AN EXPERIMENTAL DESIGN
TO DETECT DIFFERENCES OF BYCATCH RATES
IN SURFACE AND SUBSURFACE DRIFTNETS
IN NORTH PACIFIC SQUID DRIFTNET FISHERIES

by

Laura J. Simon
Auke Bay Laboratory
Alaska Fisheries Science Center
National Marine Fisheries Service
National Oceanic and Atmospheric Administration
11305 Glacier Hwy.
Juneau, AK 99801-8626, U.S.A.

and

Bruce G. Lindsay
Pennsylvania State University
Statistics Department
219 Pond Laboratory
University Park, PA 16802, U.S.A.

Submitted to the
INTERNATIONAL NORTH PACIFIC FISHERIES COMMISSION
by the United States National Section

September 1991

THIS PAPER MAY BE CITED IN THE FOLLOWING MANNER:
The current squid-fishing technique entails deploying driftnet which floats at the sea surface (Fig. 1). A newly proposed driftnet gear, called subsurface driftnet, is deployed 2 m below the sea surface (Fig. 2). The use of subsurface instead of surface driftnet in the squid fishery may reduce the bycatch of nontarget species that typically inhabit surface waters.

This paper describes a large-scale experimental survey designed to study whether differences exist in the mean bycatch rates of various species in surface and subsurface driftnets. The resulting experiment includes the use of commercial fishing vessels in the Japanese squid fishery, and will provide information on the bycatch of the following nontarget species: albacore, blue shark, salmonids, albatross, petrels, shearwaters, other seabirds, Dall's porpoise, various dolphins, and northern fur seals.

After an experimental survey, a formal test of the following hypothesis can be completed for each nontarget species:

$$H_0: \left( \mu_1/\mu_2 \right) = 1 \quad \text{versus} \quad H_a: \left( \mu_1/\mu_2 \right) < 1,$$

where:

$$\mu_1 = \text{the mean bycatch rate of species } k \text{ in subsurface driftnets, and}$$

$$\mu_2 = \text{the mean bycatch rate of species } k \text{ in surface driftnets.}$$

The hypothesis test will have a significance level of 0.05 with power of at least 0.90 against the alternative that the mean bycatch rates of subsurface driftnets are half the mean bycatch rates of the surface driftnets. That is, power of 0.90 is desired against the alternative, $H_a: \left( \mu_1/\mu_2 \right) = \frac{1}{2}$.

**THE EXPERIMENTAL DESIGN**

Squid-fishing gear and techniques play an important role in the determination of a feasible experimental design. Therefore, some of the key characteristics of squid driftnet gear and fishing are first described.

The basic unit of driftnet fishing gear, a tan, is a panel of monofilament plastic net, approximately 50 m long (Fig. 1). The corkline, a small-diameter double line with plastic floats spaced approximately 1 m apart, is connected to the top of the net. The leadline, a line weighted with lead, is attached to the bottom of the net. The corkline and leadline support the netting, allowing it to float vertically in the high seas. The distance between the corkline and leadline while the net drifts in the ocean is typically 9-10 m. Stretched-mesh sizes range from 115 to 120 mm.

Approximately 140 tans tied together compose one section of driftnet roughly 5 km long. One squid-fishing operation consists of the deployment, the soaking, and the retrieval of 6-12 sections of driftnet. (Hereafter, net sections will also be referred to as nets). The nets are deployed 2-3 km apart just before sunset, and are allowed to soak approximately 7 h. The nets are retrieved two to three hours before sunrise and the catch is processed. With this schedule, a fishing vessel usually can perform one fishing operation per day.

The nets are stored in, and deployed from, a net bin at the stern of the vessel. The net is deployed over a horizontal roller from the bin at nearly constant cruising speed. Float, radio, and light buoys mark the end of the net sections. On retrieval, the nets are hauled aboard by hydraulic equipment. After the catch is removed on the main deck, the nets are pulled
ABSTRACT

A large-scale experimental survey was designed for the Japanese squid driftnet fishery in the North Pacific Ocean. The survey was designed to study whether differences exist between mean species-bycatch rates of surface and subsurface driftnets. The paired experimental design, in which a fishing vessel deploys four sections each of subsurface and surface driftnet every night, is described in detail. On completion of the experimental survey, the ratio of the mean subsurface-bycatch rate to mean surface-bycatch rate can be studied, and the hypothesis that the ratio is equal to one can be tested. The hypothesis test, at the 0.05 significance level, will have power of at least 0.90 under the alternative that the mean subsurface-bycatch rate is one-half the mean surface-bycatch rate. Two methods provided information on the required number of fishing operations to achieve the desired power. First, a theoretical power analysis was completed assuming that surface- and subsurface-bycatch counts follow Poisson distributions. However, the Poisson model did not appear to adequately describe bycatch counts in the squid fishery. Therefore, because of the availability of a large database, a computer-intensive nonparametric power analysis was performed. Because the nonparametric model better describes bycatch counts in the fishery, it provides a better assessment of the required sample size. Based on the results of the nonparametric analysis, the number of independent squid-fishing operations necessary to achieve the desired power of at least 0.90 is presented.

INTRODUCTION

From June to December each year, over 400 Japanese fishing vessels target neon flying squid (Ommastrephes bartrami) in the North Pacific Ocean (Yatsu 1990). In the afternoon, each fishing vessel deploys 30-65 km of high-seas driftnet consisting of several unconnected sections of monofilament plastic netting. Suspended vertically at the surface from floats, these nets drift freely throughout the night, entangling the squid. However, when the driftnets are retrieved in the morning, they contain not only the target species (neon flying squid) but also several nontarget species including albacore, blue shark, salmonids, seabirds, porpoises, dolphins, and marine mammals.

In recent years, the cumulative fishing effort of the Japanese squid fleet has totalled roughly 1 500 000 km of net per season (Fisheries Agency of Japan 1990). It has been estimated that over 1 300 000 albacore, 1 100 000 blue shark, 1600 salmonids, 212 800 seabirds, 3100 Dall's porpoise, and 15 300 dolphins were accidentally caught by the Japanese squid fleet in the 1989 fishing season. In 1990, an estimated 799 400 albacore, 701 500 blue shark, 141 200 salmonids, 272 700 seabirds, 3300 Dall's porpoise, and 14 100 dolphins were caught (Yatsu and Hayase 1991). The question of whether alternative fishing methods may reduce the bycatch in the fishery is therefore an important one.

A steering committee was formed in 1989 by the International North Pacific Fisheries Commission to examine alternative fishing techniques. A new fishing technique would be considered a viable alternative if it significantly reduced the bycatch of various nontarget species without reducing the catch of the target species.
through a tube along the starboard side back to the net bin, where they are
stacked for subsequent deployment. The stacking of the nets in the stern net
bin forces a "first in-last out" deployment structure, i.e., the first net
sections retrieved in an operation are the last net sections deployed in the
next operation.

From the logistics of squid fishing as described, an appropriate
experimental design can now be considered. The basic unit of the experimental
design is a fishing operation consisting of four surface and four subsurface
nets. The net sections should be as similar as possible. For example, the
stretched-mesh sizes, the length of the sections, and the stretched distances
between the leadlines and the corklines should be kept constant. The only
thing that should differ between the two net types is the depth below the sea
surface at which they are deployed; surface nets should be deployed at the
surface, whereas subsurface nets should be deployed 2 m below the surface.

Each fishing operation should be independent and follow the same plan as
outlined above. An independent fishing operation means that no vessel works
in conjunction with any other vessel. (Independence eliminates the
possibility of array fishing, in which several vessels arrange their nets in
an array formation to maximize their catch in an area. Nearby driftnet
operations may affect the bycatch rates of the two net types.) The number of
vessels used may be fewer than the number of fishing operations, i.e., one
vessel may deploy several operations. The location of each fishing operation
should not be restricted to one area but be representative of areas fished by
all commercial vessels during the sampling period.

Ideally, net section types should be alternated—subsurface, surface,
subsurface, etc.—during deployment; however, this is not feasible at sea.
Therefore, net sections should be arranged so that the first four deployed are
of one type, e.g. surface, and the last four are the remaining type, e.g.
subsurface. Upon retrieval, the nets should be hauled aboard in the same
order in which they were deployed, to minimize differences in the time that
the two net types soak.

The order of the net deployment should be randomized before each vessel
embarks. It suffices to randomly select the first net type initially stacked
in the net bin of each vessel. The order of net deployment for the duration
of each vessel's fishing cruise is predetermined by this initial selection.
The "first in-last out" structure of the net sections stacked in the net bin
and the requirement that the net sections be retrieved in the same order of
deployment imply that the first net type deployed in a fishing operation
alters daily between surface and subsurface. For example, if the first
four nets deployed on the first fishing operation are subsurface nets (that
is, surface nets are the first net type initially stacked in the net bin),
then they will be the last four nets deployed on the second fishing operation,
the first four nets deployed on the third fishing operation, and so on
(Fig. 3).

All other factors that may affect the bycatch rates within an operation
must be controlled as much as possible. Vessel operators should fish the nets
as consistently as possible to keep extraneous factors constant between the
two net types. Ensuring that the two net types are as similar as possible and
exposed to areas as similar as possible is imperative to the success of this
experiment.

One scientific observer should be assigned to each vessel to monitor the
fishing operations. During retrieval, observers should count and record the
total number of individuals of each bycatch species caught in each of the two net types. That is, the surface-driftnet bycatch count for salmonids is the total number of salmonids caught in the surface-driftnet sections, and the subsurface-driftnet bycatch count for salmonids is the total number of salmonids caught in the subsurface-driftnet sections. These two measurements provide the information regarding possible differences in the mean bycatch rates of the two net types.

POWER-ANALYSIS METHODOLOGY

The hypothesis tests are to have power of at least 0.90 against the alternative that the mean bycatch rates of subsurface driftnets are one-half the mean bycatch rates of the surface driftnets. Two separate analyses—a theoretical one and a computer-intensive one—provide information on the required sample sizes to attain the desired power of the tests under the alternative hypothesis.

AVAILABLE DATA

In 1990, 2876 squid-fishing operations were observed as a result of the U.S.-Japanese-Canadian joint observer program (Int. North Pac. Fish. Comm. 1991). For each fishing operation, a random subset of the deployed net sections was observed. If 6-9 sections were deployed, the observer on board monitored 6 sections; and if 10-12 sections were deployed, the observer monitored 7 sections. For each section, the number of individuals of each entangled species was counted and recorded. In addition, physical and environmental conditions, such as vessel location, sea-surface temperature, and lengths of net sections, were recorded.

The resulting 1990 "sectional database" contains valuable information regarding the experimental nature of bycatch counts in surface driftnets. Therefore, a subset of this database was selected to create the power-analysis database. Only data from June, July, and August fishing operations—the assumed time of the 1991 research—were included. Only observed operations with six to nine deployed nets, and with net sections of the same length, were included. In all, 1394 fishing operations satisfied the above criteria.

As discussed later (in Results of Nonparametric Reference Distribution Analysis), any operation that is an outlier for one species may unduly influence power calculations. Five operations were removed from the database because they contained a species-bycatch count that was more than twice as large as the remaining bycatch counts of that species. The database therefore contains 1389 fishing operations.

THEORETICAL POWER ANALYSIS

The Poisson distribution is sometimes used to characterize biological count data. If the events of interest occur "randomly and rarely", the Poisson probability model generally fits biological counts well; that is, the Poisson model is suitable if two assumptions hold: (1) The occurrence of the event in any space or time interval has no effect on the further occurrence of events in that same interval; and (2) the probability of two or more occurrences of the event in a small space or time interval is almost zero.
(Haight 1967; Sokal and Rohlf 1969; Consul 1989). If the species of interest are distributed at random over the region fished in an operation, the Poisson model will adequately describe bycatch counts in the squid driftnet fishery.

**Power Calculations**

Assume that surface- and subsurface-net bycatch counts in operation $i$ for species $k$ follow independent Poisson distributions. Also assume that the ratio of their means, which are related, is constant. That is, let $X_{ik}$ and $Y_{ik}$ be the surface- and subsurface-bycatch counts, respectively, of species $k$ in operation $i$. Then we assume that

$$X_{ik} - P(\lambda_{ik}) \text{ and } Y_{ik} - P(r\lambda_{ik})$$

are independent,

and $r$ is the ratio—subsurface to surface—of the two Poisson means. Let $S_{Nk}$ and $W_{Nk}$ be the total surface- and subsurface-bycatch counts, respectively, in $N$ operations. Then: $S_{Nk} = \sum_{i=1}^{N} X_{ik} - P(\lambda_{Nk})$ and $W_{Nk} = \sum_{i=1}^{N} Y_{ik} - P(r\lambda_{Nk})$,

independent, where $\lambda_{Nk} = \sum_{i=1}^{N} \lambda_{ik}$. If $T_{Nk} = S_{Nk} + W_{Nk}$, the total bycatch count of species $k$ in $N$ operations, $T_{Nk} - P(\lambda_{Nk}(1 + r))$, and the distribution of $W_{Nk}$, given $T_{Nk} = t$, is binomial; that is, $(W_{Nk}|T_{Nk} = t) - B(t, p)$, where $p = \frac{1}{1 + r}$ (Lehmann 1959).

We are interested in testing the following hypothesis:

$$H_0: r = 1 \text{ versus } H_a: r < 1,$$

or equivalently,

$$H_0: p = \frac{1}{2} \text{ versus } H_a: p < \frac{1}{2}.$$  

The hypothesis test, with $\alpha = 0.05$, will have power of 0.90 against the alternative $r = \frac{1}{4}$ or, equivalently, $p = \frac{1}{4}$.

Let $\Phi_S(S_{Nk}, W_{Nk}|T_{Nk} = t)$ be the test function for the conditional hypothesis test with size $\alpha = 0.05$. Then:

$$\Phi_S(S_{Nk}, W_{Nk}|T_{Nk} = t) = \begin{cases} 1 & \text{if } W_{Nk} \leq w_{0.05} \\ 0 & \text{if } W_{Nk} > w_{0.05} \end{cases},$$

where $w_{0.05}$ is the critical value, an integer obtained from a binomial table with $n = t$ and $p = \frac{1}{4}$. If $\Phi_S(S_{Nk}, W_{Nk}|T_{Nk} = t) = 1$, then the hypothesis test is rejected; if $\Phi_S(S_{Nk}, W_{Nk}|T_{Nk} = t) = 0$, then the hypothesis test is accepted. That is, when $T_{Nk} = t$, the null hypothesis is rejected if $W_{Nk}$—the total subsurface-bycatch counts in $N$ operations—is smaller than the critical value $w_{0.05}$. (Note that, since the binomial random variable is discrete, the actual size, $\alpha^*$, is typically smaller than the nominal size 0.05.)

Now, provided $T_{Nk} = t$ is sufficiently large, one may instead use an approximate test statistic, the normal approximation to the binomial distribution (Feller 1957). That is,
\[
\Phi_H((S_{W_i}, W_{W_i}) | T_{W_i} = t) = \begin{cases} 
1 & \text{if } Z^* \leq z_{0.05}, \\
0 & \text{if } Z^* > z_{0.05}, 
\end{cases}
\]

where \( z_{0.05} = -1.645 \) and

\[
Z^* = \frac{W_{W_i} - \frac{1}{2}t}{\sqrt{\frac{1}{4}t}} \sim N(0, 1).
\]

The conditional power of the hypothesis test for the alternative \( p = \theta \), given \( T_{W_i} = t \), is:

\[
\text{POWER}(\Phi_H((S_{W_i}, W_{W_i}) | T_{W_i} = t)) = P(H_0 \text{ is rejected given } p = \theta) = P(Z^* \leq z_{0.05} | p = \theta) = E_{p=\theta}[\Phi_H((S_{W_i}, W_{W_i}) | T_{W_i} = t)].
\]

Now, when \( p = \theta \), \( (W_{W_i}|T_{W_i} = t) \sim B(t, \theta) \). Again, provided \( T_{W_i} = t \) is large, the normal approximation to the binomial distribution may be used (Feller 1957), i.e.,

\[
Z^{**} = \frac{W_{W_i} - \theta t}{\sqrt{\theta(1-\theta)t}} \sim N(0, 1).
\]

Then, the conditional power under the alternative \( p = \theta \) is approximately

\[
\text{POWER}(\Phi_H((S_{W_i}, W_{W_i}) | T_{W_i} = t)) \approx P(Z^* \leq z_{0.05} \text{ given } Z^{**} \sim N(0, 1))
\]

\[
= P \left[ Z^{**} \leq \frac{z_{0.05}}{2\sqrt{\theta(1-\theta)}} + \frac{\left(\frac{1}{2}-\theta\right)\sqrt{\theta}}{\sqrt{\theta(1-\theta)}} \right].
\]

Thus far, the power calculations have depended on the conditional value of \( T_{W_i} = t \). To determine the power of the hypothesis test, \( \text{POWER}(\Phi_H(S_{W_i}, W_{W_i})) \), the conditional power, \( \text{POWER}(\Phi_H((S_{W_i}, W_{W_i}) | T_{W_i} = t)) \), is averaged over all possible values of \( T_{W_i} \). That is,

\[
\text{POWER}(\Phi_H(S_{W_i}, W_{W_i})) = \sum_{T_{W_i}} P(T_{W_i} = t) \cdot \text{POWER}(\Phi_H((S_{W_i}, W_{W_i}) | T_{W_i} = t))
\]

6
Therefore, because $E(Z) \geq 0$ and $E(Z^2) \leq 1$, 

$$g(0) + \int g'(0) + \frac{1}{2} g''(0) \leq g(0) + Zg'(0) + \frac{1}{2} Z^2 g''(0).$$

The second term in this power approximation is negligible (Fig. 4). Therefore, the power calculation can be further simplified, i.e., 

$$POWER[\phi_{S_{nk}, W_{nk}}] = \Phi(c_0) - \frac{1}{5} b_0^2 \Phi(c_0) \left( \frac{1 + c_0 b_0}{\mu_{nk}} \right).$$

Results

The value $\mu_{nk}$ is estimated by $\hat{\mu}_{nk} = N_k \bar{X}_k$, where $N_k$ is the sample size and $\bar{X}_k$ is the sample mean operation bycatch count obtained from the database (Table 1). Based on this estimate, plots of $POWER[\phi_{S_{nk}, W_{nk}}]$ versus sample size were obtained for $r = 1/4$ for each species of interest (Figs. 5-9). The sample size $N_k$ that ensures power of 0.90 is the value $N_k$ at which the curve crosses the horizontal reference line at $POWER = 0.90$ (Table 2). The sample sizes obtained are similar to sample sizes obtained from published results of power calculations (see Table, Design A, of Gail (1974)).

Criticisms and Improvements

The power analysis based on the Poisson model provides a benchmark for comparison with the nonparametric approach. Its weakness is a heavy dependence on two assumptions.

First, if the Poisson model is to properly describe bycatch counts, then the species of interest must be distributed randomly throughout the ocean region covered by a fishing operation. That is, if the individuals of a species school together in a smaller area than that of an operation, the assumption will be violated.

The coefficient of dispersion, $s^2/\bar{X}$, provides a measure of the distribution of species catch within a fishing operation. If $s^2/\bar{X}$ is much
where $\Phi$ is the cumulative distribution function of the standard normal random variable.

Because $T_{nk} \sim P(\mu_{nk})$, where $\mu_{nk} = \lambda_{nk}(1+r)$, then, provided $\mu_{nk}$ is sufficiently large, the distribution of $T_{nk}$ is approximated by the normal distribution (Feller 1957). That is, $T_{nk} = (\mu_{nk} + Z\sqrt{\mu_{nk}}) \sim N(\mu_{nk}, \mu_{nk})$, where $Z$ is the standard normal random variable. Then the power may be further approximated by

$$
\text{POWER}[\Phi_{H}(S_{nk}, W_{nk})] = \int \left[ \Phi \left( \frac{Z_{0.05}}{2\sqrt{\theta(1-\theta)}} + \frac{\left(\frac{1}{2}-\theta\right)\sqrt{\theta(1-\theta)}}{\sqrt{\theta(1-\theta)}} \right) \right] \phi(z) \, dz
$$

where $\phi(z)$ is the standard normal density function.

Further simplification arises by approximating the function $g(Z)$ with the Taylor series expansion about $Z = 0$, where

$$
g(Z) = \Phi \left( \frac{Z_{0.05}}{2\sqrt{\theta(1-\theta)}} + \frac{\left(\frac{1}{2}-\theta\right)\sqrt{\theta(1-\theta)}}{\sqrt{\theta(1-\theta)}} \right).
$$

That is, if

$$
a_{\theta} = \frac{Z_{0.05}}{2\sqrt{\theta(1-\theta)}}, \quad b_{\theta} = \frac{\left(\frac{1}{2}-\theta\right)\sqrt{\theta(1-\theta)}}{\sqrt{\theta(1-\theta)}}, \quad \text{and} \quad c_{\theta} = a_{\theta} + b_{\theta},
$$

then
greater than one, then the species are not randomly distributed throughout the operation, but rather are "clumped" together (Sokal and Rohlf 1969).

Operations exist in the 1990 database in which the coefficient of dispersion is much greater than one. For example, one vessel operation caught 8, 48, 101, 49, 85, and 53 albacore in each of its 6 observed net sections. The mean catch is 57.3 and the variance is 1056.3, resulting in a coefficient of dispersion of 18.4. Thus, the data behave as if \( \lambda_{ik} \) is not constant across the nets, but rather varies within the operation. If the assumption of constant \( \lambda_{ik} \) is violated, then the theoretically determined sample sizes are too small.

Second, the above theoretical power analysis is based on the assumption that \( \lambda_{ik} = \Sigma_{i=1}^{N} \lambda_{ik} \) is fixed for each sample of size \( N \). This assumption is clearly violated. Fishing operations are to be randomly selected from an ocean that exhibits a great amount of variability in the distribution of species over the squid-fishing area. Therefore, \( \lambda_{ik} \) depends strongly on the location of the \( N \) vessels within a sample. That is, \( \lambda_{ik} \) is itself a random variable, which affects the distribution of the total bycatch \( T_{ik} \).

One sees the violation of the assumption of fixed \( \lambda_{ik} \) by studying the distribution of \( T_{ik} \). Let \( T_{ik} = (X_{ik} + Y_{ik}) \), the amount of species \( k \) bycatch in all nets in operation \( i \). Then, under the null hypothesis, for fixed \( \lambda_{ik} \),

\[
T_{ik} = \Sigma_{i=1}^{N} (X_{ik} + Y_{ik}) - P(2\lambda_{ik}).
\]

That is, \( E(T_{ik}) = \text{Var}(T_{ik}) = 2\lambda_{ik} \).

However, the variance of \( T_{ik} \) is much larger than expected. As described in the next section, the empirical distribution of \( T_{ik} \) under the null hypothesis for sample size \( N \) was generated. The variance from this empirical distribution is greater than the mean. For example, 1000 different samples of 28 fishing operations were randomly selected from the 1990 database. For each sample, the total blue shark bycatch count was calculated. The mean of \( T_{ik} \) was 1189, and the variance of \( T_{ik} \) was 194 242. This exorbitant difference in the mean and variance of \( T_{ik} \) can be explained by the variation in \( \lambda_{ik} \) over the samples.

One may attempt to improve the theoretical power analysis by including the increased variation in \( T_{ik} \) due to the variation in \( \lambda_{ik} \). Assume the random variables \( \lambda_{ik} \) come from a distribution \( \mathcal{O} \) with mean \( \mu_{\mathcal{O}} \) and variance \( \sigma_{\mathcal{O}}^{2} \). Then

\[
\text{Var}(T_{ik}) = \text{Var}[E(T_{ik} | \lambda_{ik})] + E[\text{Var}(T_{ik} | \lambda_{ik})]
\]

\[
= \text{Var}[2\lambda_{ik}] + E[2\lambda_{ik}]
\]

\[
= 4\sigma_{\mathcal{O}}^{2} + 2\mu_{\mathcal{O}}, \quad \text{and}
\]

\[
\text{Var}(T_{ik}) = \text{Var}[E(\Sigma_{i=1}^{N} T_{ik} | \lambda_{1k}, \lambda_{2k}, \ldots, \lambda_{Nk})] + E[\text{Var}(\Sigma_{i=1}^{N} T_{ik} | \lambda_{1k}, \lambda_{2k}, \ldots, \lambda_{Nk})]
\]

\[
= \text{Var}[\Sigma_{i=1}^{N} 2\lambda_{ik}] + E[\Sigma_{i=1}^{N} 2\lambda_{ik}]
\]

\[
= 4N \cdot \sigma_{\mathcal{O}}^{2} + 2N \cdot \mu_{\mathcal{O}} = N \cdot \text{Var}(T_{ik}) = N \cdot \sigma_{\mathcal{O}}^{2}.
\]
That is, $T_{nk} \sim N(N \mu_{T}, N \sigma_{T}^2)$, where $\mu_{T}$ is the mean operation bycatch count and $\sigma_{T}^2$ is the variance of the operation bycatch count. Recomputing the previous power approximation with $N \mu_{T}$ replacing the mean $\mu_{nk}$, and $N \sigma_{T}^2$ replacing the variance, $\mu_{nk}$, we have

$$\text{POWER}(\phi_{E}(S_{nk}, W_{nk})) = \Phi(c_{0}) - \frac{1}{8} b_{p} N \sigma_{T}^2 \Phi(c_{0}) \left(1 + c_{p} B_{\theta} \right).$$

(Note that for $\text{Var}(T_{nk})$ the previous power analysis employed only the first term, $N \cdot \mu_{Q}$. However, based on the 1990 data, the second term $N \cdot \sigma_{T}^2$ is much larger than the first term.) Now, one can determine new estimates of power based on estimates of $\mu_{T}$ and $\sigma_{T}^2$ from the 1990 database. The sample mean bycatch operation count, $\bar{X}_{T}$, provides an unbiased estimate of $\mu_{T}$, and the sample variance bycatch operation count, $s_{T}^2$, provides an unbiased estimate of $\sigma_{T}^2$ (Table 1). Based on these estimates, the power was determined, and new optimal sample sizes were obtained for $r = \frac{1}{2}$ (Table 3).

Although the power calculations may have been improved by accounting for the extra variation in $T_{nk}$, this new analysis still does not account for the nonnormality of the distribution of $T_{nk}$. A histogram of the generated blue shark operation bycatch counts illustrates a strong skewness in the distribution of $T_{nk}$ (Fig. 10). Because the distribution of $T$ is strongly skewed and possesses larger variation than expected, we conclude that the Poisson model does not adequately describe bycatch counts in the squid fishery.

What model appropriately describes the data? Clearly the answer to this question is not simple. The many sources of potential variation make it very difficult to create a realistic model. This fact leads one to consider a computer-intensive nonparametric analysis in which no assumptions about the distribution of bycatch counts are made. The availability of a large data set from which one can gain insight into the nature of the experimental data makes this a feasible alternative.

**NONPARAMETRIC REFERENCE DISTRIBUTION ANALYSIS**

Assume that in 1990 eight net sections were deployed and observed for each fishing operation in the squid fishery. Let $X_{ijk}$ = the bycatch count of species $k$ in net $j$ of operation $i$. Assume that the expected value of the bycatch count of species $k$ is constant for each net in operation $i$, i.e., $E(X_{ijk}) = \mu_{ik}$. Let no further assumptions be made regarding the distribution of bycatch counts. Consider two test statistics—the $t$-statistic and the "ratio statistic", the ratio of total subsurface-bycatch counts to total surface-bycatch counts in $N$ operations.

**Determining Power**

Suppose, for now, that the first four net sections in operation $i$ are surface nets, and the last four net sections in operation $i$ are subsurface
nets. Let \( S_{ik} = \sum_{jk=1}^{6} X_{ijk} \) and \( W_{ik} = \sum_{jk=1}^{6} X_{ijk} \), the total surface- and subsurface-net bycatch counts, respectively, for species \( k \) in operation \( i \). Then, \( E(S_{ik}) = 4\mu_{ik} \) and \( E(W_{ik}) = 4\mu_{ik} \). The ratio of the expected subsurface- and surface-bycatch counts is then 1, the value of the ratio under the null hypothesis.

If the null hypothesis is true, then the experimental outcome should be very similar to the data collected in the 1990 observer program. That is, the subsurface-net bycatch counts should be similar to the surface-net bycatch counts. One may select a subset of \( N \) operations from the 1990 database, randomly assign the labels "treatment" and "control" to each of the two sets of four nets, and calculate the values of the test statistic. This procedure realistically describes the distribution of the test statistic when the null hypothesis is true. We call this empirically generated distribution the null reference distribution (or null reference set) of the test statistic.

The null reference distribution provides a set of possible outcomes of the test statistic to which the actual outcome of an experiment can be compared. If the actual value of the test statistic, \( t \) or ratio, is less than 95% of the values of its null reference distribution, the outcome is considered extreme, and the null hypothesis is rejected at the 0.05 significance level (Box et al. 1978).

If each fishing operation in 1990 had eight sections deployed and eight sections observed, computer procedures can generate the relevant null reference distributions for the test statistics of interest. The computer procedures randomly select \( N \) fishing operations from the 1990 observer-program database. For each fishing operation selected, the first four net sections are randomly assigned to be either surface or subsurface nets, and the last four sections to be the other type. That is, either \( S_{ik} = \sum_{jk=1}^{4} X_{ijk} \) and \( W_{ik} = \sum_{jk=1}^{6} X_{ijk} \), or \( S_{ik} = \sum_{jk=1}^{6} X_{ijk} \) and \( W_{ik} = \sum_{jk=1}^{4} X_{ijk} \). The values of the two test statistics, \( t \) and ratio, are calculated based on the \( N \) selected pairs of surface- and subsurface-bycatch counts. Performing this process 1000 times generates two null reference distributions—one for each test statistic—consisting of 1000 possible values. The critical value for each test statistic is the 50th smallest value of the 1000 possible values of the corresponding null reference distribution. If the actual value of the test statistic is small, we claim that the subsurface-bycatch rate was less than the surface-bycatch rate.

In order to determine the power of the hypothesis tests, one must determine the distribution of the test statistic when the alternative hypothesis is true. The alternative distribution of either test statistic is obtained for the 1990 squid-fishery observations by comparing bycatch of four surface nets to two other surface sections in each of \( N \) operations. That is, suppose the first four net sections in operation \( i \) are called surface nets and the last two net sections in operation \( i \) are called subsurface nets. Then, \( S_{ik} = \sum_{jk=1}^{4} X_{ijk} \) and \( W_{ik} = \sum_{jk=1}^{2} X_{ijk} \), and \( E(S_{ik}) = 4\mu_{ik} \) and \( E(W_{ik}) = 2\mu_{ik} \). The ratio of the expected subsurface- and surface-bycatch counts is \( \frac{2\mu_{ik}}{4\mu_{ik}} = \frac{1}{2} \), the value of the ratio under the alternative hypothesis for which high power is desired.

Computer procedures, while generating the null reference distributions of the test statistics, can simultaneously generate respective alternative reference distributions. For each fishing operation selected, the four net
sections randomly assigned to be surface-net sections when generating the null reference distribution are assigned to be surface-net sections for the alternative reference distribution. Then, two of the four remaining nets are randomly selected to be subsurface nets for the alternative reference distribution. For example, suppose the first four nets in operation \( i \) are randomly classified as surface nets, so \( S_{ik} = \sum_{j=1}^{4} X_{ijk} \). Then, \( S_{ik}(alt) = \sum_{j=1}^{4} X_{ijk} \), where \( S_{ik}(alt) \) is the total surface-net bycatch count under the alternative for species \( k \) in operation \( i \). Let \( W_{ik}(alt) \) be the total subsurface-net bycatch count under the alternative for species \( k \) in operation \( i \). Then \( W_{ik}(alt) = X_{ijk} + X_{imk} \), where \( (j,m) \) is randomly selected to be either \((5,6), (5,7), (5,8), (6,7), (6,8), \) or \((7,8)\).

The 1000 values of the two test statistics based on the above sampling scheme generates their null and alternative reference distributions. Each of the 1000 test-statistic values of the alternative reference distribution is compared to its critical value obtained from its null reference distribution. The null hypothesis is rejected if the test statistic is smaller than the critical value, and accepted otherwise. The power of each test is the proportion of the 1000 times that the null hypothesis is rejected.

The computer procedures described above were coded in FORTRAN. The computer program also includes the code to handle a detail thus far ignored, namely, that when eight sections were deployed in a fishing operation, only six sections were observed. In fact, for all operations with six to nine deployed sections, only six were observed. Therefore, two net-bycatch counts needed to be estimated for each operation.

An elaborate sampling scheme was coded to handle "filling" the unobserved nets. First, operations with six, seven, or nine deployed nets were treated as "pseudo-operations with eight nets deployed". Then, the same basic premise was used to fill the missing cells. Bycatch counts of unobserved net sections were estimated with the bycatch counts of a randomly selected adjacent observed net section (Fig. 11). Although estimating the bycatch counts in this manner probably did not reflect the heterogeneity of the nets in the fullest, it seemed to be a practical solution, given the other difficulties inherent in the modelling.

For each species \( k \), the computer program was run for increasing sample sizes \( N_k \) until the approximate sample size, \( N_k^* \), required to achieve power 0.90 was determined. Because of the randomness in selecting sample sizes of size \( N_k \) from the database, the obtained power can vary for a fixed \( N_k \). Therefore, the computer program was run three times each for various sample sizes \( N_k \) in the vicinity of \( N_k^* \). Then, for each \( N_k \), the average power attained for the three runs of the program was determined.

Results

As indicated earlier, the presence of outliers in the database may unreasonably influence power. In order to investigate this possibility, the computer program was run once for some of the species on two separate databases--one with the outliers included and one with the outliers excluded. For a given species, the power of both test statistics was greater when the outliers were excluded from the database (Table 4). The power of the ratio statistic appears to be especially affected by the presence of the outliers.
The t-statistic was comparatively robust to outliers. Therefore, as is the case in the theoretical analysis, the database in which outliers were excluded was used in the nonparametric analysis.

The time necessary to complete one execution of the computer program was somewhat prohibitive. For a sample size of 1000, the computer program took approximately 10 h to execute on a Compaq Deskpro 486/33L. Estimating the two unobserved net sections per selected operation probably contributed to the program execution time. Furthermore, for most species of interest, very large sample sizes were needed to achieve power of 0.90. Therefore, an exorbitant amount of time (over 1200 total hours) was necessary to gain information regarding power for each species. Despite these time constraints, valuable information was obtained for all species of interest except petrels and other seabirds.

Plots of the average power (for the t-statistic and the ratio statistic) versus the sample size \( N_k \) were obtained for all species except petrels and other seabirds (Figs. 12-19). The required sample size for each species is the sample size at which the plotted curve of the more-powerful statistic (i.e. the statistic with higher power) crosses the horizontal reference line of 0.90 (Table 5). For petrels and other seabirds, a minimum sample size was obtained by running the program once for each species. For petrels, at \( N_k = 1000 \), the power of the ratio statistic was 0.65, and the power of the t-statistic was 0.75. For other seabirds, at \( N_k = 1500 \), the powers of the ratio and t-statistic were 0.72 and 0.78, respectively (Table 5).

Neither test statistic is consistently more powerful than the other. The t-statistic is more powerful for albacore, blue shark, salmonids, and shearwaters, whereas the ratio statistic is more powerful for Dall's porpoise, dolphins, and northern fur seals. For albatross, the more-powerful statistic alternated between the two statistics in the range of the sample sizes investigated.

For each species, the required sample sizes based on the nonparametric analysis are greater than the sample sizes based on the theoretical analysis (Table 6). The sample sizes for albacore, blue shark, salmonids, and shearwaters increased by at least 460\% over the theoretically determined sample sizes. The sample sizes for Dall's porpoise and northern fur seals had the smallest increase over the theoretical results, 31.4\% and 31.9\%, respectively. The tremendous increase in sample sizes provides yet further indication of unfulfilled assumptions underlying the theoretical analysis. The nonparametric model more accurately describes the variation in species bycatch within and across fishing operations, and hence provides a better assessment of the necessary sample sizes to achieve the desired power.

**CONCLUSION**

A random sample of 900 independently observed squid-fishing operations will ensure that the desired hypothesis tests have power of at least 0.90 for all species of interest except petrels and other seabirds. That is, assume that 20 different Japanese commercial squid-fishing vessels that fish June

---

1Reference to trade names does not imply endorsement by National Marine Fisheries Service, NOAA.
through August are randomly selected. Then, 20 scientific observers, one randomly assigned to each vessel, would each need to monitor 45 independent squid-fishing operations to attain the necessary sample size.

ACKNOWLEDGEMENTS

We appreciate the ever-helpful personnel of the Auke Bay Fisheries Laboratory, National Marine Fisheries Service: Dr. Jerome Pella, Dr. Mike Dahlberg, and Steve Ignell for their review of the paper, and Rose Rumbaugh for her support and expert SAS advice. We also thank Dr. Jim Rosenberger of Pennsylvania State University for his review of the paper.

REFERENCES

Table 1. Descriptive statistics of the species-bycatch counts. The statistics are based only on the 1389 fishing operations in the subset database.

<table>
<thead>
<tr>
<th>Variable</th>
<th>N</th>
<th>Mean</th>
<th>Std. deviation</th>
<th>Min.</th>
<th>Max.</th>
<th>Proportion of operations with positive catch</th>
</tr>
</thead>
<tbody>
<tr>
<td>Albacore</td>
<td>1389</td>
<td>34.30</td>
<td>72.32</td>
<td>0</td>
<td>789</td>
<td>0.674</td>
</tr>
<tr>
<td>Blue shark</td>
<td>1389</td>
<td>31.73</td>
<td>63.27</td>
<td>0</td>
<td>693</td>
<td>0.857</td>
</tr>
<tr>
<td>Salmonids</td>
<td>1389</td>
<td>0.94</td>
<td>10.19</td>
<td>0</td>
<td>238</td>
<td>0.071</td>
</tr>
<tr>
<td>Albatross</td>
<td>1389</td>
<td>0.35</td>
<td>1.14</td>
<td>0</td>
<td>20</td>
<td>0.191</td>
</tr>
<tr>
<td>Petrels</td>
<td>1389</td>
<td>0.13</td>
<td>1.06</td>
<td>0</td>
<td>29</td>
<td>0.044</td>
</tr>
<tr>
<td>Shearwaters</td>
<td>1389</td>
<td>7.45</td>
<td>21.82</td>
<td>0</td>
<td>318</td>
<td>0.652</td>
</tr>
<tr>
<td>Other seabirds</td>
<td>1389</td>
<td>0.06</td>
<td>0.53</td>
<td>0</td>
<td>13</td>
<td>0.027</td>
</tr>
<tr>
<td>Dall's porpoise</td>
<td>1389</td>
<td>0.11</td>
<td>0.43</td>
<td>0</td>
<td>6</td>
<td>0.081</td>
</tr>
<tr>
<td>Dolphins</td>
<td>1389</td>
<td>0.46</td>
<td>1.10</td>
<td>0</td>
<td>12</td>
<td>0.230</td>
</tr>
<tr>
<td>Northern fur seals</td>
<td>1389</td>
<td>0.18</td>
<td>0.72</td>
<td>0</td>
<td>12</td>
<td>0.107</td>
</tr>
</tbody>
</table>

Table 2. Sample sizes necessary to achieve power of 0.90 against the alternative $r = \frac{1}{2}$. The approximate power was determined theoretically under the Poisson model.

<table>
<thead>
<tr>
<th>Species</th>
<th>Sample size</th>
</tr>
</thead>
<tbody>
<tr>
<td>Albacore</td>
<td>3</td>
</tr>
<tr>
<td>Blue shark</td>
<td>3</td>
</tr>
<tr>
<td>Salmonids</td>
<td>79</td>
</tr>
<tr>
<td>Albatross</td>
<td>211</td>
</tr>
<tr>
<td>Petrels</td>
<td>569</td>
</tr>
<tr>
<td>Shearwaters</td>
<td>10</td>
</tr>
<tr>
<td>Other seabirds</td>
<td>1322</td>
</tr>
<tr>
<td>Dall's porpoise</td>
<td>666</td>
</tr>
<tr>
<td>Dolphins</td>
<td>160</td>
</tr>
<tr>
<td>Northern fur seals</td>
<td>398</td>
</tr>
</tbody>
</table>
Table 3. New optimal sample sizes determined by including the extra variation in the distribution of $T_{Nk}$. The sample sizes are based on achieving power of 0.90 against the alternative $r = 4$.

<table>
<thead>
<tr>
<th>Species</th>
<th>Sample size</th>
</tr>
</thead>
<tbody>
<tr>
<td>Albacore</td>
<td>5</td>
</tr>
<tr>
<td>Blue shark</td>
<td>5</td>
</tr>
<tr>
<td>Salmonids</td>
<td>151</td>
</tr>
<tr>
<td>Albatross</td>
<td>224</td>
</tr>
<tr>
<td>Petrels</td>
<td>652</td>
</tr>
<tr>
<td>Shearwaters</td>
<td>17</td>
</tr>
<tr>
<td>Other seabirds</td>
<td>1432</td>
</tr>
<tr>
<td>Dall's porpoise</td>
<td>685</td>
</tr>
<tr>
<td>Dolphins</td>
<td>166</td>
</tr>
<tr>
<td>Northern fur seals</td>
<td>417</td>
</tr>
</tbody>
</table>

Table 4. Outliers in the database unduly influence power. The computer program was run for three species on two separate databases—with and without outliers. The power of both test statistics was greater when the outliers were excluded from the database.

<table>
<thead>
<tr>
<th>Species</th>
<th>$N$</th>
<th>$t$ with outliers</th>
<th>Ratio with outliers</th>
<th>$t$ without outliers</th>
<th>Ratio without outliers</th>
</tr>
</thead>
<tbody>
<tr>
<td>Albacore</td>
<td>40</td>
<td>0.85</td>
<td>0.74</td>
<td>0.92</td>
<td>0.82</td>
</tr>
<tr>
<td>Salmonids</td>
<td>1000</td>
<td>0.87</td>
<td>0.29</td>
<td>0.90</td>
<td>0.76</td>
</tr>
<tr>
<td>Shearwaters</td>
<td>140</td>
<td>0.89</td>
<td>0.44</td>
<td>0.91</td>
<td>0.65</td>
</tr>
</tbody>
</table>
Table 5. Sample sizes necessary to achieve power of 0.90 against the alternative $r = \frac{1}{2}$. The approximate power was determined nonparametrically.

<table>
<thead>
<tr>
<th>Species</th>
<th>Sample size</th>
</tr>
</thead>
<tbody>
<tr>
<td>Albacore</td>
<td>40</td>
</tr>
<tr>
<td>Blue shark</td>
<td>30</td>
</tr>
<tr>
<td>Salmonids</td>
<td>850</td>
</tr>
<tr>
<td>Albatross</td>
<td>400</td>
</tr>
<tr>
<td>Petrels</td>
<td>&gt;1000</td>
</tr>
<tr>
<td>Shearwaters</td>
<td>125</td>
</tr>
<tr>
<td>Other seabirds</td>
<td>&gt;1500</td>
</tr>
<tr>
<td>Dall's porpoise</td>
<td>900</td>
</tr>
<tr>
<td>Dolphins</td>
<td>350</td>
</tr>
<tr>
<td>Northern fur seals</td>
<td>550</td>
</tr>
</tbody>
</table>

Table 6. For each species, the required sample sizes based on nonparametric analysis are greater than the sample sizes based on theoretical analysis.

<table>
<thead>
<tr>
<th>Species</th>
<th>Theoretical sample size</th>
<th>Nonparametric sample size</th>
<th>% increase</th>
</tr>
</thead>
<tbody>
<tr>
<td>Albacore</td>
<td>5</td>
<td>40</td>
<td>+700.0</td>
</tr>
<tr>
<td>Blue shark</td>
<td>5</td>
<td>30</td>
<td>+500.0</td>
</tr>
<tr>
<td>Salmonids</td>
<td>151</td>
<td>850</td>
<td>+462.9</td>
</tr>
<tr>
<td>Albatross</td>
<td>224</td>
<td>400</td>
<td>+78.6</td>
</tr>
<tr>
<td>Petrels</td>
<td>652</td>
<td>&gt;1000</td>
<td>+53.4</td>
</tr>
<tr>
<td>Shearwaters</td>
<td>17</td>
<td>125</td>
<td>+635.3</td>
</tr>
<tr>
<td>Other seabirds</td>
<td>1432</td>
<td>&gt;1500</td>
<td>+4.7</td>
</tr>
<tr>
<td>Dall's porpoise</td>
<td>685</td>
<td>900</td>
<td>+31.4</td>
</tr>
<tr>
<td>Dolphins</td>
<td>166</td>
<td>350</td>
<td>+110.8</td>
</tr>
<tr>
<td>Northern fur seals</td>
<td>417</td>
<td>550</td>
<td>+31.9</td>
</tr>
</tbody>
</table>
Figure 1. One tan of surface driftnet, and its dimensions.
Figure 2. One tan of subsurface driftnet, and its dimensions. Suspension lines allow the net to be deployed below the sea surface.
Figure 3. The "first-in/last-out" net-deployment method. The first net type in the stern net bin is the last net type deployed.
Figure 4. The value of the second term of the theoretical-power approximation plotted for $r = \frac{1}{3}$ and for $\mu_{\text{Kr}}$ between 0 and 300.
Figure 5. Plots of theoretical power versus sample size for albacore, blue shark, and shearwaters.
Figure 6. Plots of theoretical power versus sample size for salmonids.
Figure 7. Plots of theoretical power versus sample size for the albatross and dolphin species.
Figure 8. Plots of theoretical power versus sample size for northern fur seals and petrels.
Figure 9. Plots of theoretical power versus sample size for other seabirds and Dall’s porpoise.
Figure 10. A histogram of the generated total blue shark bycatch counts in 28 operations.
<table>
<thead>
<tr>
<th>1990 operation:</th>
<th>&quot;Filling&quot; of missing cells:</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>C</td>
</tr>
<tr>
<td>1</td>
<td>2</td>
</tr>
<tr>
<td>A</td>
<td>A</td>
</tr>
<tr>
<td>A</td>
<td>C</td>
</tr>
<tr>
<td>A</td>
<td>C</td>
</tr>
</tbody>
</table>

Figure 11. Example of how unobserved sections are estimated for an operation with 8 sections deployed and 6 sections observed. One of the 4 configurations on the right is randomly selected for the estimate.
Figure 12. Power, determined nonparametrically, versus sample size for albacore.
Figure 13. Power, determined nonparametrically, versus sample size for albatross.
Figure 14. Power, determined nonparametrically, versus sample size for blue shark.
Figure 15. Power, determined nonparametrically, versus sample size for Dall's porpoise.
Figure 16. Power, determined nonparametrically, versus sample size for dolphin.
Figure 17. Power, determined nonparametrically, versus sample size for northern fur seals.
Figure 18. Power, determined nonparametrically, versus sample size for salmonids.
Figure 19. Power, determined nonparametrically, versus sample size for shearwaters.