseline coverage rates were established following the results of simulated sampling of past fishing events (Gasper et al., 2019). Second, if annual funding is available for sampling effort (sea days converted to fishing trips) above these baseline observer coverage rates, additional effort is allocated according to a blended weighting for optimized sample allocation following Cochran (1977). The blended allocation weighting is the average of the 3 separate optimal allocation weightings derived from discarded total catch of groundfish and catch of the prohibited species Pacific halibut and Chinook salmon. Each weighting takes into account average trip cost and variance for the stratum during the year (Sullivan and Faunce, 2018).

Results of simulated sampling and analyses of past fishing events indicate that, at deployment rates of 15% or greater, the number of gear and area combinations with fewer than 3 trips observed decreased substantially (Gasper et al., 2019). With the use of this 2-step process, the deployment rates employed since 2018 are the combination of the base sampling effort (15% of trips) and additional optimized sampling effort (sea days) allocated to each stratum (NMFS[5]). In contrast, deployment of electronic monitoring is based on voluntary vessel participation and is set at 30% following NPFMC guidance (NMFS[4]).

---

[4] NMFS (National Marine Fisheries Service). 2017. 2018 annual deployment plan for observers in the groundfish and halibut fisheries off Alaska, 20 p. Natl. Mar. Fish. Serv., NOAA, Juneau, AK. [Available from: website.]

[5] NMFS (National Marine Fisheries Service). 2019. 2020 annual deployment plan for observers and electronic monitoring in the groundfish and halibut fisheries off Alaska, 18 p. Natl. Mar. Fish. Serv., NOAA, Juneau, AK. [Available from website.]

Selection of sampling units

Trips Within each partial coverage stratum, the sampling frame consists of all fishing trips that are required by law to be logged into an online application (the Observer Declare and Deploy System, available from website) or through a 24-h call center. Each logged trip represents a sampling unit, and for each sampling unit an observer or electronic monitoring system is assigned on the basis of a random 4-digit number between 0 and 1 generated when the trip is logged. If the number is at or below the selection rate stipulated in the ADP, that trip is selected for monitoring; otherwise it is not. The system in Alaska is 1 of 2 such declaration and selection systems in the country (see Palmer et al., 2016, for a description of the other). In Alaska, although selection rates differ between strata as a result of the allocation process defined in the ADP, each logged trip within a stratum receives the same selection probability. The rate of trip selection within each stratum is held constant for the calendar year.

Hauls Once an observer is deployed, the monetary sampling costs do not change (because the cost of deploying an observer has already been incurred), and estimates of costs are measured by an observer's time and energy. Factors to consider when developing methods for sampling on a vessel include randomizing sample selection, maximizing both the number and size of sampling units selected, and integrating sampling activities into a streamlined observer workflow. Sample size recommendations reflect what is logistically possible, and sampling rates are limited by an observers' access to the catch, the time available to sample, and their physical capacity.

Randomization of haul selection is achieved through the use of 2 alternative tools. The first tool is a modified table of random numbers that assigns a random number of hauls to be sampled followed by a number of hauls to be skipped. The second tool is a table that provides an observer with a 6-h break starting at a random hour of the day. Both of these tools give the observer periods during which non-sampling duties can be completed, provide rest periods for the observer, and decrease the ability of the vessel crew to predict which hauls will be sampled (as would be possible if a systematic design was used). Because it is required that all large organisms, such as marine mammals and sharks, caught in bycatch be available to the observer for all sampled hauls, randomization at the haul or set level is critical to accurate estimation of bycatch for these species.

Catch An observer is responsible for assessing the fishing activities on a vessel and for determining how sampling is best conducted. They define the population to be sampled, establish a sampling frame, define the sampling units, and randomly select a set of sampling units for data collection at each level of the hierarchy. Sampling frames and units are defined differently depending on fishing gear, observer experience, fishing operations and catch handling practices on a particular vessel, environmental conditions, and the general situation on the particular vessel. Observers are trained to maximize the fraction of the catch sampled in order to increase detection probabilities of rare species (e.g., seabirds, deep sea corals, and sharks) and to minimize sampling variance associated with haul-specific estimates of species-specific bycatch.

The sampling methods used by observers vary between fishing trips and between hauls within the same fishing trip, as a result of the varied sampling environments. Options for defining sampling frames and units vary dramatically across fisheries, and opportunities for randomizing selection of those units can be constrained by the gear used and by deck or factory configurations of individual vessels. Hence, the ability to adjust sampling in response to changing fishing conditions and vessel operations is critical to successful implementation of a credible randomized sampling design.

Since 2010, observers have recorded the type of sampling unit and the selection process used to randomize collection of their catch composition samples. Samples can be selected by using *strictly randomized methods* (simple or systematic random sampling in which sampling units are unambiguously defined), *randomized methods* in which randomization is incorporated into sample selection but either sampling units are not well defined or the entire catch is not available to be sampled (with a large difference between sampling frame and target populations), or non-random methods (which include both *opportunistic* sampling unit selection and complete enumeration [i.e., a *census* of the catch]).

Despite the range and diversity of fishing operations in Alaska, fishing operations that present similar sampling challenges can be grouped into 3 broad categories: 1) fixed-gear vessels (longline and pots) of any type, 2) catcher-processors that use trawl nets, and 3) catcher vessels that use trawl nets. The catch on fixed-gear vessels is sampled at one location (the roller on longline vessels or the sorting table or tote on pot vessels) in an orderly fashion as the catch is either brought on board or discarded. The observer defines sampling units and the sampling frame in terms of units of retrieved gear, such as segments of longline gear or predetermined numbers of pots. Catch is tallied at a vessel's rail from a systematic random selection of units (collection by using a strictly randomized method). Although defining sampling frames and randomizing selection of sampling units are straightforward, observers on these vessels work on deck often in severe weather conditions (extreme cold), limiting the size of individual samples. On larger vessels, gear retrieval can take in excess of 10 h, and in order to ensure adequate time to process samples (e.g., time to sort catch items and to collect weight and length data) and in order to allow an adequate rest period, time management can become a component of an observer's sampling design.

On trawl vessels, catch comes aboard in a more disorderly fashion than the process used on fixed-gear vessels, and sampling units and frames are defined in terms of catch weight. On catcher-processor vessels that use trawl nets, hauls can be in excess of 100 t. These catcher-processor and mothership vessels have a factory, and an observer has access to tools not available on catcher vessels. These tools include a motion-compensated platform scale, a flow scale that provides a continuous and cumulative weight of catch as it passes over the scale on a conveyor belt, and a dedicated sample station with specified minimal space requirements (4.5 m^2; Equipment…2019) to provide an area to store and process collected samples. Because catch passes into the factory on a series of conveyor belts past the observer's sample station, randomization is generally achieved by using systematic random sampling methods (strictly randomized collection). Although an observer works in the vessel factory and not on deck, sampling is conducted in an industrial setting designed and operated to move catch through the processing plant to the freezer as quickly as possible, limiting the amount of time an observer has to select and process samples.

Catcher vessels that use trawl nets often pose the most challenging sampling design scenarios. Similar to catcher-processors that use trawl nets, haul weight on these vessels can be over 100 t; however, catches typically range between 10 t and 70 t. Observers work on deck in a dynamic environment, and catch is available to the observer in an unorganized manner. On some catcher vessels, catch comes aboard directly into the trawl alleyway, and the net is often reset before the completion of sampling. Only a few tools, such as spring scales or small platform scales, are available to an observer, and totes or bin space for storing individual samples is limited. An observer may have little time to collect samples before the catch is stowed. Observers who work on deck are exposed to weather and sample under dynamic conditions in which the catch shifts within the trawl alley. In spite of this chaotic sampling environment on deck, methods have been developed to randomize sample collections on catcher vessels, and alternative methods have been explored (Cahalan et al., 2016). For example, some catcher vessels may dump catch from the codend of the trawl net onto the deck where the crew sorts and discards unwanted catch or fish can flow directly into the hold.

Using time as a proxy for fish weight, observers can define sampling units and a systematic sampling design and then randomly select proxy sampling units until the entire catch has been moved below deck (strictly randomized collection). In the more common situation, however, an observer samples catch directly from the trawl alley. An observer might randomly choose several places adjacent to the alleyway to select samples (collection by using a randomized method), or a captain might determine where the observer can safely collect a single sample (opportunistic collection). Scenarios in which observers resort to opportunistic sampling pose the most challenges, in terms of decisions on how to best define sampling units and to randomize selections given the working environment on deck.

Individual fish Lastly, individual fish are randomly selected from the samples of catch by using either a systematic or simple random sampling design to select fish of a specified species. Biological data, such as fish lengths, as well as otoliths and tissue samples (stomachs, gonads, and fins taken for genetics analysis) are collected from the randomly selected individuals. Bycatch of all prohibited species, including the Pacific halibut, the Pacific herring (*Clupea pallasii*), all Pacific salmon species (*Oncorhynchus* spp.), steelhead (*O. mykiss*), all king crab species (*Lithodes* spp. and *Paralithodes* spp.), and Tanner crab species (*Chionoecetes* spp.), have biological data collected by the observer at sea. Through an annual process, the scientists at the NMFS Alaska Fisheries Science Center, working with the North Pacific Observer Program, determine the species for which, and quantities of, collections of biological data and specimens will be specified in the observer sampling manual for the following year. The collection methods for most species are integrated into the sampling hierarchy.

Landed catch is not sampled by observers in the North Pacific Ocean with the single exception of the fisheries that target walleye pollock and for which there has been long-standing concerns over salmon bycatch (Faunce, 2015). To address these concerns, a separate, parallel sampling design has been implemented in which trips on catcher vessels within these fisheries are monitored at the point of shoreside delivery. From these monitored shoreside landings, observers count and obtain biological data from salmon caught as bycatch.

Program evaluation

The North Pacific Observer Program evaluates the past performance of the implementation of the sampling design and provides the results in an annual report to the NPFMC in June. The results of the evaluations in annual reports inform decisions about recommendations for improving the deployment of observers and electronic monitoring systems in the next ADP cycle. This annual cycle of review, revision, and refinement leads to continual improvement of stakeholder engagement and deployments of observers and electronic monitoring systems. Here, we focus on review of the 2 stages of the sampling design with the most utility to other programs: deployment of observers (trip selections) and catch sampling (Table 2).

Observer and electronic monitoring deployments The ability to deploy observers at prespecified rates is evaluated in annual reports by comparing the rate of trip selection programmed into the online trip logging application to the initial selection rate for each sampling stratum (Table 3). The magnitude of non-response and actual on-the-ground coverage rates are assessed by comparing the proportion of trips observed in each stratum (actual coverage rate) with the prespecified rate (Table 3).

Temporal and spatial patterns of observed trips should be representative of the total fishing effort to provide inference about catch and bycatch. Avoidance of observer coverage during periods of high bycatch or during fishery openings typically is reflected in departures from expected patterns of observer coverage over time. Differences in temporal fishing patterns between observed and unobserved fishing trips are assessed by comparing the cumulative distributions of the number of observed trips and the expected number of observed trips given the deployment

Table 2

Structure of the sampling hierarchy used in 2018 by the North Pacific Observer Program, National Marine Fisheries Service, to monitor groundfish fisheries off Alaska and to evaluate on an annual basis the performance and design of sampling. The evaluation method informs efforts to improve the deployment of observers and electronic monitoring systems in the next annual deployment plan (ADP). Action items address questions of whether or not collected data are representative of fishing activities.

Question	Representativeness question and evaluation methods	Action item
Primary sampling units: trips		
How well was fishing effort predicted in the ADP last year?	Comparison of actual fishing effort vs. distribution of simulated outcomes from ADP	If actual effort not within simulated distribution, corrective or alternative methods for predicting future fishing effort are made, and the density value chosen to set overage risk is reassessed.
Is the realized observation rate reasonable given the selection rate?	Binomial test and evaluation of trip cancellation rates (selected, not selected)	Disproportionate cancellations may be evidence of a deployment effect. Policy allowing multiple trips to be logged is recommended to be reevaluated.
Is the temporal distribution of observed trips reasonable given fishing activity?	Cumulative distribution of selected trips vs. all trips	Together with cancellation rates, departure may be evidence of a deployment effect. Evaluated with opening and closing of fisheries.
Are data representative in space?	Probability densities resulting from hypergeometric distribution	Single test results not given much value in isolation. If the suite of results point to observer effect, look at prior years. If such a pattern found, recommend higher coverage rates for strata.
Is the information on observed trips comparable to that on unobserved trips?	Permutation test	
Secondary sampling units: hauls		
Is the sample rate consistent with expectations outlined in the observer sampling manual (AFSC[1])?	Inspection of proportion of hauls sampled per trip	Feedback to the observer both at sea and during data quality control process.
Tertiary sampling units: portions of total catch		
Are randomized sample selection methods used?	Inspection of proportions of hauls sampled using different randomization schemes	Feedback to observer and additional coaching related to specific sampling situations. Recommendations for additional training may be made.
Are multiple samples collected?	Inspection of numbers of samples per haul for different sampling scenarios	
Are sampling units well defined?	Inspection of sample unit definitions recorded by observers. Inspection of consistency of size of samples (catcher vessels)	
Quaternary sampling units: individual catch items (fish)		
Are randomized sample selection methods used?	Inspection of randomization schemes used to select fish; amount of data deleted because of inappropriate selection methods	Feedback to observer and additional coaching related to specific sampling situations. Recommendations for additional training may be made.
Are lengths of individual randomly selected fish consistent with mean size of fish in the catch sample (tertiary sample)?	Comparisons of mean fish weights by species	

[1] AFSC (Alaska Fisheries Science Center). 2018. 2018 observer sampling manual. North Pac. Groundfish Obs. Program, Fish. Monit. Analysis Div., AFSC, Seattle, WA. [Available from website.]

Table 3

Table of results from evaluation of trip selection in 2018 for 3 main gear-based strata for the category of vessels with partial coverage. These results illustrate the ability of the North Pacific Observer Program, National Marine Fisheries Service, to assess whether or not random selection was achieved in sampling of the groundfish fisheries off Alaska by using the programmed rate of trip selection, realized rate expected, and realized rate. Programmed and expected selection rates are derived from the trip logging application, and realized rates are derived from landing reports and the observer program data on observer deployments. The observer program does not report selection rates realized beyond the nearest tenth of a percentage. Entries in the "Within expected range?" column refer to whether or not the actual value was within the 95% binomial confidence limits of the numbers in the "Programmed value" column. Evaluation of whether or not the programmed rate was implemented correctly includes tests of whether or not the random number generator of the trip selection application was functioning correctly. Evaluation of the expected rate includes tests of whether or not the trip selection application rate is equal to the programmed rate after user influence (e.g., trip cancellations and trip duration changes). Evaluation of the realized rate includes tests of whether or not the number of trips actually monitored was equal to the programmed rate. For complete results, see Ganz et al., 2019.

Stratum	Rate	Programmed value	Actual value	Within expected range?
Hook and line				
	Programmed	17.26	16.32	Yes
	Expected	17.26	17.67	Yes
	Realized	17.26	15.5	No
Hook and line electronic monitoring				
	Programmed	30.00	30.15	Yes
	Expected	30.00	34.10	No
	Realized	30.00	22.7	No
Pot				
	Programmed	16.21	16.38	Yes
	Expected	16.21	16.45	Yes
	Realized	16.21	15.5	Yes
Trawl				
	Programmed	20.18	20.07	Yes
	Expected	20.18	20.88	Yes
	Realized	20.18	20.3	Yes

rate, the total number of fishing trips that occurred, and the 95% confidence interval based on Bernoulli selection of individual trips (AFSC[3]). Deviations of the number of observed trips from the expected number of observed trips that coincide with fishery openings or closings are evidence of deployment effects.

Differences in the spatial distribution of the sample (observed trips) and the total population (all fishing trips in sampled strata) are also evaluated. The probability that the sample contains the number of trips actually observed in a geographic area given the number of trips that occurred in the area (from landings data) and the total number of trips fished is computed for each area (within each sampling stratum) and used to assess the spatial representativeness of observed trips. A low probability of the sample containing the number of trips actually observed (or a number of trips greater than expected) is indicative of departures from representative deployments. However, because multiple areas are assessed simultaneously, we must avoid the risk of making erroneous conclusions as a result of multiple comparisons (Greenland et al., 2016). Therefore, a finding of a greater number of areas with probabilities lower than our nominal error rate predicts (i.e., if the probability of the sample containing the number of trips actually observed is less than 0.05) is interpreted to be evidence of an overall deployment effect. For example, in an assessment of 20 areas, we expect that one of those areas could have a low (<0.05) probability (or one area with a more extreme result). If more areas were found in the current and prior years, then it is recommended that coverage rates for that stratum be increased.

Characteristics of observed and unobserved trips are compared to evaluate whether observed fishing trips are similar to (and exchangeable with) unobserved fishing trips. Currently, 6 characteristics that are known for both the observed and unobserved trips are assessed: trip duration, number of geographic areas fished, total landed weight, number of species landed, proportion of the landing accounted for by the predominant species, and vessel length (Ganz et al., 2019). Permutation tests are used to test the

Table 4

Percentage of hauls sampled in 2018 by the North Pacific Observer Program, National Marine Fisheries Service, within each gear and vessel type that achieved the specified randomization. In strictly randomized methods of sampling, sampling units are unambiguously defined, all units are available to be selected, and sample selection is either simple random or systematic random. In randomized methods, randomization is introduced to the sampling process, but sampling units cannot be well defined, all units are not available to the observer, or the entire sampling unit is not collected. Opportunistic sample unit selection is a non-random method. A census, another type of non-random method, is a complete enumeration or weight for the entire haul.

		Percentage of hauls by gear type			
Vessel type	Method	Pelagic trawl	Non-pelagic trawl	Longline	Pot
Catcher	Strictly randomized	33.5%	38.7%	91.5%	91.9%
	Randomized	42.5%	21.0%	1.2%	1.1%
	Opportunistic	24.0%	40.0%	0.8%	2.8%
	Census	0.1%	0.2%	6.4%	4.3%
Catcher-processor	Strictly randomized	98.4%	96.9%	99.1%	99.1%
	Randomized	0.1%	0.5%	0.5%	
	Opportunistic	0.2%	2.5%	0.3%	
	Census	1.4%	0.1%	0.1%	0.9%

probability that the observed difference in trip metrics between observed and unobserved trips derives from the same population. A preponderance of low P-values among tests is interpreted as evidence of an observer effect. However, low sample sizes can result in spurious results, and large sample sizes can lead to low P-values for very small differences. For these reasons, the North Pacific Observer Program presents the effect size (magnitude of the difference in the metric for observed and all trips) as well as the P-values to increase interpretability of test results with low P-values (to avoid confusion between no difference and no significant difference, as per Greenland et al., 2016).

Catch sampling In spite of the logistical difficulties of sampling on commercial fishing vessels, most hauls are sampled by using strictly randomized or randomized methods (Table 4). As expected on longline and pot vessels, close to 100% of hauls are sampled by using strictly randomized methods (Table 4). The most difficult sampling situation for observers is on catcher vessels that use trawl nets because defining sampling units and randomizing which units to select is complicated by catch handling and vessel operations. In spite of the logistical difficulties, observers resorted to opportunistic sampling in fewer than half the sampled hauls on catcher vessels that use trawl nets, and 60% of hauls were sampled with some randomization incorporated into sample selection.

Discussion

The Magnuson–Stevens Fishery Conservation and Management Reauthorization Act of 2006 (Magnuson–Stevens…2018), the Marine Mammal Protection Act (Marine…2018), and the Endangered Species Act (Endangered…2018) of the United States mandate not only collection of data for fisheries management of target species but also monitoring of the effects of fisheries on protected species (Moore et al., 2009). The observer programs tasked with this monitoring mandate have multiple, disparate sampling objectives that often change over time. In Alaska, we have chosen not to design the monitoring program to a specific fishery or to meet specific target precision goals. Because species-specific catches (including discards) have different underlying distributions and population structures that vary both with the time of year and in geographic space, it can be difficult to design sampling programs to meet species-specific (e.g., achieve a target-level precision for species-specific bycatch estimates) or short-term goals without decreasing the utility of the collected data for estimation specific to other species or analytic objectives. Sampling schemes that are optimal for less common species will not necessarily be optimal for target species. A diverse group of researchers and stakeholders, including statisticians, stock assessment scientists, social scientists (economists and anthropologists), members of the fishing industry, and fisheries managers, use data collected by observers, and they all have very different data needs and objectives. In all cases, analytic data users, and fisheries assessment and management, are affected by sparse and missing data. This issue of a lack of or low quality of data is particularly problematic for low-effort fisheries in which observers may not be able to sample from specific strata (i.e., areas and time of year).

The challenge in designing sampling programs that collect representative data useful to researchers and fisheries managers is surmountable. Sampling designs that do not make assumptions about the underlying population being sampled are preferred to model-based designs that incorporate assumptions (Cotter and Pilling, 2007). A hierarchical probability-based design for the federally managed fisheries off Alaska is used because it is feasible and necessary (e.g., it is impossible to randomly deploy observers to individual hauls or by weight of catch). This type of design makes efficient use of observer funds and sampling effort, and it allows flexibility to the data user in the estimators they chose. Because data collected under a probability-based sampling scheme can be used in both model-based estimation and more traditional design-based estimation processes, an analyst can choose to use design-based estimators when estimating bycatch of relatively common groundfish species or to use a model-based estimator (e.g., zero-inflated parametric models) to estimate the bycatch of a rare species (Allen et al., 2002; Dietrich et al., 2009; Stock et al., 2019).

Stratification decisions in the sampling design for deployment of observers or electronic monitoring systems can be based on logistical considerations aimed at reducing cost (e.g., Sun and Fine, 2016). Defining sampling strata on the basis of fishery (i.e., fishery-specific deployment of observers) can be problematic because factors describing the sampling units and strata need to be defined prior to fishing (Cochran, 1977), and although the intended target species is known prior to fishing, the realized target is not. Melvin et al. (2019) and Moore et al. (2009) have advocated fishery-specific solutions to monitoring for bycatch of seabirds and marine mammals, respectively, but we intentionally avoid inclusion of species in the stratum definitions for sampling in the North Pacific Ocean with the notable exception of dockside monitoring in the trawl fishery that targets walleye pollock. Because fisheries are not clearly defined prior to fishing and the North Pacific Observer Program cannot deploy into specific fisheries, the terms *fishery* or *métier*, which are often conflated, are not used (Ulrich et al., 2012).

Stratification decisions in the sampling design for deployment of observers or electronic monitoring systems can be based on statistical considerations aimed at reducing variance in the estimator (e.g., Wigley and Tholke, 2018). Although this objective is achieved by maximizing the differences between strata and minimizing differences within strata (Thompson, 2012), if poorly constructed, stratified sampling designs can lead to estimates with variances greater (not lower) than would be achieved with an unstratified design and the same overall sample size. In addition, for situations in which multiple estimation objectives (e.g., multiple species) exist, sample size and allocation decisions based on the minimization of variance must rely on assignment of relative importance of objectives, de facto decisions of which species are important and which are not. For example, Miller et al. (2007) used variance minimization to allocate sampling effort among gear-based strata to identify the ideal stratification of "small" and "large" vessels in the North Pacific Observer Program in the North Pacific Ocean. They synthesized multiple sampling objectives (i.e., species-specific estimates of catch and discards) into a single objective by weighting the individual species according to the management priorities at the time. This weighting required prioritizing the importance of species and fisheries not only in the current year but also in future years when sampling will take place.

Although this method of Miller et al. (2007) decreased variance of the species- and fishery-specific estimates of catch and discards that were investigated, this sample allocation may not be optimal or even appropriate for other data users who have sampling goals that are not related to catch or discards (e.g., collection of tissue samples for genetic analysis). Unlike the approach used by Miller et al. (2007), the method used by the North Pacific Observer Program addresses multiple species (not only a few high-profile species) across fisheries and multiple estimation goals, including catch estimation for in-season quota management, catch-at-age estimation used in the stock assessment process, estimation of bycatch of rare species such as marine mammals, and identification of population substructure through genetic analysis of tissue samples.

The question of how much coverage is enough is reasonable to ask but difficult to answer for any monitoring program. Babcock et al.[2] used simulation to show how variance decreases with increasing sample size and how more samples are required for species with increasing rarity, and they advocated setting precision-based targets to guide analyses of sample sizes. This type of analysis was used by NMFS (2004) to set a precision target for U.S. fisheries. The observer program in New England established a 30% coefficient of variation for its species within numerous fishery complexes. The result was a sample size that the program could not afford, and the program was scaled back. However, the NMFS was challenged in a legal suit on the grounds that it failed to achieve its own performance standard. In an eventual compromise, species that constituted small portions of total catch or for whom the recommended sampled sizes would have been large were eliminated from sample size recommendations and other filtering rules, prior to scaling back to affordable levels (Wigley et al.[6]; Wigley and Tholke, 2018).

In the North Pacific Observer Program, a different approach was taken to setting target sample rates and to making sample allocation decisions. The North Pacific Observer Program does not set coverage rates on the basis of predetermined precision criteria. Instead, it evaluates alternative stratifications and allocations in its draft ADP review process and compares the relative performance of each against the stratification used in the previous year and an equal allocation of available sample effort. Performance scores are based on the severity of data gaps generated—the greater the gap, the lower the score

[6] Wigley, S. E., P. J. Rago, K. A. Sosebee, and D. L. Palka. 2007. The analytic component to the Standardized Bycatch Reporting Methodology Omnibus Amendment: sampling design end estimation of precision and accuracy, 2nd ed. Northeast Fish. Sci. Cent. Ref. Doc. 07-09, 156 p. [Available from website.]

(NMFS[7]). Selection of the current baseline rate of 15% was derived in large part from an analysis that evaluated how the probability of having no data in a catch estimation post-stratum (defined by the dominant species in the catch during periods of 2–4 weeks) changed over a range of deployment rates (Gasper et al., 2019).

By instituting a baseline rate, the North Pacific Observer Program effectively curtailed the ability for managers to annually tailor coverage to meet short-term goals because only funds available after the baseline rate threshold had been reached were available for tailored use. Instead, the establishment of the base rate by the NMFS in Alaska has helped ensure that representative data are collected by the North Pacific Observer Program. Doing so also has created incentive for the NMFS and the NPFMC to obtain funding for coverage above 15% and retained the ability for targeted deployment to meet specific and often changing management priorities.

This concept of the use of different deployment strategies for observer programs is not a new one. The most striking example is when observers are also used to aid in enforcement programs. In Alaska, observers provide an invaluable resource for law enforcement agents by providing statements of potential violations (Porter, 2010). Furlong and Martin (2001) describe a program in which observer deployment is optimized to be greater in proportion to a vessel's demonstrated or suspected violation rate in cases where the deployment cost of an observer is directly borne by the vessel on a trip-by-trip basis. In this system, the incentive to comply is tied to the cost of observation. Such a deployment scheme would tackle the largest potential sources of bias, but it would not result in representative data at the fishery level (Benôit and Allard, 2009). The partial coverage fleet in Alaska funds observer coverage like insurance, wherein all participants pay into the program but only those that are selected have observer coverage. This funding system removes the economic penalties normally associated with industry-funded monitoring programs, and in our opinion greatly reduces the incentive for vessels to change fishing behavior when observers are onboard. Although evidence of this observer effect has been found in the past (Faunce and Barbeaux, 2011), recent complaints about observer deployment are often limited to new participants in a program and are generally focused on the logistical aspects of observation, such as the need for bunk space, food, or life-raft capacity.

To address the logistical difficulties of physical observation, the NPFMC began implementing electronic monitoring on longline vessels in 2014 and has expanded its use every year since. The NPFMC elected to use electronic monitoring for bycatch estimation rather than the less cumbersome compliance monitoring (monitoring only for compliance with regulations such as no discards; e.g., the use of electronic monitoring in Europe to help with enforcement of a discard ban; Borges et al., 2016). Vessels participating in electronic monitoring constitute their own deployment and estimation stratum within the current sample design of the North Pacific Observer Program (Ganz et al., 2019). However, electronic monitoring does not include collection of biological tissues or weight information; therefore, biological samples and data must be obtained elsewhere or inferred. Specifically, on vessels in the electronic monitoring stratum, staff review the video imagery of fishing activities recorded by camera systems, documenting the number and disposition (retained or discarded) of each species. The number of at-sea discards for these strata is estimated by using data derived from electronic monitoring in combination with species-specific mean weights of fish from the observed strata.

Although we have focused here on the deployment of observers, this design is equally applicable to the design of electronic monitoring programs and incorporation of electronic monitoring into observer-based monitoring programs. The reliance on other sampling strata to represent the electronic monitoring stratum highlights the need to set observer coverage rates so that data gaps are minimized and the necessary data elements from the observed strata can be used to "complete" data collected in the electronic monitoring strata, effectively reducing the number of gaps in biological data (inclusive of fish length and weight).

Although deployment rates in Alaska are set to ensure coverage in time and space, the intensity of sampling at sea is largely determined by the sampling environment. Once an observer is deployed, sampling design considerations shift to include the abilities of the observer, and decisions are based on what can be physically and logically achieved given that sampling occurs on vessels actively engaged in commercial fishing. The ability to adjust sampling in response to changing fishing conditions and vessel operations is critical to successful implementation of a credible randomized sampling design. Defining sampling units and having access to those sampling units in order to randomize selections is difficult in many situations. In spite of this challenge, the majority of samples are collected by using credible randomization methods.

Fully randomized designs have the added advantage of allowing sampling to be evaluated at all levels. Instituting randomized trip selection through the restructure of the North Pacific Observer Program in 2013 resulted in the ability to evaluate how well observers and electronic monitoring systems were deployed. Patterns and trends in observer coverage in the previous year are used to assess the effectiveness of the North Pacific Observer Program and to make recommendations for changes to be instituted the following year (as part of the processes for annual review and development of an ADP). Similarly, explicitly capturing information about how observers sample while they are deployed on fishing vessels allows evaluation of the effectiveness of sampling in terms of both accuracy of collected data and statistical robustness of sampling methods. These data are used to provide feedback directly to the observer after deployment and to adjust training and sampling methods for future years.

[7] NMFS (National Marine Fisheries Service). 2019. Draft 2020 annual deployment plan for observers and electronic monitoring in the groundfish and halibut fisheries off Alaska, 16 p. Natl. Mar. Fish. Serv., NOAA, Juneau, AK. [Available from website.]

Randomization, which has long been known to lower costs of sampling (relative to conducting a census of a population), results in representative data that can be used for a variety of purposes and estimators. Our experiences with the North Pacific Observer Program in Alaska have indicated that randomization can be achieved at each level and that these theoretical benefits can be realized in practice. The trip selection process used in Alaska has proven to be a robust method for the randomization of deployments of observers and electronic monitoring systems and, even though active fishing vessels are not perfect sampling platforms, credible at-sea sampling methods can be designed and implemented. We are not aware of another fishery monitoring program that makes annual evaluations of its performance and determines whether sampling objectives were met and whether departures from randomization could be detected. This annual cycle of evaluation and improvement of sampling methods is unique to Alaska and contributes to the success of the North Pacific Observer Program and the positive state of sustainable fisheries management in Alaska. The success of the North Pacific Observer Program in implementing random sampling while using state-of-the-art technology and efficiently collecting data for a variety of uses indicates that this observer program can be used as a model for other fishery monitoring programs with similar objectives.

Acknowledgments

We thank the many observers who are deployed to commercial fishing vessels that target groundfish species in Alaska and the staff of the Fisheries Monitoring and Analysis Division of the Alaska Fisheries Science Center who support observers and make observer data available to our many data users. Our gratitude goes to our colleagues J. Gasper and B. Mason, whose discussions have helped to form this manuscript, and to S. Lowe, C. Tribuzio, and 2 anonymous reviewers whose contributions greatly improved this manuscript.

Literature cited

Allen, M., D. Kilpatrick, M. Armstrong, R. Briggs, G. Course, and N. Pérez.
2002. Multistage cluster sampling design and optimal sample sizes for estimation of fish discards from commercial trawlers. Fish. Res. 55:11–24. Crossref

American Fisheries Act. Pub. L. 105-277, div. C, title 2. 112 Stat. 2681 (21 October 1998). [Available from website.]

Benôit, H. P., and J. Allard.
2009. Can the data from at-sea observer surveys be used to make general inferences about catch composition and discards? Can. J. Fish. Aquat. Sci. 66:2025–2039. Crossref

Borges, L.
2015. The evolution of a discard policy in Europe. Fish Fish. 16:534–540. Crossref

Borges, L., L. Cocas, and K. N. Nielsen.
2016. Discard ban and balanced harvest: a contradiction? ICES J. Mar. Sci. 73:1632–1639. Crossref

Cahalan, J., C. Faunce, J. Bonney, and R. Swanson.
2016. A field test of fisheries observer sampling methods for estimation of at-sea discards. Fish. Res. 174:219–233. Crossref

Cochran, W. G.
1977. Sampling techniques, 3rd ed., 448 p. John Wiley and Sons Inc., New York.

Cotter, A. J. R., and G. M. Pilling.
2007. Landings, logbooks and observer surveys: improving the protocols for sampling commercial fisheries. Fish Fish. 8:123–152. Crossref

Cotter, A. J. R., G. Course, S. T. Buckland, and C. Garrod.
2002. A PPS sample survey of English fishing vessels to estimate discarding and retention of North Sea cod, haddock, and whiting. Fish. Res. 55:25–35. Crossref

Davies, S. L, and J. E. Reynolds (eds.).
2003. Guidelines for developing an at-sea fishery observer programme. FAO Fish. Tech. Pap., 414, 116 p. FAO, Rome.

DiCosimo, J., R. D. Methot, and O. A. Ormseth.
2010. Use of annual catch limits to avoid stock depletion in the Bering Sea and Aleutian Islands management area (Northeast Pacific). ICES J. Mar. Sci. 67:1861–1865. Crossref

Dietrich, K. S., J. K. Parrish, and E. F. Melvin.
2009. Understanding and addressing seabird bycatch in Alaska demersal longline fisheries. Biol. Conserv. 142:2642–2656. Crossref

Endangered Species Act. 16 U.S.C. 1531–1544 (2018). [Available from website.]

Equipment and operational requirements. 50 C.F.R. 679.28 (2019). [Available from website.]

Faunce, C. H.
2015. Evolution of observer methods to obtain genetic material from Chinook salmon bycatch in the Alaska pollock fishery. NOAA Tech. Memo. NMFS-AFSC-288, 28 p.

Faunce, C. H., and S. J. Barbeaux.
2011. The frequency and quantity of Alaskan groundfish catcher-vessel landings made with and without an observer. ICES J. Mar. Sci. 68:1757–1763. Crossref

Fissel, B., M. Dalton, B. Garber-Yonts, A. Haynie, S. Kasperski, J. Lee, D. Lew, A. Lavoie, C. Seung, K. Sparks, et al.
2019. Stock assessment and fishery evaluation report for the groundfish fisheries of the Gulf of Alaska and Bering Sea/Aleutian Islands area: economic status of the groundfish fisheries off Alaska, 2017, 320 p. [Available from website.]

Furlong, W. J., and P. M. Martin.
2001. Observer deployment in the fishery and regulatory self-enforcement. *In* Microbehavior and macroresults: proceedings of the tenth biennial conference of the International Institute of Fisheries Economics and Trade; Corvallis, OR, 10–14 July 2000 (R. S. Johnston and A. L. Shriver, eds.), 8 p. Int. Inst. Fish. Econ. Trade, Corvallis, OR. [Available from website.]

Ganz, P., and C. H. Faunce.
2019. An evaluation of methods used to predict commercial fishing effort in Alaska. NOAA Tech. Memo. NMFS-AFSC-395, 19 p.

Ganz, P., S. Barbeaux, J. Cahalan, J. Gasper, S. Lowe, R. Webster, and C. Faunce.
2019. Deployment performance review of the 2018 North Pacific Observer Program. NOAA Tech. Memo. NMFS-AFSC-397, 73 p.

Gasper, J., C. Tide, G. Harrington, J. Mondragon, J. Cahalan, S. Bibb, and G. Merrill.
2019. Supplement to the environmental assessment for restructuring the program for observer procurement and deployment in the North Pacific. NOAA Tech. Memo. NMFS-F/AKR-19, 168 p.

Gilman, E. L., P. Dalzell, and S. Martin.
2006. Fleet communication to abate fisheries bycatch. Mar. Policy 30:360–366. Crossref

Gilman, E., M. Weijerman, and P. Suuronen.
2017. Ecological data from observer programmes underpin ecosystem-based fisheries management. ICES J. Mar. Sci. 74:1481–1495. Crossref

Greenland, S., S. J. Senn, K. J. Rothman, J. B. Carlin, C. Poole, S. N. Goodman, and D. G. Altman.
2016. Statistical tests, *P* values, confidence intervals, and power: a guide to misinterpretations. Eur. J. Epidemiol. 31:337–350. Crossref

Hilborn, R., J. M. Orensanz, and A. M. Parma.
2005. Institutions, incentives and the future of fisheries. Philos. Trans. R. Soc. Lond., B 360:47–57. Crossref

Karp, W. A., M. Breen, L. Borges, M. Fitzpatrick, S. J. Kennelly, J. Kolding, K. N. Nielsen, J. R. Viðarsson, L. Cocas, and D. Leadbitter.
2019. Strategies used throughout the world to manage fisheries discards—lessons for implementation of the EU landing obligation. *In* The European landing obligation: reducing discards in complex, multi-species and multi-jurisdictional fisheries (S. Uhlmann, C. Ulrich, and S. J. Kennelly, eds.), p. 3–26. Springer, Cham, Switzerland.

Lewis, S. G., and M. Boyle.
2017. The expanding role of traceability in seafood: tools and key initiatives. J. Food Sci. 82(S1):A13–A21. Crossref

Lewison, R. L., L. B. Crowder, A. J. Read, and S. A. Freeman.
2004. Understanding the impacts of fisheries bycatch on marine megafauna. Trends Ecol. Evol. 19:598–604. Crossref

Magnuson–Stevens Fishery Conservation and Management Reauthorization Act of 2006. 16 U.S.C. 1801–1891d (2018). [Available from website.]

Marine Mammal Protection Act. 16 U.S.C. 1361–1421 (2018). [Available from website.]

Melvin, E. F., K. S. Dietrich, R. M. Suryan, and S. M. Fitzgerald.
2019. Lessons from seabird conservation in Alaskan longline fisheries. Conserv. Biol. 33:842–852. Crossref

Miller, T. J., J. R. Skalski, and J. N. Ianelli.
2007. Optimizing a stratified sampling design when faced with multiple objectives. ICES J. Mar. Sci. 64:97–109. Crossref

Moore, J. E., B. P. Wallace, R. L. Lewison, R. Žydelis, T. M. Cox, and L. B. Crowder.
2009. A review of marine mammal, sea turtle and seabird bycatch in USA fisheries and the role of policy in shaping management. Mar. Policy 33:435–451. Crossref

NMFS (National Marine Fisheries Service).
2004. Evaluating bycatch: a national approach to standardized bycatch monitoring programs. NOAA Tech. Memo. NMFS-F/SPO-66, 108 p.

NRC (National Research Council).
1998. Improving fish stock assessments, 177 p. Natl. Academy Press, Washington, DC.

O'Keefe, C. E., S. X. Cadrin, and K. D. E. Stokesbury.
2014. Evaluating effectiveness of time/area closures, quotas/caps, and fleet communications to reduce fisheries bycatch. ICES J. Mar. Sci. 71:1286–1297. Crossref

Palmer, M. C., P. Hersey, H. Marotta, G. R. Shield, and S. B. Cierpich.
2016. The design and performance of an automated observer deployment system for the northeastern United States groundfish fishery. Fish. Res. 179:33–46. Crossref

Porter, R. D.
2010. Fisheries observers as enforcement assets: lessons from the North Pacific. Mar. Policy 34:583–589. Crossref

Stock, B. C., E. J. Ward, J. T. Thorson, J. E. Jannot, and B. X. Semmens.
2019. The utility of spatial model-based estimators of unobserved bycatch. ICES J. Mar. Sci. 76:255–267. Crossref

Roberson, L. A., J. J. Kiszka, and J. E. M. Watson.
2018. Need to address gaps in global fisheries observation. Conserv. Biol. 33:966–968. Crossref

Sullivan, J., and C. Faunce.
2018. Alternative sampling designs for the 2018 annual deployment plan of the North Pacific Observer Program. NOAA Tech. Memo. NMFS-AFSC-364, 30 p.

Sun, C.-H. J., and L. Fine.
2016. A cost-effective discards-proportional at-sea monitoring allocation scheme for the groundfish fishery in New England. Mar. Policy 66:75–82. Crossref

Sutinen, J. G.
1999. What works well and why: evidence from fishery-management experiences in OECD countries. ICES J. Mar. Sci. 56:1051–1058. Crossref

Thompson, S. K.
2012. Sampling, 3rd ed., 472 p. John Wiley and Sons Inc., Hoboken, NJ.

Ulrich, C., D. C. K. Wilson, J. R. Nielsen, F. Bastardie, S. A. Reeves, B. S. Andersen, and O. R. Eigaard.
2012. Challenges and opportunities for fleet- and métier-based approaches for fisheries management under the European Common Fishery Policy. Ocean Coast. Manage. 70:38–47. Crossref

van Helvoort, G.
1986. Observer program operations manual. FAO Fish. Tech. Pap. 275, 207 p. FAO Rome.

Vølstad, J. H., P. S. Afonso, A. P. Baloi, N. de Premegi, J. Meisfjord, and M. Cardinale.
2014. Probability-based survey to monitor catch and effort in coastal small-scale fisheries. Fish. Res. 151:39–46. Crossref

Wigley, S. E., and C. Tholke.
2018. 2018 discard estimation, precision, and sample size analyses for 14 federally managed species groups in the waters off the northeastern United States. NOAA Tech. Memo. NMFS-NE-243, 182 p.

Zollett, E. A., R. J. Trumble, J. H. Swasey, and S. B. Stebbins.
2015. Guiding principles for development of effective commercial fishery monitoring programs. Fisheries 40:20–25. Crossref