INTER-AMERICAN TROPICAL TUNA COMMISSION
COMISIÓN INTERAMERICANA DEL ATÚN TROPICAL

WORKSHOP ON TURTLE BYCATCH MITIGATION FOR LONGLINE FISHERIES: EXPERIMENTAL DESIGN AND DATA ANALYSIS

7-8 November 2007
San Ramón, Alajuela, Costa Rica

REUNIÓN SOBRE LA MITIGACIÓN DE LA CAPTURA INCIDENTAL DE TORTUGAS EN LAS PESQUERÍAS DE PALANGRE: DISEÑO EXPERIMENTAL Y ANÁLISIS DE DATOS.

7-8 noviembre 2007
San Ramón de Alajuela, Costa Rica

La Jolla, California
2008

OFCF

CIAT
IATTC

NORR

The Inter-American Tropical Tuna Com-mission (IATTC) operates under the authority and direction of a convention originally entered into by Costa Rica and the United States. The convention, which came into force in 1950, is open to adherence by other governments whose nationals fish for tropical tunas in the eastern Pacific Ocean. Under this provision Panama adhered in 1953, Ecuador in 1961, Mexico in 1964, Canada in 1968, Japan in 1970, France and Nicaragua in 1973, Vanuatu in 1990, Venezuela in 1991, El Salvador in 1997, Guatemala in 2000, Peru in 2002, Spain in 2003, the Republic of Korea in 2005. Canada withdrew from the Commission in 1984, and Colombia in 2007.

Additional information about the IATTC and its publications can be found on the inside back cover of this report.

La Comisión Interamericana del Atún Tropical (CIAT) funciona bajo la autoridad y dirección de una convención establecida originalmente por Costa Rica y los Estados Unidos. La Convención, vigente desde 1950, está abierta a la afiliación de otros gobiernos cuyos ciudadanos pescan atunes en el Océano Pacífico oriental. Bajo esta estipulación, Panamá se afilió en 1953, Ecuador en 1961, México en 1964, Canadá en 1968, Japón en 1970, Francia y Nicaragua en 1973, Vanuatu en 1990, Venezuela en 1991, El Salvador en 1997, Guatemala en 2000, Perú en 2002, España en 2003, la República de Corea in 2005, y Colombia en 2007. Canadá se retiró de la Comisión en 1984.

En la otra contraportada de este informe se presenta información adicional sobre la CIAT y sus publicaciones.

COMMISSIONERS—COMISIONADOS

COLOMBIA
Diego Felipe Cardena Montenegro

COSTA RICA
Bernal Alberto Chavarría Valverde
Asdrubal Vásquez Nuñez
Carlos Villalobos Sole

ECUADOR
Jimmy Martínez Ortiz
Ramón Montaño Cruz
Guillermo Morán Velásquez
Luis Torres Navarrete

EL SALVADOR
Manuel Calvo Benivides
Manuel Ferín Oliva
Sonia Salaverría
José Emilio Suadi Hasbun

ESPAÑA—SPAIN
Rafael Centenera Ulecia
Fernando Curcio Ruigómez
Samuel J. Juárez Casado

FRANCE—FRANCIA
Marie-Sophie Dufau-Richet
Christiane Laurent-Monpetit
Delphine Leguerrier
Michel Sallenave

GUATEMALA
Fraterno Diaz Monge
Carmen Sandoval

JAPAN—JAPÓN
Masahiro Ishikawa
Shingo Ota
Ryotaro Suzuki

MÉXICO
Marío Aguilar Sanchez
Miguel Ángel Cisneros Mata
Ramón Corral Ávila
Michel Dreyfus León

NICARAGUA
Steadman Fagoth Müller
Manuel Pérez Moreno
Armando Segura Espinoza
Edward E. Weissman

PANAMÁ
María Patricia Díaz
Arnulfo Franco Rodríguez
Leika Martínez
George Novey

PERÚ
Gladys Cárdenas Quintana
Alfonso Miranda Eyzaguirre
Doris Sotomayor Yalan
Jorge Vértiz Calderón

REPUBLIC OF KOREA— REPÚBLICA DE COREA
Chiguk Ahn
Young-Hoon Chung
Jeongseok Park

USA—EE.UU.
Robert Fletcher
Rodney McInnis
Patrick Rose

VANUATU
Christophe Emelee
Roy Mickey Joy
Dmitri Malvirlani

VENEZUELA
Alvin Delgado
Gilberto Gimenez
Nancy Tablante

DIRECTOR

Dr. Guillermo A. Compeán

HEADQUARTERS AND MAIN LABORATORY—OFICINA Y LABORATORIO PRINCIPAL
8604 La Jolla Shores Drive
La Jolla, California 92037-1508, USA
www.iattc.org

INTER-AMERICAN TROPICAL TUNA COMMISSION
COMISIÓN INTERAMERICANA DEL ATÚN TROPICAL

Special Report 17
Informe Especial 17

WORKSHOP ON TURTLE BYCATCH MITIGATION FOR LONGLINE FISHERIES: EXPERIMENTAL DESIGN AND DATA ANALYSIS

7-8 November 2007
San Ramón, Alajuela, Costa Rica

REUNIÓN SOBRE LA MITIGACIÓN DE LA CAPTURA INCIDENTAL DE TORTUGAS EN LAS PESQUERÍAS DE PALANGRE: DISEÑO EXPERIMENTAL Y ANÁLISIS DE DATOS

7-8 noviembre 2007
San Ramón de Alajuela, Costa Rica

La Jolla, California, USA

2008

TABLE OF CONTENTS

PARTICIPANTS

Mary Christman, University of Florida, U.S.A.
Daniel Hall, University of Georgia, U.S.A.
Martín Hall, Inter-American Tropical Tuna Commission
Paul Kinas, Fundacáo Universidade do Rio Grande, Brazil
Cleridy Lennert-Cody, Inter-American Tropical Tuna Commission
Bryan Manly, Western Ecosystem Technology, Inc, U.S.A.
Marti McCracken, NMFS/NOAA-Hawaii, U.S.A.
Mihoko Minami, Institute of Statistical Mathematics, Japan
Michelle Sims, Duke University, North Carolina, U.S.A.
Steven Thompson, Simon Fraser University, Canada

1. EXECUTIVE SUMMARY

Large numbers of fishing vessels operating from ports in Latin America participate in surface longline fisheries in the eastern Pacific Ocean (EPO), and several species of sea turtles inhabit the grounds where these fleets operate. The endangered status of several sea turtle species, and the success of circle hooks ('treatment' hooks) in reducing turtle hookings in other ocean areas, as compared to J-hooks and Japanese-style tuna hooks ('control' hooks), prompted the initiation of a hook exchange program on the west coast of Latin America, the Eastern Pacific Regional Sea Turtle Program (EPRSTP)[1]. One of the goals of the EPRSTP is to determine if circle hooks would be effective at reducing turtle bycatch in artisanal fisheries of the EPO without significantly reducing the catch of marketable fish species. Participating fishers were provided with circle hooks at no cost and asked to replace the J/Japanese-style tuna hooks on their longlines with circle hooks in an alternating manner. Data collected by the EPRSTP show differences in longline gear and operational characteristics within and among countries. These aspects of the data, in addition to difficulties encountered with implementation of the alternating-hook design, pose challenges for analysis of these data.

The discussions at this workshop were focused on methods of analysis of the existing EPRSTP data, with some discussion as to variations of the current sampling design that might be considered for future experiments. It is important to note that there was not a unanimous opinion among workshop participants as to appropriateness of the current alternating-hook sampling design for the estimation of hook performance in an actual fishery setting (*i.e.*, fishing with longlines of predominantly or exclusively one hook type). The differences in opinion stemmed from the fact that there is a fundamental lack of information on: 1) animal behavior at longlines, and 2) whether hooks on a given longline can be assumed to fish independently. Specifically, some participants were of the opinion that no data collected from longlines containing both control and treatment hook types would be appropriate for estimation of differences in hooking rates under actual fishing conditions unless the hooks were fishing completely independently. Further, these participants were of the opinion that hooks do not fish independently. Other participants did not believe that interactions were likely to be an issue, and thought the current sampling design was appropriate. Still other participants were of the opinion that the current sampling design was appropriate as a starting place to determine if there existed a relative difference in hooking rates between hook types, but that the design would lead to biased estimates of differences in hooking rates in the presence of interactions. The workshop discussion did not expressly take up the question of how to determine whether a longline of predominantly (or exclusively) treatment hooks behaves the same with respect to bycatch and catch rates as a longline of predominantly (or exclusively) control hooks, adjusting for all the likely sources of variation such as longline length, spacing between hooks, bait type and location of fishing.

The workshop discussion on data analysis included simple procedures for testing hypotheses about relative differences in hook performance, and on the development of complex models for hooking rates. Specifically, there was discussion of how to construct randomization procedures for testing hook performance when the order of placement of treatment and control hooks on the longline was unknown. Other simple tests for complex data, such as the Mantel-Haenszel test,

---

[1] The Eastern Pacific Regional Sea Turtle Program is an 'umbrella' name that describes a set of activities carried on, on a voluntary basis, by several institutions working in the region, to reduce the incidental mortality of sea turtles in the region's fisheries, with emphasis on artisanal longlining.

were discussed, and there was a limited discussion on simple Bayesian procedures. Case-control match-pairs analyses were also debated as a means of improving the detection of differences in the hooking rates of rare species when hook placement was unknown and when it was of importance to take into account the catch of other species. In regards to more complex models, discussion focused on the feasibility of developing zero-inflated models for these types of data when most hooks (longlines) have no animals hooked, and on spatial modeling. There was also limited discussion on testing hypotheses of equivalence, specifically for the purpose of comparing catch rates of marketable fish among hook types. In regards to sampling design, most of the discussion pertained to the use of alternating sections of homogeneous hook types, as compared to alternating individual hooks. There was a limited discussion on the estimation of sample size and determination of appropriate sampling units.

There was a diversity of opinions on the most appropriate approaches for extending the existing EPRSTP sampling design, and on the analysis of the existing EPRSTP data. Nonetheless, the main conclusions of the workshop discussion were as follows.

1) The requested sampling design of alternating individual hook types is probably sufficient if the goal is to establish whether there is a difference in hook performance. However, in the presence of interactions between hooks on the same longline, this sampling design is likely to yield biased estimates of the magnitude of the difference in hook performance. If that is the case, a follow-up experiment with a hook placement design less susceptible to hook interactions should be considered, particularly if estimating the magnitude of the difference of hook performance is an objective.

2) Collection of data on individual hooks and development of case-control methods for longline data were suggested as a means of improving ability to detect differences in hook performance for rare events, particularly when the placement of hook types along the longline does not follow a regular pattern and/or when it may be important to consider interactions among species.

3) Because the dynamics of turtle interactions with longlines are not well understood, and hookings are relatively rare events, developing complex models for hooking data will likely be difficult. Moreover, complex models may require assumptions unlikely to be valid in most situations when using the current design for testing differences in hooking rates. Simple tests of hook performance, based on a randomization procedure or a Mantel-Haenszel test, should be conducted before attempting to develop complex models.

4) Simulations with existing data should be conducted to develop a better understanding of any effect of differences in longline configuration, or departures from the specified sampling design, on measured differences in hook performance. Simulations with existing data should also be conducted to determine the best sampling unit, and to estimate sample size for future experiments. As part of these simulations, tests for spatial-temporal homogeneity in the data should be performed.

2. INTRODUCTION AND BACKGROUND INFORMATION

Section 2 was provided to workshop participants as background information prior to the workshop in order to motivate discussion at the workshop.

2.1. Longlines and turtles

Sea turtles are wide-ranging species (Polovina *et al*., 2004), and as a result they come into

contact with many types of fishing gear across the Pacific Ocean, including longlines. Large numbers of fishing vessels operating from ports in Latin America participate in surface longline fisheries in the eastern Pacific Ocean (EPO) targeting tunas, billfishes, sharks, mahi-mahi, and other fish species (Largacha *et al*., 2005; Hall *et al*., 2006). Several species of sea turtles inhabit the fishing grounds of these fleets. Concern for sea turtle populations because of decreasing trends in the numbers of turtles reported at nesting beaches around the Pacific (Steyermark *et al*., 1996; Sarti *et al*., 1996, 2000; Kamezaki *et al*., 2003; Limpus and Limpus, 2003) has led to increased efforts to reduce sea turtle bycatch in longlines in the EPO and elsewhere.

A typical longline (Figure 1) consists of a mainline (at least several miles long), to which floatlines and branchlines are attached (Beverly and Chapman, 2007, and references therein). Floatlines are used to suspend the mainline from surface floats, in order to place the mainline at a desired depth below the water's surface. Branchlines attach the hooks to the mainline at regular intervals between floatlines. The actual fishing depth of a hook depends on the length of the floatline and the branchline, and on the hook's position in the catenary between floatlines. Depending on environmental characteristics (*e.g.*, water masses, fronts, currents), some hooks may be exposed to different conditions than other hooks. Fishers use different lengths of floatlines and branchlines, depending on the fish species they hope to catch (*i.e.*, depending on their 'target' species) and the environment in which they fish. For instance, the high-seas fisheries targeting bigeye tuna in the EPO fish at greater depths because the target species is more abundant at 200-400 m. By contrast, the coastal fisheries described by Largacha *et al*. (2005) and Hall *et al*. (2006) fish at 10-30 m depth. Hooks are typically baited as the longline is deployed. Bait may be either whole or sectioned animals, depending on bait/hook size and on bait availability. Bait can be a specific taxon (*e.g.*, squid) or it may be made from any number of animals caught on a previous fishing trip. Once deployed, longlines are allowed to fish ('soak') for several hours to overnight. Bait may not be retained on all hooks over the soak period, and diffusion of bait odor may be asymmetrical, orienting down current. Retrieval of the longline can take hours.

The factors that influence sea turtle interactions with longlines are not yet well known (Beverly and Chapman, 2007). Although turtles occasionally dive, they are primarily surface dwellers, spending most of their time in the upper 20-30 m of the water column. Turtles encounter longlines in regions where they forage and along migration routes to nesting beaches (Beverly and Chapman, 2007). Experiments suggest that turtles have better color vision than fishes, permitting them to see a range of colors, and that they respond to light sources (Sea Turtle and Pelagic Fish Sensory Physiology Workshop, 2006). Limited work has been done to determine if turtles are attracted to different colored surface floats (Beverly and Chapman, 2007). The diets of sea turtles differ by species; most turtle species encountered in the areas of operation of Latin American artisanal longline fleets feed on fish and squid, except for the leatherback which has a diet of jellyfish and other gelatinous organisms. Field trials have found differences in turtle hooking rates with different bait types (*e.g.*, squid bait *versus* mackerel; Watson *et al*., 2005; Boggs and Swimmer, 2007). However, it is not known if this difference is related to a diet preference or differences in the ease with which turtles can remove different bait types from hooks (Gilman *et al*., 2006). Thus, it is currently unknown to what extent turtles interact with longline gear because of an attraction to bait, because of a visual cue in response the fishing gear (*e.g.*, surface floats), or simply by chance encounter. Moreover, once a turtle has encountered a longline, it is unknown to what extent the turtle may travel the mainline from one part of the longline to the next or whether the turtle only investigates the immediate area where it

intercepted the longline. Differences in responses to longlines may differ among turtle species (*e.g.*, because of differences in diets, or because of differences in attraction to floating objects) or between juveniles and adults (*e.g.*, because of ontonogenic changes in behavior).

2.2. The hook exchange program of the Eastern Pacific Regional Sea Turtle Program

Most hooks used in the world's longlines are of two types: J-hooks and Japanese-style tuna hooks (Figure 2). Research into longline hook design has shown that wider hooks ('circle' hooks; Figure 2) can have lower capture rates of turtles than traditional J/Japanese-style tuna hooks, and are less likely to be swallowed (Watson *et al*., 2005; Gilman *et al*., 2006; Boggs and Swimmer, 2007). The success of the use of wider hooks in reducing turtle hookings prompted the initiation of a 'hook exchange' program in Latin America, the Eastern Pacific Regional Sea Turtle Program (EPRSTP)[2], in 2004. One of the goals of the EPRSTP is to determine if circle hooks would be effective at reducing turtle bycatch in artisanal fisheries of the EPO without significantly reducing the catch of marketable fish species (Largacha *et al*., 2005; Hall *et al*., 2006). The main purpose of the EPRSTP was to conduct experiments on the effects of different hook types on turtle bycatch and fish catch within normal longline fishing operations throughout Latin America. (Bycatch is defined as any species hooked that is not marketable.)

The motivation behind the formation of the EPRSTP is the fact that several sea turtle species are endangered (Beverly and Chapman, 2007, and references therein), and it is necessary to develop an effective mitigation scheme. The first management question is: do circle hooks result in significantly lower hooking rates of sea turtles than J/Japanese-style tuna hooks? A positive answer to this question would trigger a huge tide of change to replace the hooks in the whole EPO region. If the answer is not clear or negative, it would result in a major re-direction of research efforts towards other, not easy to visualize, alternatives. The poverty of the fishing communities, and the lack of options in the economies of the countries of the region does not leave many options open. Another question that is critical to the success of the EPRSTP is: are the catch rates of circle hooks equal to or greater than the catch rates of the J/Japanese-style tuna hooks? If the answer to this question is negative, then the fishing community will resist the change, and it will become unfeasible. Given the very low abundance estimates of some turtle species, a major effort cannot be wasted in the wrong direction.

The hook exchange program of the EPRSTP is a voluntary program intended to encourage fishers to test the use of circle hooks in daily operations. Participating fishers are provided with circle hooks at no cost and requested to exchange the J/Japanese-style tuna hooks on their longlines with circle hooks in an alternating manner. For example, for the comparison of one type of circle hook with J/Japanese-style tuna hooks, the request to fishers was to replace every other J/Japanese-style tuna hook ('control') on the longline with a circle hook ('treatment') (*i.e*., J, C, J, C, J, C….). This pairing of control/treatment was to be repeated the length of the line. If two types of circle hooks were to be tested on the same mainline, the trio of control-treatment-treatment was to be repeated the length of the line. As part of participation in the hook exchange program, fishers agree to take observers to sea with them during their fishing trips. Observers record characteristics of the fishing gear, the fishing operation, the amounts and species of catch and bycatch by hook type, characteristics of turtle entanglements, and limited data on turtle

---

[2] The Eastern Pacific Regional Sea Turtle Program is an 'umbrella' name that describes a set of activities carried on, on a voluntary basis, by several institutions working in the region, to reduce the incidental mortality of sea turtles in the region's fisheries, with emphasis on artisanal longlining.

sightings. In every other aspect, fishers are left to fish as they choose. For example, bait types and bait placement are not part of the EPRSTP hook exchange program, because of the impossibility to control them. To date, hook exchange programs have been initiated in nine countries in Latin America.

2.3. Examples of data collected by the EPRSTP

Data have been collected as part of the EPRSTP since 2004. These data represent longline fisheries for several different target species, although there is some ambiguity regarding the definition of target species. The level of fisher participation in the EPRSTP and program resources vary by country, not necessarily in proportion to fleet size. (Estimates of fleet size are generally not available.) The following is a presentation of data collected during 2005 for longline fisheries targeting tunas of three different countries (hereafter referred to Countries A-C). These data were selected because they illustrate some of the challenges posed by data collected by the EPRSTP as a whole. For simplicity, the example data have been limited to longlines that involved comparisons between J and C16 hook types (single 'treatment'; but see Sections 2.6.1-2.6.2).

There is a hierarchical structure inherent in the data as a result of longline fishing activities. Vessels fish in units referred to as trips, which may last one or several days. A trip begins when the vessel leaves port and ends when the vessel returns. A trip may involve one set of a longline (a 'set') or several sets. Thus, sets are nested within trips which are nested within vessels. The same vessel may be sampled more than once over the course of the year. Measures of sample size for the example data are shown in Table 1. Vessels in the example data ranged in length from 6 to 32 m. Fishing months are shown in Table 2. Fishing took place between the coast and about 90°W, from roughly 10°S to 10°N. The fishing locations of vessels of the three countries did not overlap.

For Country A, it is known that some smaller vessels may have operated in cooperation with 'motherships.' Although motherships may make sets themselves, apparently more commonly they act as tender vessels, towing smaller vessels to fishing grounds, storing the smaller vessels' catch, and providing sleeping/eating quarters for the fishers. Because it is thought to be unlikely that the distinction between mothership, independent small vessel, and small vessel operating with a mothership, affects a particular vessel's fishing activities in ways that are not already captured by other aspects of the data, these vessel categories have been ignored in this treatment of the example data. In the case of a small vessel operating with a mothership, not all of the small vessel's sets may be sampled by an observer because the observer may have switched daily among small vessels. Nonetheless, the data structure remains consistent in the sense that a trip of a small vessel is based on the small vessel's departure from and arrival to port, and the sets of that trip were only sets made exclusively by that particular small vessel.

2.3.1. Longline gear and operational characteristics

The available data show differences among countries in terms of both gear characteristics and operational characteristics. Longlines sampled for Country A typically had the greatest distance between hooks and the fewest hooks in the water, whereas sampled longlines from Country B had the shortest distance between hooks and the most hooks in the water (Figure 3). These differences translate into differences among countries in terms of the estimated length of mainline in the water (Figure 3). This raises the question as to whether longlines with different numbers of hooks in the water and/or total length of mainline in the water equate to different

sampling devices or whether these differences can be ignored. Most sampled longline sets were largely daytime activities (Figure 4). Median approximate soak times among the countries ranged over a factor of about two times (Figure 4). Bait was more homogeneous within sampled longlines of Country A, and perhaps of Country B, than of Country C (Table 3).

Not only were sampled longlines of different countries of differing lengths and hook density, but the balance of J hooks to C16 hooks per line differed among countries (Figure 5). Countries A and C achieved a relatively balanced number of hooks of each type per longline, but most longlines of Country B had more J hooks than C16 hooks. This raises the question as to whether there is an 'acceptable' range of numbers of control and treatment hooks on the same longline (or whether the numbers should be exactly equal). An imbalance in the number of different types of hooks on a longline might be particularly problematic if interactions between hooks occur. In the case of Country B, the fact that a number of longlines had many more J hooks than C hooks implies that hook placement could not have been alternating at the level of individual hooks, although it may have been alternating in groups of hooks. (No data are available on the details of hook placement.) This suggests that any interactions between hooks on the line would be more likely to involve a J hook, regardless of the hook type first encountered, and thus, interactions between hooks might lead to more J hookings than expected if hooks acted independently.

2.3.2. Hookings of turtles and fish

Turtle hookings were relatively rare events, both in terms of numbers of longlines that had turtles and in terms of turtles per line, regardless of the hook type (Figures 6-7). For all three countries, many lines had no turtles and many hooks on a given line had no turtles. By contrast, marketable fish catch varied considerably more among countries, and was generally more abundant (Figures 6-7). Regardless, the frequency distributions of numbers of turtles and catch of marketable fish per hook are skewed toward small values for all three countries (Figures 8-9). The low proportion of hooks with either catch or bycatch are consistent with longline data from fisheries in other oceans (*e.g.*, Watson *et al*., 2005; Kerstetter and Graves, 2006; Kim *et al*. 2007; Yokota *et al*. 2006). The paucity of hookings per line raises the question as to what is the appropriate measure of sample size as regards obtaining sufficient data for a comparison of hook performance. Animals, hooks, lines? It also raises the question as to whether sets with no catch of the species of interest should be considered as successful trials, and given this, what sampling designs and methods of estimation should be considered in order to make the best use of experimental data that appear to yield a low number of actual trials.

To note is that once a hook has caught a turtle, it cannot catch a fish (and *vice versa*). This would seem, at least in principle, to impart some dependence into the hookings data of turtles and fishes, and as a result, comparisons of hook performance for turtles and for fish catch. However, because relatively few hooks caught either turtles or marketable fish (Figure 10), perhaps this dependence can be ignored.

2.3.3. A preliminary consideration of hookings as independent Bernoulli events

As a starting point for modeling hookings to compare hook performance, one possibility would be to assume that hooks on a longline represent independent Bernoulli trials. In other words, the number of successes (hooked animals) for a given hook type might be modeled as binomial(n, p), where n is the number of hooks of a given type on the longline and p is the probability that a hook of a given type catches an animal. The application of this model to the example data is discussed below.

For turtles, the example data suggest that in some instances number of hooks on the longline may not have an observable effect on the probability that a hook catches a turtle (at least over the range of n sampled). For both Countries A and B, an estimate of the probability of at least one J hook per line with a turtle appears to be consistent with the assumption of hooks having the same probability of catching a turtle, regardless of the number of J hooks on the longline (Figure 11). For country C, the probability of at least one J hook per line with a turtle appears to vary with total hooks in the water (Figure 11); however, given the range of numbers of hooks (as compared to country B), this pattern may reflect something else entirely.

The overall patterns exhibited in these plots are consistent with results of fitting a simple logistic mixed-effects model to the data of each country, assuming hooks per line to be independent Bernoulli trials (Table 4). The model fitted separately to the data of each country was: logit(probability of turtle on a hook) = constant + hook type effect + β·total hooks in the water + line effect. The hook type effect was a fixed effect (0 = J, 1 = C16), and the line effect was a random effect. The coefficient associated with total hooks in the water was not significant for countries A and B, but was for Country C. Other than actually indicating an effect of total hooks on the probability of catching a turtle, this may reflect model misspecification, such as the need for spatial effects or other effects (*e.g.*, bait-type effects). Estimated line effects (random effect), grouped by trips and vessels (Figures 12-13), suggest that differences among vessels, and trips within vessels, may need to be taken into account, even with the lack of balance (Table 1b). Differences among trips within vessels may reflect differences in fishing location, bait, or other unidentified factors (which may or may not have been measured). The estimated scale parameter for all three fits was less than 1.0 (Table 4), which raises the question that, if after having accounted for other factors, such as nested effects of sets within trips (vessels), does an estimated scale less than 1.0 indicate significant under-dispersion or under-estimation with respect to the binomial model, and, if so, how would this best be addressed. Depending on the method of approximation (penalized quasi-likelihood, Laplace), variance estimates of the line effects varied considerably; estimated coefficients for fixed effects were more similar.

Marketable fish hookings may depart more from a simple Bernoulli model than do turtle hookings. As the number of J hooks on the line increased, the probability of at least one J hook per line with a fish appears to depart from the assumption of hooks having the same probability of catching a fish, regardless of the number of J hooks on the longline (Figure 14). However, a logistic mixed-effects model fitted to the data on presence/absence of fish catch (Table 5) did not always yield a positive coefficient for 'hooks in the water.' In contrast to models fitted to the turtle data, the estimated scale parameters were greater than one for the fits of Countries B-C, based on penalized quasi-likelihood and Gaussian quadrature. The estimates based on the Laplace approximation were enormous for all the fits of all three countries and the estimated variance of the random effects exactly the same, suggesting computational difficulties.

It may be more difficult to describe fish hookings with a simple logistic mixed-effects model than turtle hookings because some marketable fish species may be more likely to be aggregated (*i.e.*, occur in schools), and thus, when one fish is hooked, there are more likely to be other fish in the vicinity that get hooked. In addition, if densities of marketable fish species, collectively, are greater than turtle densities, it may be that the greater the number of hooks in the water (*i.e.*, the longer the longline), the more likely the longline is to be encountered by a school of fish and have at least one hook with a fish. Thus, the appropriateness of the simple logistic mixed-effects model may depend on local animal abundance, animal behavior (schooling *versus* solitary), and

possibly smaller scale habitat variability.

2.4. Methods used to estimate hook performance

Several different methods of comparing hook performance can be found in the recent reports and the peer-reviewed literature (*e.g.*, Bolton and Bjorndal, 2005; Watson *et al*. 2005, Kerstetter and Graves, 2006; Yokota *et al*. 2006; Kim *et al*. 2007). These have been pioneering studies on hook performance to reduce bycatch. As such, the statistical methods that have been used are listed below only to inform for the workshop discussion, with no intent to evaluate individual efforts. Some studies have first transformed the data on animals per hook (per line) using a natural logarithmic transformation (plus a constant) and then applied a paired t-test (or ANOVA model) to compare hook performance. Other studies have used a Friedman's rank sum test with blocking (on the line), contingency table and exact binomial tests, or a logistic model (without line effect). In these studies, as with the EPRSPT data, hookings were relatively uncommon (*i.e.*, a large percentage of the hooks had no animals), some species were likely to be schooling species (*e.g.*, yellowfin tuna), and control and treatment hooks were paired on the line (either by alternating individual hooks or alternating sections of a single hook type). In a few cases, the total number of hooks in the water varied among sets, and in some cases sets may have been nested within trips or vessels (although this aspect of the design was not always clear). In contrast to the EPRSTP data, these experiments were controlled, and in all cases the numbers of treatment and control hooks per line were similar. In addition to the methods mentioned above, for previous analysis of the EPRSTP data, a paired t-test on data limited to lines that caught at least one turtle was used. Because the EPRSTP data are highly skewed, longline gear differs among boats, and treatment and control hooks were sometimes, but not always, paired on each longline, it is believed that other methods are needed to adequately address these aspects of the data.

McCracken (in review) used Hawaiian longline catch data collected by observers to simulate sea turtle mitigation experiments for different sample sizes and experimental designs. The recorded catch for each line was randomly assigned to a treatment according to the experimental design being simulated. The catches of 10 different species were used to examine the behavior of different models and corresponding tests of treatment effects over different catch distributions. Linear and generalized linear models and expansions of these models for mixed models, overdispersed data, and zero-inflated data where examined. For moderate overdispersion, the majority of commonly-used tests of a treatment effect failed to achieve the nominal test level. Diagnostics indicated problems estimating variance and/or achieving the assumed asymptotic distribution. However, the estimate of the magnitude of the treatment effect typically appeared to have negligible bias.

2.5. A starting point for a model for hookings?

In order to be able to compare hook performance for the types of data collected by the EPRSTP (and other hook performance studies), it seems useful to try to develop a probabilistic model for hookings. In particular, as regards data collected by the EPRSTP, there is specific interest in understanding how best to accommodate unbalanced data (with respect to trips and vessels), and data from longlines of varying total hooks and numbers of treatment and control hooks. A simple starting point for a conceptual model might be the following:

p(hook of type k on line i caught an animal) =

      p(hook of type k caught an animal | hook of type k was encountered by an animal) ·

p(hook of type k was encountered by a animal | animal(s) encountered line i) ·
p(animals(s) encountered line i)

Some comments with regard to the components of this model follow.

p(hook of type k caught an animal | hook of type k was encountered by an animal) may be what ideally would be compared between two hook types to measure their relative performance. However, it seems this quantity is not likely to be directly measurable.

Animal behavior in response to encountering the line may affect p(hook of type k was encountered by an animal | animal(s) encountered line i). If animals do not move along the line, but merely interact with the line at the point that the line was intercepted, it might be assumed that hooks on a given longline function independently. However, if animals move along the line, investigating multiple hooks, individual hook outcomes may not be independent. If such is the case, it may be that the arrangement of hooks on the line (alternating individual hooks, alternating sections of hooks of the same type) and the total numbers of treatment and control hooks on the line should be considered. Whether hooks act independently may also be affected by the distance between hooks.

The total number of hooks on the longline (total longline length), perhaps in concert with animal density, may affect p(animals(s) encountered line i). For example, intuitively it seems plausible that a 10 nm longline may be more likely to be encountered than a 1 nm longline. Whether any effect of longline length on p(animals(s) encountered line i) can be detected may depend on animal density, which is unknown. If a species for which hook performance is to be quantified is relatively rare or has a patchy distribution, this may lead to a considerable proportion of the lines never encountering animals. In other words, there may be two types of zeros in the data: zeros that arose because animals were not in the area during the time the longline fished, and zeros that arose because no animals were caught, even though animals were in the area and investigated the line. With regard to hook performance, no experiment could have been conducted when animals were known not to have been in the area. On the other hand, very low numbers of encounters may not have any relationship to animal abundance, but may instead indicate that p(hook of type k caught an animal | hook of type k was encountered by an animal) is very small for all types of hooks.

2.6. Questions for the workshop

2.6.1. Sampling design/data collection

1) What is the optimal placement of control and treatment hooks on the longline when the goal is to compare hook performance: alternating individual hook types along the length of the line; same hook type per section (*i.e.*, between floatlines), but alternating hook types between sections, or some other design? What is an acceptable deviation from the perfect alternation?

2) Should there be restrictions on the range of total length of longline (total number of hooks) in the experiment and the total numbers of treatment and control hooks allowed on each longline? How would suitable ranges of line length and numbers of each hook type be determined?

3) Are there sampling concerns when only part of the longline in the water is used for a hook performance experiment; *i.e.*, when only part of the line has alternating hook types and the rest of the longline remains unchanged from its pre-experiment configuration?

4) Should (can) additional data be collected (*e.g.*, on turtles present but not hooked) to try to improve knowledge of when turtles were actually present during the fishing operations (*i.e.*, knowledge as to when a hook performance experiment was more likely to have been conducted)?

5) In terms of estimating necessary sample size for experiments, what is the appropriate sampling unit for the analysis (*e.g.*, total numbers of animals hooked, regardless of hook type; total number of longlines with at least one animal, total numbers of longlines, total number of hooks)?

6) What restrictions does a lack of random/systematic sampling place on the use of the data for estimation of hook performance? For example, given that fisher participation is voluntary, and therefore results are more difficult to generalize to the fleet than if the data were a random sample, should stricter control be placed on potential confounding factors (*e.g.*, alternating which hook type is closest to floatlines, maintaining the same bait type within a longline)?

2.6.2. Statistical methods for comparing hook performance

1) Is there a conceptually nested set of models that can be used to compare hook performance for both turtles and marketable fish species (*e.g.*, mixed-effects binomial/beta-binomial models; hierarchies of zero-inflated mixed-effects models)?

2) What type of test can be done to determine if the observed ranges of hooks in the water are acceptable from the point of view of comparing hook performance?

3) What type of test can be done to determine if the observed range of numbers of control and treatment hooks per line is acceptable from the point of view of comparing hook performance?

4) Can any type of test be done to look into hook interactions (*i.e.*, the assumption that hooks on the same longline function approximately independently)?

5) Given 2)-4), how should data from longlines be handled when only part of the line was experimental (*i.e.*, control hooks were replaced with treatment hooks on only part of the longline, but the entire longline was placed in the water)?

6) Are there any changes to the overall methodologies for comparing hook performance when there was more than one type of treatment hook on the line?

7) How should the nested structure of the data (sets within trips within vessels) best be accommodated in the analysis when some trips or vessels are represented by only one longline set?

8) How can potentially confounding factors (*e.g.*, mixed bait types within a line, different bait types among lines) be best accommodated in the analysis?

3. WORKSHOP MINUTES

The following is a summary of discussions that took place during the workshop. This summary has been organized by topic.

3.1. Experimental design

3.1.1. General discussion

Animal populations are often more patchy than realized. Setting hooks in an alternating pattern increases the odds that animals in a patch are exposed to all hook types. In addition, due to the variety of sources of variability within a set, an alternating hook design should be the optimal approach for exposing the different hook types to relatively homogeneous experimental conditions within the set. It is under this premise that this design has been frequently adopted. However, it is possible that there may be an interaction between hook types on the same longline. In theory, randomly assigning hook type to the branchline would 'average out' the effect of animal behavior, but it has been deemed an impractical design for many studies. Thus, the question becomes what is the best practical experimental design. It is possible that the bias introduced by the alternating hook design may be negligible, and the increased power obtained from this design outweighs its shortcomings.

3.1.2. Specific topics

3.1.2.a   Hook type allocation

Several participants believed that the best hook type allocation might change, depending on the goal of the experiment. In other words, the hook type allocation that would be optimal for determining whether there is a difference in hook performance might be different from the allocation that was optimal for estimating the magnitude of the difference in hook performance. It was noted that any type of experiment for which the longline has multiple hook types may represent a departure from the longline configuration typically used by fishermen. If such is the case, it is not clear how well the difference in hooking rates measured in any experiment will translate to actual fishing situations. This is a dilemma with any experiment for which the fishing practice is manipulated to a point that the experiment protocol would be unlikely to be adopted in the fishery. With this in mind, several hook type allocations were discussed. It was suggested that choosing between these different types of allocations might be best done by way of a simulation (see Sections 3.1.2.c(ii) and 3.2.2.f) because in a simulation other aspects of the study can be taken into account (*e.g.*, heterogeneity in hooking rates with different gear configurations, different vessel behavior, different fishing locations). It was also noted that the required sample size may change, depending on the hook type allocation.

3.1.2.a(i)   Alternating individual hooks

Some participants believed that an alternating design at the level of the individual hook (*i.e.*, J, C, J, C, J, C…) would perform well, regardless of whether the goal of the study was to determine if a difference in hook performance existed or to estimate the magnitude of a difference in hook performance. However, other participants believed that while the alternating design might be best for determining if there was a difference in hook performance, it might lead to biased estimates of the magnitude of the difference in hook performance if hooks on the same longline did not fish independently. In particular, if one hook type is more apt to hook an animal than another hook type, then having more than one type of hook on the line may lead to a biased estimate of the difference in hook performance. For example, suppose that a J hook retains turtles at twice the rate of a C hook. A turtle biting a C hook will be more likely to get away and could then move along the longline to encounter a J hook. A related issue is whether a hooking depends more on the hook type or more on whether the hook was encountered. That is, if the line

has only C hooks, would animals change their response to C hooks, as compared to their response to a longline that had both J and C hooks? The advantage of the alternating hook design, and the reason it may be useful for detecting a difference in hooking rates, is that by having different hook types as close together as possible, nuisance covariates (*e.g.*, spatial effects, seasonal effects) can be controlled.

Several participants suggested that, for the reasons noted above, it might be best to perform experiments in a hierarchical manner. In particular, it was suggested that one could start with experiments using alternating individual hooks to determine if there was, in fact, a difference in hook performance. This experiment could be followed with another experiment for which the goal was to estimate as accurately as possible the actual difference in hooking rates that would be realized by fishermen.

3.1.2.a(ii)   Other designs that could follow

3.1.2.a(ii)a Alternating sections of hooks

It was suggested that a sampling design that involved alternating sections of homogeneous hook types (*e.g.*, J, J, J, J, J, C, C, C, C, C, J, J, J, J, J, C, C, C, C, C, …..) might provide a less biased estimate of the difference in hooking rates between hook types. Potentially this design could also be used to study if there is an interaction between hook types. With hooks arranged in sections, it is less likely that any animal behavior (*e.g.*, swimming along the longline) will immediately lead animals to encounter a hook of a different type. If sections of hooks are small enough, nuisance covariates may still be similar among hooks in neighboring sections; long sections could lead to considerable variation in nuisance covariates among sections, for example, due to trends in environmental conditions from one end of the longline to the other. There was some discussion as to the length of the sections. It was commented that perhaps the catenary in the longline between floatlines should be taken into consideration when determining section length. For example, section length could be set to the length of mainline between floatlines. In this way, the full catenary in the longline between floating lines would be all of the same hook type, possibly avoiding any confounding between hook depth and hook type. It was also noted that for species of sea turtles that tend to be entangled rather than hooked, it had been suggested previously that having the same hook type per section may be preferred because it can be difficult to determine which hook type was associated with entanglement when nearby hooks are of different types (entangled turtles may be wrapped in the longline and have multiple hookings). It was suggested that the size of the same-hook sections might be determined through simulation.

3.1.2.a(ii)b Total conversions

It was suggested that comparing hooking rates for a given hook type on a longline with alternating individual hooks (*i.e.*, J, C, J, C…) to the hooking rate of the same hook type on a line with only that hook type (*i.e.*, a longline with only J hooks or a longline with only C hooks) might be informative as to any magnitude of bias in hooking rates computed from lines with alternating hooks. It was noted that if fishermen typically use only one hook type on a longline, then experimental lines with only one hook type would most closely reflect an actual gear configuration and if data were available from longlines with only J hooks and longlines with only C hooks, the hooking rates from these two types of lines could be compared against differences obtained from lines with alternating hooks. (This is complicated by the fact that as hooks are lost in the course of fishing activities, fishermen may replace lost hooks with whatever hook type they have on hand, and these may differ in size and type from the original hooks on

the longline.) It was noted, however, that the effects of nuisance covariates (*e.g.*, spatial effects) may cause a problem in this type of a comparison because hooks of different types will not be on the same line. As a result, to make a meaningful comparison (*i.e.*, one that is not overly influenced by other factors), a very large sample size may be required. Necessary sample size requirements could be explored through simulation.

3.1.2.a(ii)c Randomized placement

It was suggested that the best way to avoid bias due to hook interactions would be to randomly allocate hook types to different places on the longline. It was noted, however, that the sample size requirement for a random hook type allocation might be larger than for a systematic allocation (*i.e.*, alternating hooks, or alternating sections of homogeneous hook types, 3.1.2.a(ii)a). It was also commented that, in spite of its attractiveness from a statistical perspective, random allocation is not likely to be practical in a voluntary hook exchange program, although it may be feasible in a controlled experiment

3.1.2.b Collection of individual-hook data

It was suggested that collecting data at the hook level might be worth consideration. One purpose for collecting such data would be to have the ability to apply a case-control matched-pairs analysis. In a designed experiment, treatment and control hooks could be randomly placed on the longline. However, it was suggested that this might not be necessary and that an alternating hook design might be acceptable once the details of selecting the 'matched pair' had been determined. (See Section 3.2.2.d for a discussion of the analysis of longline data using a case-control type of approach.) It was further noted that collection of individual hook level data would be useful in and of itself, regardless of whether a case-control type analysis were done. One reason to collect data at a finer level than the whole longline is to be able to detect and adjust for any within-line correlation among the hooks. It may or may not be possible to adjust for correlation among the hooks when the data are aggregated by longline. However, if the data are available at the hook level, then between-hook correlation could be estimated and adjusted for in an efficient manner. Of course, a possible drawback of collecting individual hook data is that it may complicate the data collection process.

3.1.2.c Sample size

3.1.2.c(i) Candidate sampling units

Several different sampling units for analysis were suggested, including number of turtles, and numbers of longline sets (numbers of trips) that will likely give a minimum number of turtles. It was also suggested that there should be an investigation as to whether to sample more trips of fewer vessels or more vessels but fewer trips per vessel, or in the case of a mothership-type fishing operation, whether to sample more sets of fewer trips or more trips but fewer sets per trip. A determination of how many sets per trip (trips per vessel) should be based on an assessment as to whether hookings were more variable within or among sets, trips, or vessels. It was suggested that the most appropriate sampling unit might be determined from simulations using existing data.

3.1.2.c(ii) Estimation of sample size

It was suggested that determination of the sample size should be based on simulations with existing data. Simulations should be designed to take into account any spatial structure in the hooking data, if it exists (see Section 3.2.2.f). The simulation model should be kept simple (at

least to start). For a hypothetical difference in hook performance, the idea would be to run the simulation model a large number of times at different sample sizes (*e.g.*, different numbers of longline sets) to see how likely it was with different sample sizes that different hypothetical hooking rate differences would be detected.

3.1.2.d   Stratified sampling

It was also noted that if there exists spatial patchiness in hooking rates or differences in hooking rates among vessels (*e.g.*, perhaps owing to longline configuration), stratified sampling (*e.g.*, by area or vessel characteristics/longline gear characteristics) should be considered. As noted in Section 2, there appears to be variability in longline configurations, both among countries and within countries. Simulations with existing data could be used to determine if stratification would be useful to improve detection of any differences in hooking rates.

3.2. Hypothesis testing and estimation

3.2.1.   General discussion

For hypothesis testing, it is a good idea to have at least two types of analyses: one simple, and one more complex. The simple test might be something such as a randomization test or a Mantel-Haenszel test, because, if the presumed experimental layout was largely implemented, individual hooks would have been naturally paired within a set, mostly adjusting for covariates without the need to expressly identify the covariates and/or their relationship to the probability that a hook caught an animal. In addition, any variation of the hierarchical design can be easily taken into account with a test such as a randomization test through the randomization procedure, as compared to more complex models (*e.g.*, mixed-effects models). More complex models would allow for explicit modeling of covariates, which might improve the power to detect differences, but only if the covariate structures are known well enough to be modeled properly. Between the simplest test, such as a randomization test, and detailed and complex models such as zero-inflated hierarchical spatial models, there is a range of testing and modeling options. Regardless, with the presumed sampling design and in the absence of any convincing way to determine whether the hooks behaved independently, only relative differences (J hooks *versus* C hooks) can be tested and estimated.

3.2.2.   Specific topics

3.2.2.a   Randomization procedures

Randomization methods make minimal assumptions about the data, and do not attempt to explicitly model the processes underlying the data. For these reasons, randomizations tests may not be optimal for estimating the relative magnitude of hooking effects. This is likely best done with a more complex model. However, to test the general null hypothesis of no difference in hook performance, a randomization test is a quick and straightforward method for testing the null hypothesis, and the results of randomization tests are easy to interpret.

3.2.2.a(i)   Hypothesis testing

Because there can be no control of potential confounding factors with a randomization approach, only the average relative effect can be tested (*i.e.*, implicit in the use of a randomization test is that it is assumed that the differences in hooking rates are constant from data unit to data unit). The question of interest is: Is there a difference in hooking rates between J and C hooks? Thus, the null hypothesis is, H_o: the 'hook type' effect = 0, or equivalently, H_o: the hook type ('label')

makes no difference to the relative hooking rate.

To test this null hypothesis, permutations of the data should be generated in accordance with the study design, assuming no hook effect. The actual method of randomizing the data will depend on the experimental design and should preserve the data structure. For example, in experiments for which actual hook placement on the line is known (*e.g.*, alternating individual hooks, alternating sections of homogeneous hooks), randomization could be done at the hook level or the section level. By way of example, suppose that exactly equal numbers of each hook type occur on the longline, and that J and C hooks are strictly alternating (*i.e.*, J, C, J, C…), with only one C hook type. In this case, once the hook type of the first hook on the longline has been determined, the placement of the rest of the hooks (by type) along the line is fixed. Thus, to randomize data collected for a longline set with this design, one needs only to randomize the first hook on the line (J *versus* C) with probability 0.50, and the layout of the rest of the hooks follows automatically. Note that in this example randomizing the hook type of the first hook is equivalent to randomly assigning, with probability 0.50, the column labels "J" and "C" of a 1 x 2 contingency table that contains the number of turtles on the line caught on each hook type. In other words, if the only data available are the numbers of hooks of each type with turtles (per longline set), the randomization procedure that assumes equal numbers of perfectly alternating hooks can still be applied. To apply the randomization procedure to a collection of sets, the hook type of the first hook of each set (or the column heading of each 1x2 table) should be randomized separately.

In experiments for which longlines have different numbers of J and C hooks within the same line, and for which the actual hook placement on the line is not known, randomization of the data must be handled differently. The null hypothesis states that turtles are equally likely to be hooked on J hooks as C hooks. If such is the case, assuming only two hook types for each set, each hook that caught a turtle would be randomly assigned a hook type, either a J or C, based on the proportion of J and C hooks on the longline. For example, suppose the longline set had 100 J hooks and 200 C hooks, and that five turtles were caught, three on J hooks and two on C hooks. To obtain a randomization of these data under the null hypothesis, each of the five hooks that caught turtles would be randomly assigned a hook type according to the following probabilities: 1/3 for J (= 100/300) and 2/3 for C (=200/300). To randomize data from multiple longline sets, where each set had different numbers of J and C hooks per longline, these probabilities must be computed separately for each longline. Note that by randomizing the labels of hooks that caught turtles (without randomizing the total hooks by type), we are testing for additive hook effects, not interactions.

It was also suggested that in the case of unequal numbers of J and C hooks on the longline (and unknown hook placement), data for each set might be randomized by assigning the two proportions of hooks with turtles to either "J" or "C" with probability 0.50 (on a per-set basis). However, some commented that this type of randomization does not appear to be in keeping with the assumption under the null hypothesis that turtles are as likely to be hooked on J hooks as they are on C hooks.

There are several possible summary statistics that could be used to test the null hypothesis. In the case of equal numbers of J and C hooks on the line, the sum (across sets) of differences in numbers of hooked animals (*i.e.*, number of animals on J hooks minus the number of animals on C hooks) could be used as a summary statistic. For lines with unequal numbers of J and C hooks, the sum across sets of the differences in the proportions of hooks with animals could be used.

Ideally, the power of various summary statistics to detect a difference should be taken into consideration when selecting a summary statistic. This is an area for further research, and would best be explored through simulation, particularly for longlines with unequal numbers of hooks by type and unequal total numbers of hooks. If the total number of hooks per longline differs greatly among sets, it is not necessarily true that different summary statistics will give the same results with a randomization test. Simulations involving different numbers of J and C hooks on the same line, and different line length (*i.e.*, total number of hooks on the line, assuming the same spacing between hooks for all longlines), should be used to identify the most statistically efficient summary statistic.

For general information on randomization tests (including the number of required randomizations and computation of p-values), see, *e.g.*, Manly, 2007.

3.2.2.a(ii)   Confidence interval

Following general randomization procedures for constructing confidence intervals (Manly, 2007) and the permutation procedures outlined above, confidence intervals based on randomization methods can also be computed.

3.2.2.b   Mantel-Haenszel test

Mantel-Haenszel (M-H) procedures provide for estimation and testing of average partial association between two categorical variables (that is, within a two-way contingency table) conditional on the levels of third variable, whose levels are often termed 'strata.' In other words, M-H methods are appropriate for analyzing the association within K 'stratified' 2 x 2 tables. In fact, the strata may be defined as the combinations of the levels of multiple variables, so these methods have application beyond the three-way table setting. Typically, variables that define the strata are nuisance factors or potential confounders in the two-way relationship of interest and may or may not be actual strata in the sense of a stratified sampling design. In this context, M-H estimators exist for a variety of measures of association in a two-way table, but the most commonly used choice is the M-H odds ratio. For K 2 x 2 tables, this statistic can be thought of as a weighted average of the K conditional odds ratios within each of the strata, and it functions as a measure of average conditional association between rows and columns of the tables. The M-H test statistic (also known as the Cochran-Mantel-Haenszel test) tests the hypothesis that this average odds ratio is one (no conditional association). The M-H test assumes additivity (*i.e.*, no interaction between strata, rows and columns). Thus, the test is appropriate when the true association is of the same sign and of similar magnitude from table to table. The p-value for a M-H test can either be based on a permutation or randomization distribution or on a large-sample chi-square approximation. When a randomization distribution is used, this becomes a randomization procedure (Section 3.2.2.a(i) above), for which the test statistic is based on sample odds ratios.

In the context of longline turtle bycatch data, the M-H odds ratio estimator and test statistic could be used to investigate conditional association between hook type and hooking events, following stratification by a number of potential covariates (or "control variables"). The finest level of stratification is by set, which would provide control for heterogeneity across all between-set factors. In other words, it may be most sensible to take K to be the number of sets. A reason for doing this would be to try to control for differences among sets, for example, in terms of longline configuration (*e.g.*, total numbers of hooks on the longline), amount of fish caught per longline, fishing strategy (*e.g.*, use of a mother ship *versus* individual small vessel), fishing location, or

other potentially confounding factors that varied among sets. It was suggested that it might also be possible to take K to be the number of trips or boats, *i.e.*, pool data across sets within a trip or trips within a boat. It is noted, however, that in pooling data across sets of the same trip, or trips of the same boat, the linkage between hooks of the same longline would be lost, and there would also be a loss in power (compared to analyzing the data at the level of the set).

It should be noted that there is a close connection between M-H methods and the permutation/randomization procedures of section 3.2.2.a(i). For stratified 2 x 2 tables, the full permutation distribution of the M-H odds ratio estimator is available and is a hypergeometric distribution. Thus, the so-called 'exact' p-values and confidence intervals for M-H tests and estimators that are available in standard software are based on the hypergeometric distribution. If the strata are defined to be individual longline sets, exact M-H results are equivalent to those that would be obtained by generating a permutation distribution by holding the locations of hooked turtles on a line fixed and permuting all the hook labels on that line. Note that the validity of such procedures is based on an assumption of independence and constant hooking probability within a line, though extensions of M-H methods do exist for cases when these conditions are violated.

For general information on the M-H test, see, *e.g.*, Agresti, 1996. Note that the M-H test can be generalized to K m x 2 tables, for data collected on longlines with m hook types (*e.g.*, two C treatment hook types, and one J control hook type).

3.2.2.c   Bayesian (simple)

Assuming only two hook types, it should be straightforward to construct a two-dimensional posterior distribution of hooking probabilities. For each longline set, the data would be the numbers of hooks of each type, with and without turtles. The conditional distribution of the data would be assumed to be binomial, and the prior a uniform distribution in two dimensions. In other words, for a given set, let X_1 = number of J hooks with a turtle, X_2 = number of C hooks with a turtle, N_1 = total number of J hooks on the longline, and N_2 = total number of C hooks on the longline. If we assume $X_1 \sim$ binomial(N_1, p_1), and independently, $X_2 \sim$ binomial(N_2, p_2), we can estimate a two-dimensional posterior distribution for the hooking probabilities p_1 and p_2, and can then see how the posterior density is distributed in relation to the line $p_1 = p_2$ to determine if there is a difference in hooking probabilities. Under this model, the prior would be P(p_1,p_2), the likelihood L($X_1,X_2|N_1,N_2,p_1,p_2$) and the posterior P($p_1,p_2|X_1,X_2,N_1,N_2$) proportional to L($X_1,X_2|N_1,N_2,p_1,p_2$)·P(p_1,p_2). Evidence in favor of a lower hooking rate for C hooks is obtained from the probability P($p_1>p_2|X_1,X_2,N_1,N_2$) or from the odds ratio (against) P($p_1 \geq p_2|X_1,X_2,N_1,N_2$)/ P($p_1<p_2|X_1,X_2,N_1,N_2$). A posterior density for the difference (p_1-p_2) could also be easily obtained.

It may be useful to compare the results obtained from Bayesian posterior inference using the two-dimensional posterior to that from the randomization test. If hooks were grouped by type in sections, this might change the Bayesian approach. However, the first step could still be to construct the simple two-dimensional prior, and then more specific conditions could be incorporated to see if they matter.

Another slightly more complex possibility would include the use of trinomial distributions since each hook can be classified as: with turtle, with marketable fish, or empty/unmarketable fish. Now, X_1 becomes a vector with two components: number of J hooks with a turtle and number of J hooks with a marketable fish. Similarly X_2 = (number of C hooks with a turtle, number of C hooks with a marketable fish). We may then assume $X_1 \sim$ Mult(N_1, p_{11}, p_{12}) with p_{11} and p_{12} the

J-hooking probabilities for turtles and for marketable fish, respectively; similarly, for C hooks we may assume $X_2 \sim \text{Mult}(N_2, p_{21}, p_{22})$. Priors can be modeled as the product of Dirichlet distributions. This model is superior to an analysis based on the binomial distribution because it takes into account the dependence among hooks. Hook performance could be evaluated through the posterior probability of lower turtle hooking rate and simultaneous higher fish hooking rate for C hooks. This is expressed as $P(p_{11} > p_{21}, p_{21} < p_{22} | X_1, X_2, N_1, N_2)$.

For general information on the Bayesian approaches, see, *e.g.*, Congdon, 2002, and Gelman *et al.*, 2004.

3.2.2.d   Case-control models (matched-pairs retrospective analysis)

If data were available at the level of each hook, for hooks *both* with and without animals, then a modification of the standard case-control retrospective analysis approach might be developed for longline data. This type of analysis might be particularly useful when the alternating hook design could not be followed exactly (*i.e.*, the placement of hook types along the longline does not follow a regular pattern), and/or when it is believed necessary to take into account catches of other species when analyzing hook performance with respect to turtles. In the medical sciences, case-control retrospective studies are often used because the case-control design ensures a sufficiently large sample of subjects with the disease or medical condition of interest, and the matched case-control pairing facilitates control of nuisance covariate effects. In the simplest case, a case-control analysis models the odds ratio of two variables, say Y and X. Because the value of the odds ratio is the same regardless of whether it is based on the conditional distribution of Y given X or *vice versa* (*e.g.*, Agresti, 1996), it is not necessary to specify which variable is to be considered the response and which is to be considered the covariate prior to data collection. For this reason, a case control analysis might be applied to existing longline data that were processed in a retrospective manner.

To apply the case-control approach to individual hook data for turtles, it is necessary to define the 'case' (and its match, the 'control') and the 'exposure.' It was suggested that each hook with a turtle could be considered a 'case.' If hook types had been randomly placed on the longline, then the match to each case might be defined as the closest neighboring turtle-free hook (in one direction or the other). The 'exposure' would be the hook type. For example, for a longline with only two hook types (*e.g.*, one type of J hook and one type of C hook), each hook with a turtle (a case) and its matched-pair turtle-free hook (a control) would generate a 2 x 2 table of 0-1 data. The table columns would be whether a turtle was hooked (*i.e.*, Y-case / N-match), and the rows would be the hook type of each hook (J or C for each of the case and the matched control). In this way, each hook with a turtle and its matched turtle-free hook defines a stratum on the longline, helping to control for nuisance covariate effects. Stratification in this manner might help to divide the longline into segments, where hooks within each segment have an approximately equal probability of hooking an animal. This might be particularly important if hooked animals are clustered on the longline.

How best to determine the 'match' needs to be given some consideration. The goal is to select a match (control) so that the matched set is as similar as possible in all other aspects. In standard case-control analysis, the marginal distribution of the variable regarded as the case is considered fixed (*e.g.*, by the sampling design) and the exposure is assumed to be a random variable. The closest neighboring turtle-free hook might be a reasonable choice for the match. However, if J and C hooks were placed in a truly alternating manner on the longline, then the exposure may not

be random because hook types were pre-determined to be alternating as part of the sampling design. It was suggested that a possible solution to this problem might be to use groups of hooks to determine the match for each case. In other words, for every hook for which there was a turtle, randomly select as the control hook a hook from among m hooks (without turtles) located either forward or backward of the target hook. It was noted that if the m hooks have greatly unequal numbers of J or C hooks, this may be a problem. It was commented that if data obtained from a designed experiment for which the hooks of different types were randomly placed on the longline, one could take m = 1 and select the control hook for each case to be either the next or previous hook with probability 0.5. The question was raised as to how to select the match when the next hook has a fish on it. (It was noted that the only factor that is impossible to control is whether a hook catches an animal.) It was suggested that the match might be the next empty hook. Concern was expressed that because of the catenary shape of the longline between floatlines, neighboring hooks may fish at different depths, and this might complicate selection of matched pairs. It was noted that a catenary effect might be minimized by careful specification of m. It was also noted that matched case control designs can be implemented by matching more than one control to each case. Such a design might help to avoid any bias caused by differences in hook depths between floatlines because each case could be matched to the next turtle-free hook and the previous turtle-free hook. Although this 1:2 matching might balance any depth effect, further consideration should be given to the choice of matches to confirm that depth and hook type effects would not be confounded.

For general information on the case-control design and its analysis, see, *e.g.*, Schlesselman, 1982. Multiple-category (*e.g.*, more than more type of treatment hook) extensions are available for the standard case-control model and might be modified for use with longline data (*e.g.*, Agresti, 1996). Multiple control to single case matching is described in Stokes *et al.*, 2001.

3.2.2.e   Complex models

In principle, development of complex models to analyze the data would be useful for the following reasons: 1) if enough is understood about the processes that generated the data, complex models can be used to control for effects of other covariates in testing hypotheses and for estimation of the relative difference in hook performance; 2) even if the complex model gets the dynamics wrong, it can be useful for simulating data to explore both the data generation processes and to see how non-parametric tests (*e.g.*, randomization procedures) work for different data structures; and 3) complex models would be useful for translating results to other oceans.

A frequentist approach to complex modeling would probably involve some type of nonlinear mixed-effects model (*e.g.*, zero-inflation; random effects to capture nested structure of hooks within sets within trips). Although no formal testing was done, data on hookings presented in Section 2 appear to be zero-inflated, at least for marketable fishes (frequency distributions appear highly skewed relative to a Poisson distribution). However, there are several challenges to fitting nonlinear mixed-effects models to these data. As an alternative, in order to deal with the apparent zero-inflation, the hooking data could be aggregated, but this has the disadvantage of aggregating over potentially important covariate effects and has the potential to introduce ecological bias. It was noted that hierarchical Bayesian spatial models might be fruitfully explored, and perhaps more tractable with currently available software than nonlinear mixed-effects models. From a Bayesian viewpoint the complexity in the data structure can be best described in a hierarchical-Bayesian (HB) model (Clark, 2005). Differences among countries, vessels, and sets within

vessels should be included to extend the simple Bayesian models described previously. Different levels of added complexity should be evaluated with DIC (Spiegelhalter *et al.*, 2002) in a similar fashion as AIC is used in a non-Bayesian context.

The following paragraphs describe concerns regarding the implementation of nonlinear mixed-effects models. First, it may be difficult to determine how to parameterize the zero-inflation because of incomplete knowledge about the processes that generated the data. A starting point might be to add zero-inflation and/or over-dispersion to a basic binomial model. The basic binomial model might take the following form for a comparison of two hook types: for the ith set, X_i = number of hooks of a given type with a turtle, N_i = total number of hooks of a given type on the line, and we assume $X_i \sim$ binomial(N_i, p), with logit(p/[1 - p]) = constant + hook type effect + set effect, for which the hook type effect is a fixed effect and the set effect is a random effect. A common version of the zero-inflated model has a constant mixing probability uninfluenced by covariates. However, this is an extremely strong assumption, and it may be more realistic to allow heterogeneity in the zero-inflation probability due to both observed and unobserved covariates. An alternative that was suggested was to fit a truncated model only to data of longlines that caught the taxa of interest (*e.g.*, fit a truncated Poisson or binomial to the positive-valued observations for marketable fishes). If the true mixture model depends on covariates, but a truncated Poisson is fitted to the data, this is equivalent to assuming the mixing probability is constant. This may undo any advantage of fitting a truncated model because the true mixing probability might depend on covariates. In some participants' experience, hurdle and zero-inflated models gave the same results. However, other participants cautioned that if there exist influential covariates, the results will not be the same for the two models. An important consideration when fitting a zero-inflated model with covariates is that specifying the form of covariate effects and covariate selection can be difficult. There could be different covariates for the mixing probability (*e.g.*, spatial effects, total numbers of hooks) and the hooking rate, but there may be little information in the data if hookings are rare events, as is the case for turtles. In other words, if there are too few non-zero observations in the data, it may not be realistic to use a zero-inflated model. Little work on model selection has been reported in the statistical literature on zero-inflated regression. Therefore, choosing a suitable model in the presence of many potentially important covariates and multiple applicable model classes (hurdle, zero-inflated Poisson, zero-inflated negative binomial, *etc*.) will be a formidable challenge.

Second, concern was expressed about the need to model potential interactions among the catches of various taxa on the longline (*e.g.*, is turtle bycatch independent of marketable fish catch?). In some longline fisheries (*e.g.*, the Hawaiian fishery for swordfish) there is a positive correlation between catches of different taxa, which may be related to habitat. On the other hand, it was wondered whether it was plausible that a negative association might result from animals thrashing on the longline or predator/prey dynamics, interfering with hookings on neighboring hooks. Such a negative association might add to under/over-dispersion. One suggestion was to model the data with a multinomial instead of a binomial distribution. However, it was noted that a multinomial could be replaced with a shared random effect (*i.e.*, a set effect) or perhaps by adding an over-dispersion factor. It was suggested that for a simpler model, all that might be needed would be to include a fixed-effect predictor for presence/absence of catches of other taxa. It was also suggested that since over-dispersion with binomial data leads to estimation based on a quasi-likelihood approach, it might be better to consider zero-inflated marginal models (*e.g.*, Hall and Zhang, 2004). Such models would account for within-cluster (*e.g.*, within-set) correlation, as would mixed models with set effects, but marginal models make weaker assumptions about the

nature of the within-cluster correlation and account for additional over/under-dispersion without making specific parametric assumptions about the data-generating mechanism that leads to this phenomenon. A further advantage of these marginal models is that they are easier to fit than their mixed model counterparts; marginal models insert a working covariance structure into the generalized estimating equation (GEE)-like score functions that arise in the 'M' step of an EM algorithm. Unlike mixed models, there is no need for the computationally intensive calculations involved in integrating over a random effects distribution.

Finally, it was noted that p-values from parametric tests based on complex models may be misleading. Simulations suggest that while complex models may capture the mean structure, nominal p-values obtained from these models may be misleading because the standard errors may not be accurately estimated or the assumed asymptotic distribution not met because of relatively small sample sizes or distributional violations. This implies that testing for covariate effects, in addition to relative hook performance, could be problematic. To deal with this problem, it was suggested that a bootstrap or randomization procedure might be used. To test relative hook performance with a complex model, one could randomize the labels of the hooks (re-fitting the model each time) and obtain the p-value for a hook effect from the randomization distribution. It was also noted that one could randomize covariate by covariate to test covariate effects. (Testing covariates effects is essential if the results are to be transportable to other oceans.) There was limited discussion on the feasibility of an empirical bootstrap to get confidence intervals for covariate effects. An empirical bootstrap would presumably be based on resampling of residuals. However, if a zero-inflated model was fitted to the data, it is not clear how to define residuals. Without a proven procedure for constructing and resampling residuals, an empirical bootstrap procedure would not be a sensible approach.

Given the above concerns, it was believed that there should be three steps in developing a complex model for this type of data: 1) randomization/M-H tests for temporal-spatial homogeneity (see also below); 2) tests from a "regression-type" model to explore correlation structure in the data; and 3) integrating results from steps 1 and 2 into a more complex model to describe the processes generating the data. Going beyond a purely frequentist approach, Step 3 might involve generating a space-time distribution of turtle relative abundance (perhaps drawing on ancillary information), adding a fishing model, and the probability of hooking due to different hook types, getting a posterior distribution of hooking probabilities and performing inference in a Bayesian manner.

General information on mixed-effects models can be found in, *e.g.*, Pinhero and Bates, 2004, and Demidenko, 2004.

3.2.2.f   Spatial modeling and simulations

It was suggested that as an alternative/complement to a nonlinear mixed-effects model or other option (*e.g.*, Bayesian inference), or as a means of studying how to improve sampling design, spatial modeling should be considered. Spatial modeling would involve development of spatial point processes for each taxa (*e.g.*, marketable fishes, turtles) separately and then integrating these point processes into one spatial model. As a first step, it was recommended that the question "Are sets made in random places or, for example, is fishing location related to the environment?" be addressed. In other words, is there spatial homogeneity in the catch and bycatch data? A randomization test or M-H test could be used to determine whether sets with turtles or other taxa were close in space (and/or in time). If there exists spatial clustering, it is

better to develop a clustered point processes for each taxa than to use a basic point process. If bycatch/catch distributions could be assumed to be related to true abundance, these distributions could be used to specify the parameters of spatial point processes.

In order to specify spatial point processes for the whole region occupied by the fishery, it was suggested that environmental data (or other ancillary information) might be used to fill in where no fishery data were available. It was noted that smoothing methods (*e.g.*, kriging, generalized additive models, *etc*.) all try to get at filling-in data in a smooth way. This may not be desirable if the environment has discontinuities (*e.g.*, strong fronts) that lead to discontinuities in catch/bycatch distributions. In this regard, ancillary information, such as environmental data, might be used to facilitate spatial modeling in a more sensible/informed manner. It was noted that many complex environmental models are deterministic, which is undesirable. For example, a deterministic model would not allow for Bayesian inference. For simplicity, it was suggested that the spatial model could be built on a rectangular grid (*e.g.*, 1° areas). It was commented that any conclusions would be predicated on the structure of the underlying point process; however, it was noted this is always a problem when constructing models.

Complexity would be best added in layers, and existing data should be used to make the simulation as realistic as possible. For example, an estimate of the mean hooking rate could be used to parameterize a Poisson distribution or the proportion of hooks with animals could be used to parameterize a binomial distribution. This information would be combined with information on spatial structure obtained from randomization/M-H tests to specify the parameters of a spatial point process. First spatial structure, then other factors such as differences in fishing behavior or longline gear or bait, could be incorporated into a spatial model. It was noted that spatial models can become extremely complex; the goal would be to capture reality, without cumbersome complexity. Ideally, the spatial simulations would run relatively quickly so that it would be practical to use simulation results with Bayesian inference.

In terms of studying sampling design, the suggestion was to simulate data and explore the ability of different sampling designs to recover known (hypothetical) differences in hook performance. Simulated data from a spatial point process would be "sampled" to see how sampling design, including differences in longline configurations (*e.g.*, numbers of control and treatment hooks on the same line or differences among longlines in terms of line length (total numbers of hooks)), and number of samples, affect detectability of hypothetical differences in hook performance. Assuming that the true probability that a hook catches an animal (given animals were in the area) does not depend on turtle abundance, it may be possible to use this spatial model to establish whether a test for hook performance is sensitive to spatial patchiness or other factors. With a spatial simulation one can explore design questions such as: with a particular sampling design (and/or range of longline gear configurations), can a difference in hook performance be detected, given spatial structure?

3.2.3. Miscellaneous topics

3.2.3.a   Testing hypotheses of equivalence

It was noted that the formulation of the null hypothesis for testing hook performance with respect to catches of marketable fish should be different than the formulation of the null hypothesis for testing hook performance with respect to catches of turtles. Because the desire is to establish that C hooks can be substituted for J hooks with no appreciable disadvantage to the fishing industry, the appropriate null hypotheses should be that the hook types are 'not equivalent,' (or more

specifically, that C hooks produce lower catch rates) the reverse of the usual null hypothesis. The issue here is similar to that which arises in the comparison of generic equivalents to name-brand drugs in the pharmaceutical industry. In that setting it is desired to establish that a generic drug is `bioequivalent' to the existing drug. To establish equivalence between the two hook types, an alternative hypothesis must be formulated that assumes that the catch rates for the hook types are within a tolerance of each other that is small enough to be practically insignificant. It was cautioned that such a hypothesis may be difficult to agree upon because of the small profit margins in the artisanal fishery which is likely to make fishers reluctant to accept any reduction in marketable fish catch rates, no matter how small. An additional challenge to testing equivalence is that the necessary sample size to test such a null hypothesis depends on how "equivalence" is defined and is typically much larger than in traditional hypothesis testing. It was also commented that equivalence may not be testable using the data collected under the sampling designs currently in use. Some participants were of the opinion that the current sampling method requires the assumption that the fish, turtles and hooks are acting independently. It was believed by some participants that this may not be tenable and should be addressed explicitly. It was suggested that because fishers were unlikely to tolerate any reduction in marketable fish catch rates with the new hooks, it may be suitable to use a null hypothesis of equal or lower catch rates versus a one-sided alternative of higher catch rates with C hooks. In this case the increased sample size requirements and other complications of equivalence testing could be avoided.

It was noted that an equivalence hypothesis can also be tested using a Bayesian approach. A Bayesian approach has the advantage that one demonstrates the chance of obtaining a smaller catch by assigning it a probability. A Bayesian approach would involve generating a posterior distribution for the probabilities of hooking (*i.e.*, a two-dimensional posterior distribution for hooking probabilities (p_1, p_2) if there were only two hook types on the longline). The proportion of the area in the 'tail' of this surface provides information on the likelihood of the null hypothesis. A Monte Carlo simulation could be carried out to obtain a measure of uncertainty as to this 'tail' area.

It was suggested that fish size could also be considered when evaluating a null hypothesis of no effect of hook type on catch rates. Length information should be part of any equivalence testing. This could be done by incorporating length information into a loss function. It would be nice to incorporate information on market value into the loss function as well; however, this may not be practical. Length information may serve as a good proxy for economic value and conservation value. The loss function could be based on the expected value of catch. It was commented that constructing a specific loss function would be important if one wanted to decide among different alternatives. However, at this point, this was perhaps a step ahead.

General information on testing hypotheses of equivalence can be found in, *e.g.*, Patterson and Jones, 2006, or Wellek, 2002.

4. CONCLUSIONS

The main conclusions of the discussions of the workshop participants were as follows.

1) The requested sampling design of alternating individual hook types is probably sufficient if the goal is to establish whether there is a difference in performance. However, in the presence of interactions between hooks on the same longline, this sampling design is likely to yield biased estimates of the magnitude of the difference in hook performance. If that is the case, a follow-up experiment with a hook placement design less susceptible to hook interactions should be

considered, particularly if accurate estimation of the magnitude in the difference of hook performance is an objective.

2) Collection of data on individual hooks and development of case-control methods for longline data were suggested as a means of improving ability to detect differences in hook performance for rare events, particularly when the placement of hook types along the longline does not follow a regular pattern and/or when it may be important to consider interactions among species.

3) Because the dynamics of turtle interactions with longlines are not well understood, and because hookings are relatively rare events, developing complex models for hooking data will likely be difficult. Moreover, complex models may require assumptions unlikely to be valid in most situations when using the current design for testing differences in hooking rates. Simple tests of hook performance, based on a randomization procedure or a Mantel-Haenszel test, should be conducted before attempting to develop complex models.

4) Simulations with existing data should be conducted to develop a better understanding of any effect of differences in longline configuration, or departures from the specified sampling design, on measured differences in hook performance. Simulations with existing data should also be conducted to determine the best sampling unit, and to estimate sample size for future experiments. As part of these simulations, tests for spatial-temporal homogeneity in the data should to be performed.

ACKNOWLEDGEMENTS

REFERENCES

Agresti, A. 1996. An Introduction to Categorical Data Analysis. Wiley Series in Probability and Statistics. 290pp.

Bates, D. 2007. lme4: Linear mixed-effects models using S4 classes. R package version 0.99875-9.

Beverly, S. and Chapman, L. 2007. Interactions between sea turtles and pelagic longline fisheries. Document WCPFC-SC3-EB SWG/IP-01 of the Scientific Committee Meeting of the Western and Central Pacific Fisheries Commission, 13-24 August, 2007, Hawaii, USA.

Boggs, C.H. and Swimmer, Y. 2007. Developments (2006-2007) in scientific research on the use of modified fishing gear to reduce longline bycatch of sea turtles. Document WCPFC-SC3-EB SWG/WP-7 of the Scientific Committee Meeting of the Western and Central Pacific Fisheries Commission, 13-24 August, 2007, Hawaii, USA.

Bolton, A.B. and Bjorndal, K.A. 2005. Experiment to evaluate gear modification on rates of sea turtle bycatch in the swordfish longline fishery in the Azores – Phase 4. Final Project Report, February, 2005, submitted to the U.S. National Marine Fisheries Service. Archie Carr Center for Sea Turtle Research, University of Florida, Gainesville, Florida, USA.

Broström, G. 2007. glmmML: Generalized linear models with clustering. R package version 0.71-4.

Clark, J.S. (2005) Why environmental scientists are becoming Bayesians. *Ecology Letters* 8: 2-14.

Congdon, P. 2002. Bayesian Statistical Modelling. John Wiley and Sons.

Demidenko, E. 2004. *Mixed Models, Theory and Applications*. Wiley Series in Probability and Statistics. 704pp.

Gelman, A., Carlin, D.B., Stern, H.S. and Rubin, D.B. 2004. Bayesian Data Analysis, Second Edition. Chapman and Hall. 698 pp.

Gilman, E., Zollett, E., Beverly, S., Nakano, H., Davis, K., Shiode, D., Dalzell, P. and Kinan, I. 2006. Reducing sea turtle by-catch in pelagic longline fisheries. Fish and Fisheries 7:2-23.

Hall, D.B. and Zhang, Z. 2004. Marginal models for zero inflated clustered data. Statistical Modelling 4:161-180.

Hall, M.A., Vogel, N., and Orozco, M. 2006. Eastern Pacific Regional Sea Turtle Program, Final Report, June, 2006.

Kamezaki, N., Matsuzawa, Y., Abe, O., Asakawa, H., Fujii, t., Goto, K., Hagino, S., Hayami, M., Ishii, M., Iwamoto, T., Kamata, T., Kato, H., Kodama, J., Kondo, Y., Miawaki, I., Mizobuchi, K., Makamura, Y., Nakashima, Y., Naruse, H., Omuta, K., Samejima, M., Saganuma, H., Takeshita, H., Tanaka, T., Toji, T., Uematsu, M., Yamamoto, Q., Yamato, T., and Wakabayashi, I. 2003. Loggerhead turtles nesting in Japan. p. 210-217. In: Bolten, A.B., and Witherington, B.E., (eds.). Loggerhead Sea Turtles. Smithsonian Institution Press, Washington D.C. 352 pp.

Kerstetter, D.W. and Graves, J.E. 2006. Effects of circle versus J-style hooks on target and non-target species in a pelagic longline fishery. Fisheries Research 80:239-250.

Kim, S.-S., A., D.-H., Moon, D.Y., and Hwang, S.-J. 2007. Comparison of circle hook and J hook catch rate for target and bycatch species taken in the Korean tuna longline fishery in 2006. Document SAR-8-12g of the Working Group on Stock Assessment of the Inter-American Tropical Tuna Commission, May 7-11, 2007, La Jolla, California, USA.

Largacha, E., Parrales, M., Rendón, L., Velásquez, V., Orozco, M., and Hall, M. 2005 Working with the Ecuadorian fishing community to reduce the mortality of sea turtles in longlines: The first year, March 2004-March 2005.

Limpus, C.J. and Limpus, D.J. 2003. Loggerhead turtles in the Equatorial and Southern Pacific Ocean: A species in decline. p. 199-209. In: Bolten, A.B., and Witherington, B.E., (eds.). Loggerhead Sea Turtles. Smithsonian Institution Press, Washington D.C. 352 pp.

McCracken, M.L. (in review). Analysis of longline catch data from bycatch mitigation studies: A comparative study.

Manly, B.F.J. 2007. Randomization, Bootstrap and Monte Carlo Methods in Biology, 3rd Edition. Chapman and Hall/CRC. 455pp.

Patterson, S., and Jones, B. (2006), Bioequivalence and Statistics in Clinical Pharmacology, Chapman & Hall/CRC.

Pinhero, J.C. and Bates, D.M. 2004. *Mixed-effects Models in S and S-PLUS*. Springer. 528pp.

Polovina, J.J., Balazs, G.H., Howell, E.A., Parker, D.M., Seki, M.P., and Dutton, P.H. 2004. Forage and migration habitat of loggerhead (*Caretta caretta*) and olive ridley (*Lepidochelys olivadea*) sea turtles in the Central North Pacific Ocean. Fisheries Oceanography 13:36-51.

R Development Core Team 2007. R: A language and environment for statistical computing. R

Foundation for Statistical Computing, Vienna, Austria. ISBN 3-900051-07-0, URL http://www.R-project.org

Sarti, L., Eckert, S.A., García, N.T. and Barragán, A.R. 1996. Decline of the world's largest nesting assemblage of leatherback turtles. Marine Turtle Newsletter. Number 74. July, 1996.

Sarti, L., Eckert, S.A., Dutton, P.H., Barragán, A.R. and García, N.T. 2000. The current situation of the leatherback population on the Pacific coast of Mexico and Central America, abundance and distribution of the nestings: an update. p. 85-87. In: Kalb, H.J., and Wibbels, T. (compilers), Proc. Nineteenth Annual Symposium on Sea Turtle Biology and Conservation. U.S. Dept. Commerce NOAA Tech. Memo. NMFS-SEFS-443. 291 pp.

Schlesselman, J.J. (1982), *Case-Control Studies. Design, Conduct, Analysis*. Oxford University Press, New York.

Sea Turtle and Pelagic Fish Sensory Physiology Workshop. 2006. Swimmer, Y. and Wang, J., summary compilers. September 12-13, 2006, Virginia Institute of Marine Sciences, Eastern Shore Laboratory, Wachapregue, Virginia.

Spiegelhalter, D.J., Best, N.G., Carlin, B.P and Van der Linde, A. (2002) Bayesian measures of model complexity and fit. *Journal of the Royal Statistical Society, Series B:* 64: 583-639.

Steyermark, A.C., Williams, K., Spotila, J.R., Paladino, F.V., Rostal, D.C., Morreale, S.J., Koberg, M.T., and Arauz, R. 1996. Nesting leatherback turtles at La Baulas National Park, Costa Rica. Chelonian Conserv. Biol. 2(2):173-183.

Stokes, M.E., Davis, C.S. and Koch, G.G. 2001. Categorical Data Analysis Using the SAS System, Second edition. Wiley InterScience. 648 pp.

Watson, J.W., Epperly, S.P., Shah, A.K., and Foster, D.G. 2005. Fishing methods to reduce sea turtle mortality associated with pelagic longlines. Canadian Journal of Fisheries and Aquatic Sciences 62:965-981.

Wellek, S. (2002), Testing Statistical Hypotheses of Equivalence, Chapman & Hall/CRC.

Yokota, K., Kiyota, M., Minami, H. 2006. Shark catch in a pelagic longline fishery: Comparison of circle and tuna hooks. Fisheries Research 81:337-341.
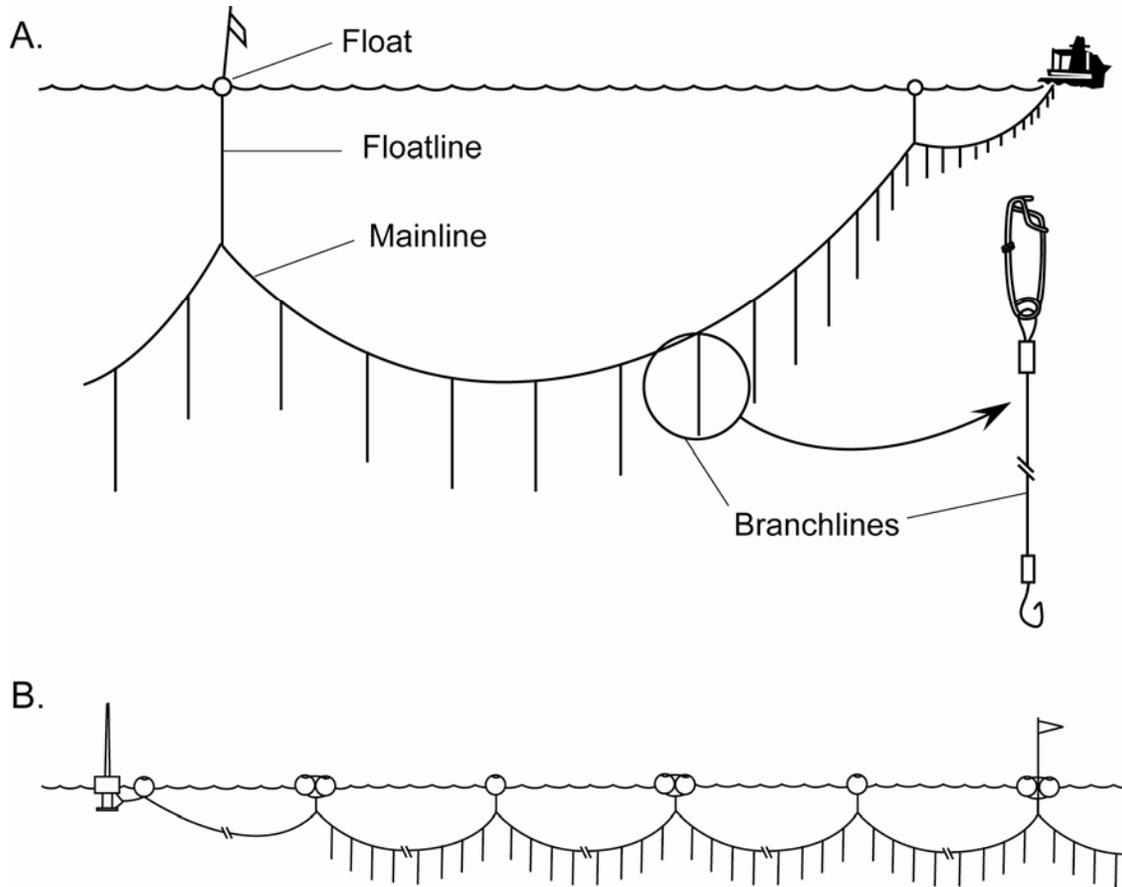
FIGURE 1. Example of a generic longline (from Beverly and Chapman (2007), with permission): a) shows one section of a longline in detail; b) shows an example of multiple sections.
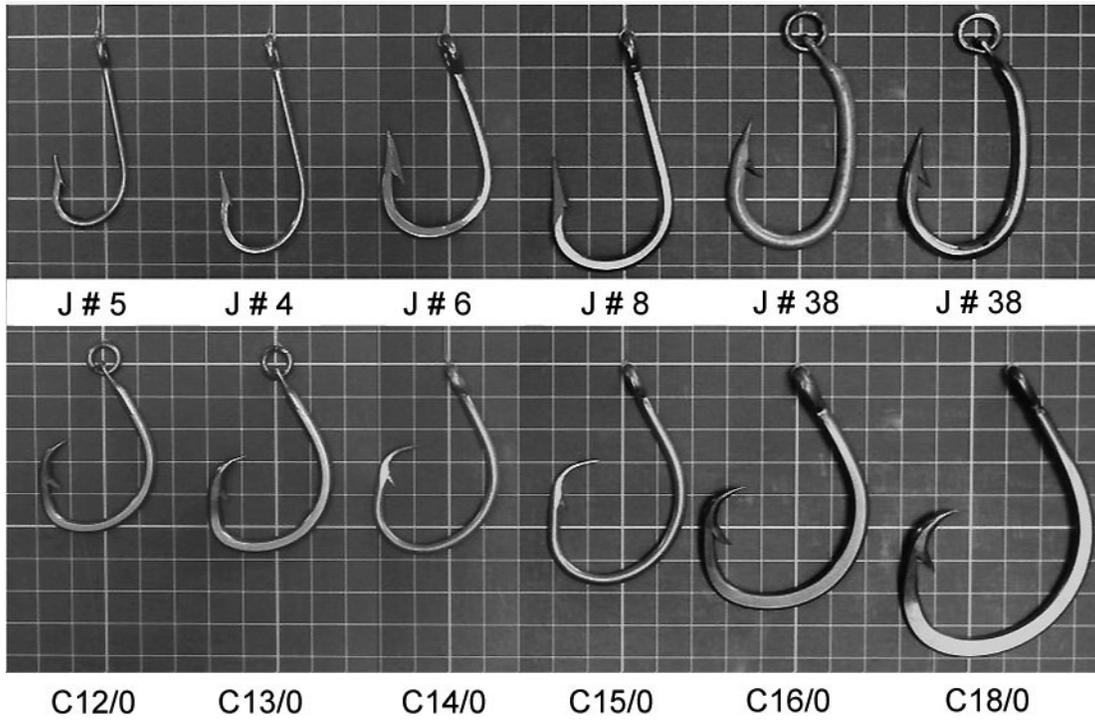
FIGURE 2. Examples of different sizes of J hooks (top row; the J#38 hooks shown in the top row are Japanese-style tuna hooks) and circle hooks (bottom row).
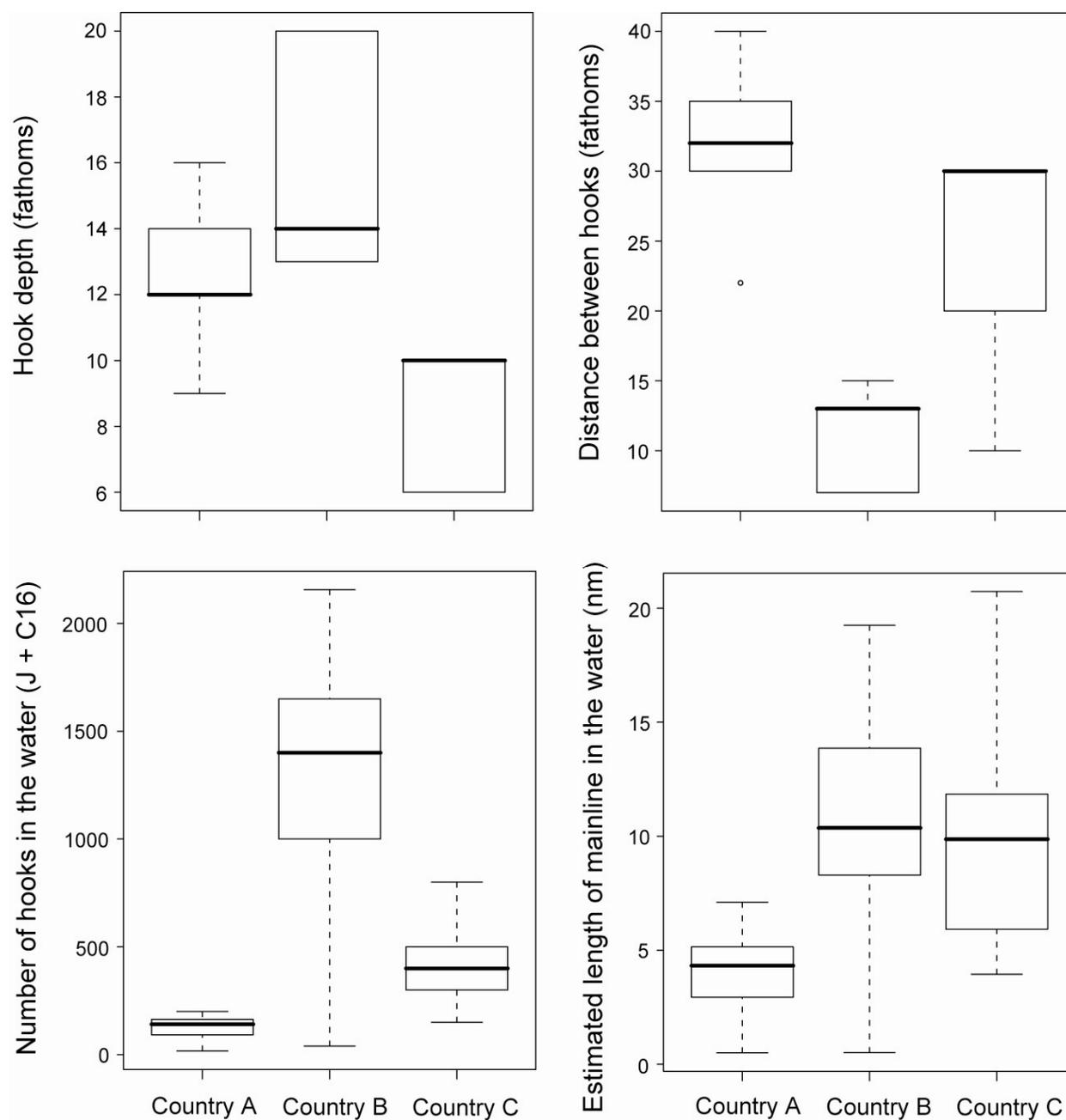
FIGURE 3. Boxplots of various longline gear characteristics, by country. The 'hook depth' is the sum of the floatline and the branchline lengths (Figure 1), and represents the nominal fishing depth of the hook, *ignoring* catenary effects.
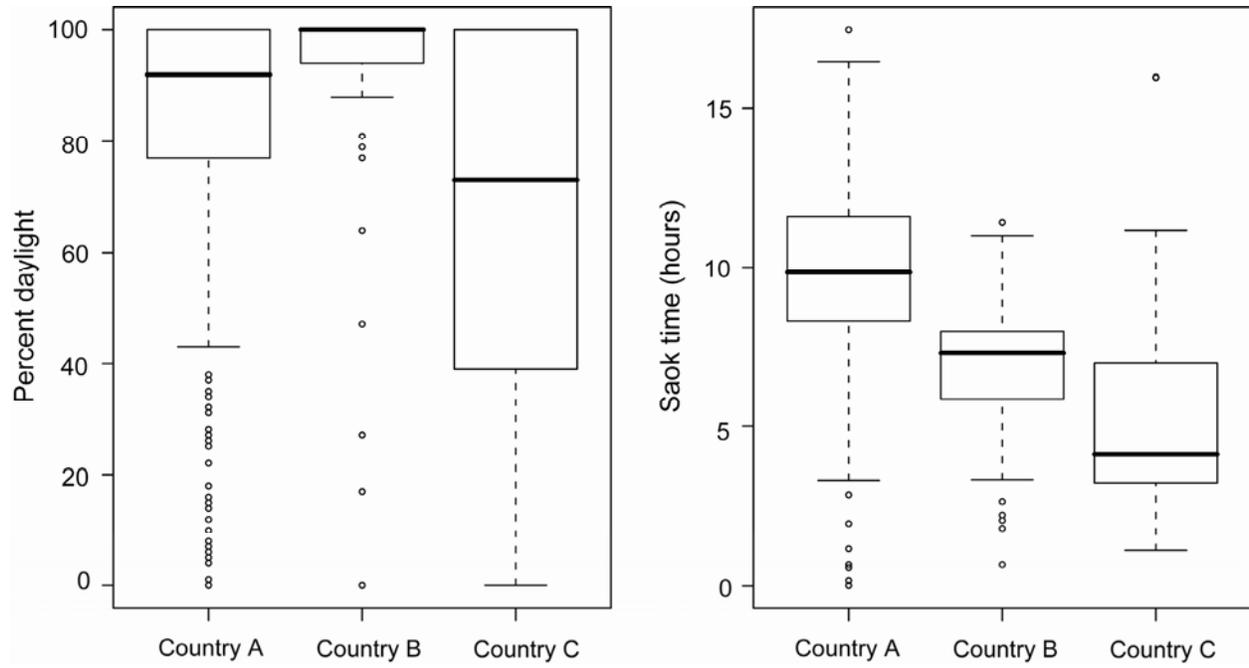
**FIGURE 4.** Operational characteristics, by country.
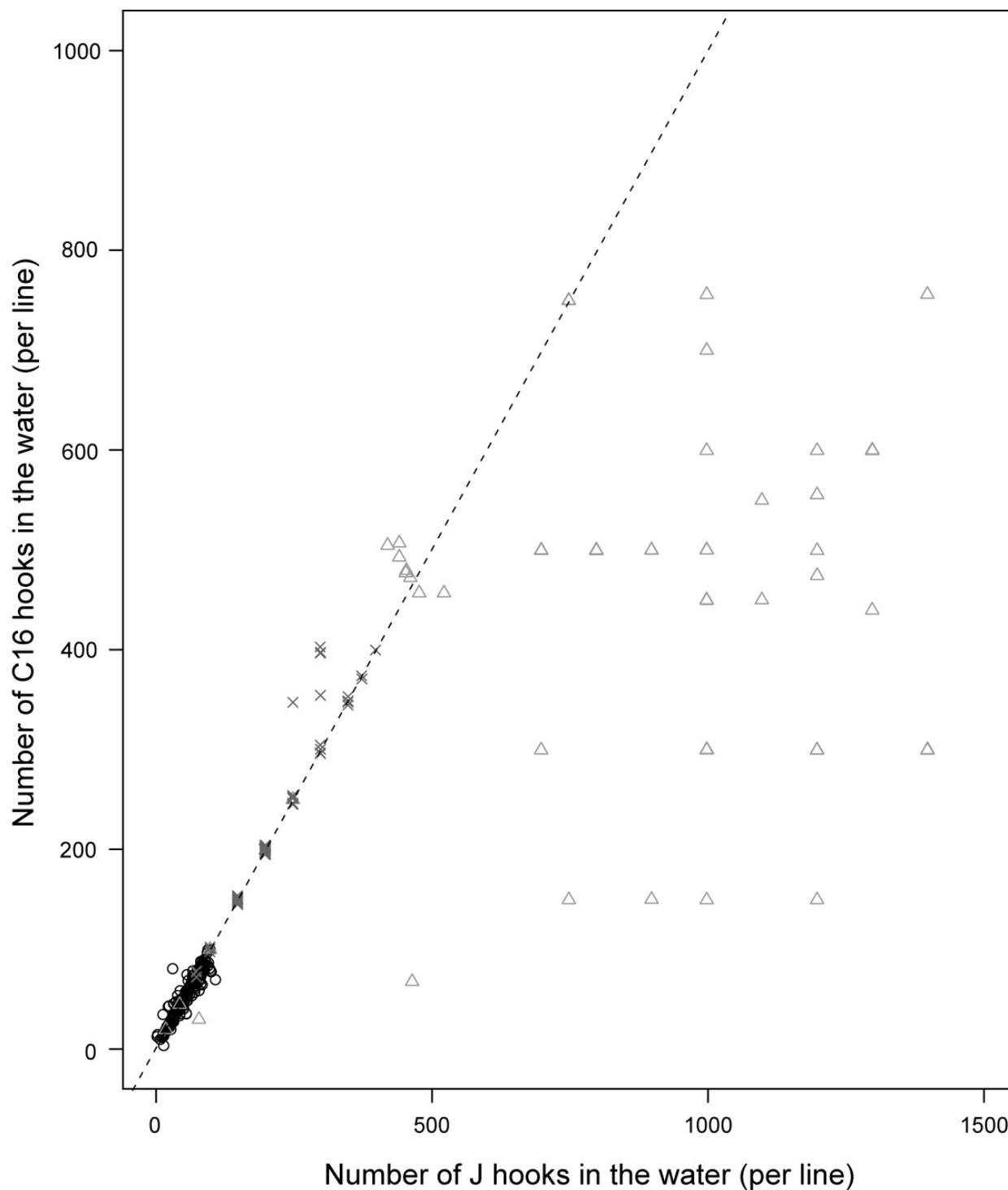
FIGURE 5. Number of J and C16 hooks in the water (per line), by country (black circles: Country A; light gray triangles: Country B; dark gray crosses: Country C) (points jittered to show duplicity of samples at various numbers of hooks).
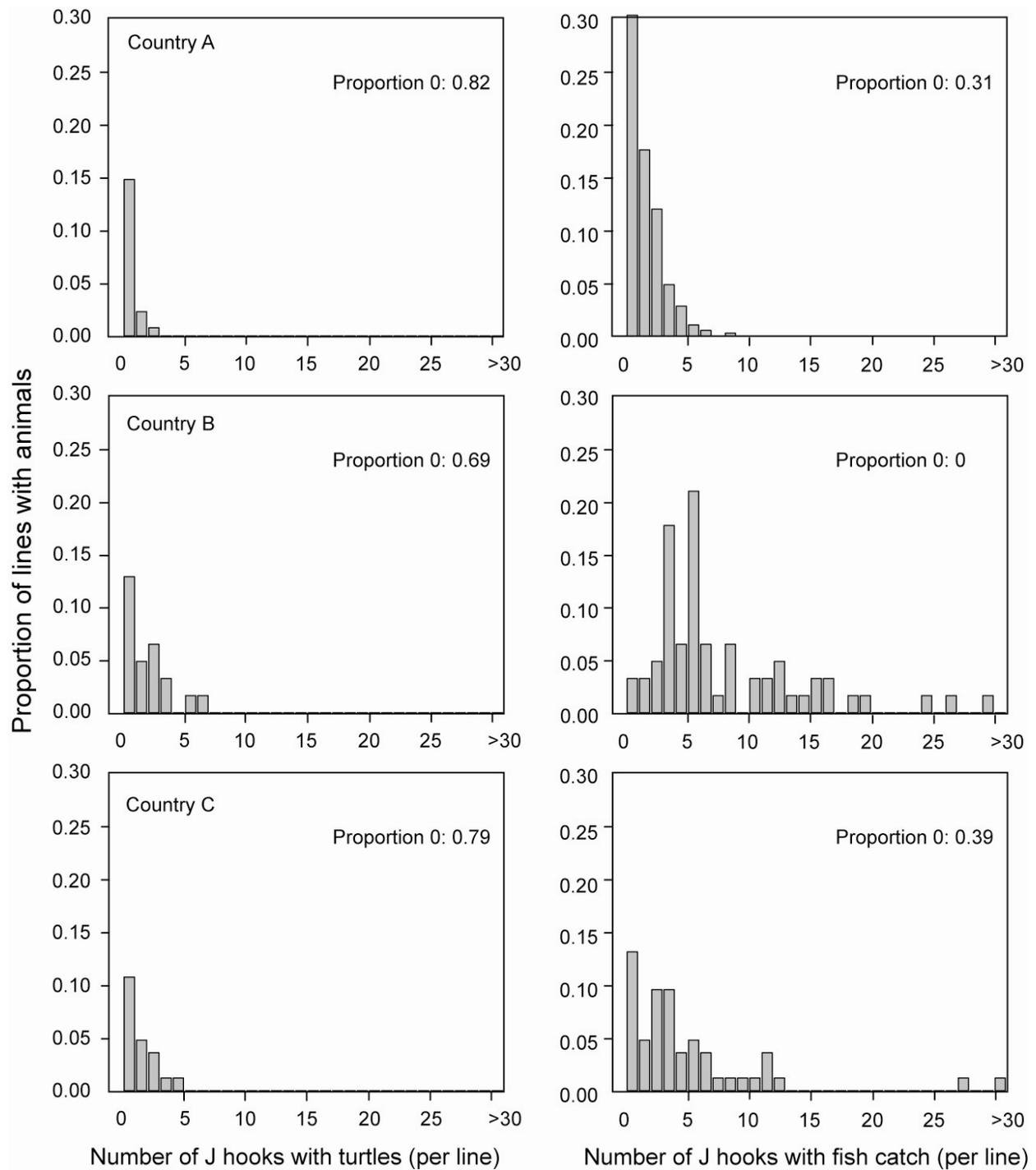
FIGURE 6. Frequency distributions of numbers of turtles and numbers of marketable fishes on J hooks per longline, by country. The proportion of longlines with no turtles (fish) on J hooks is shown for each country.
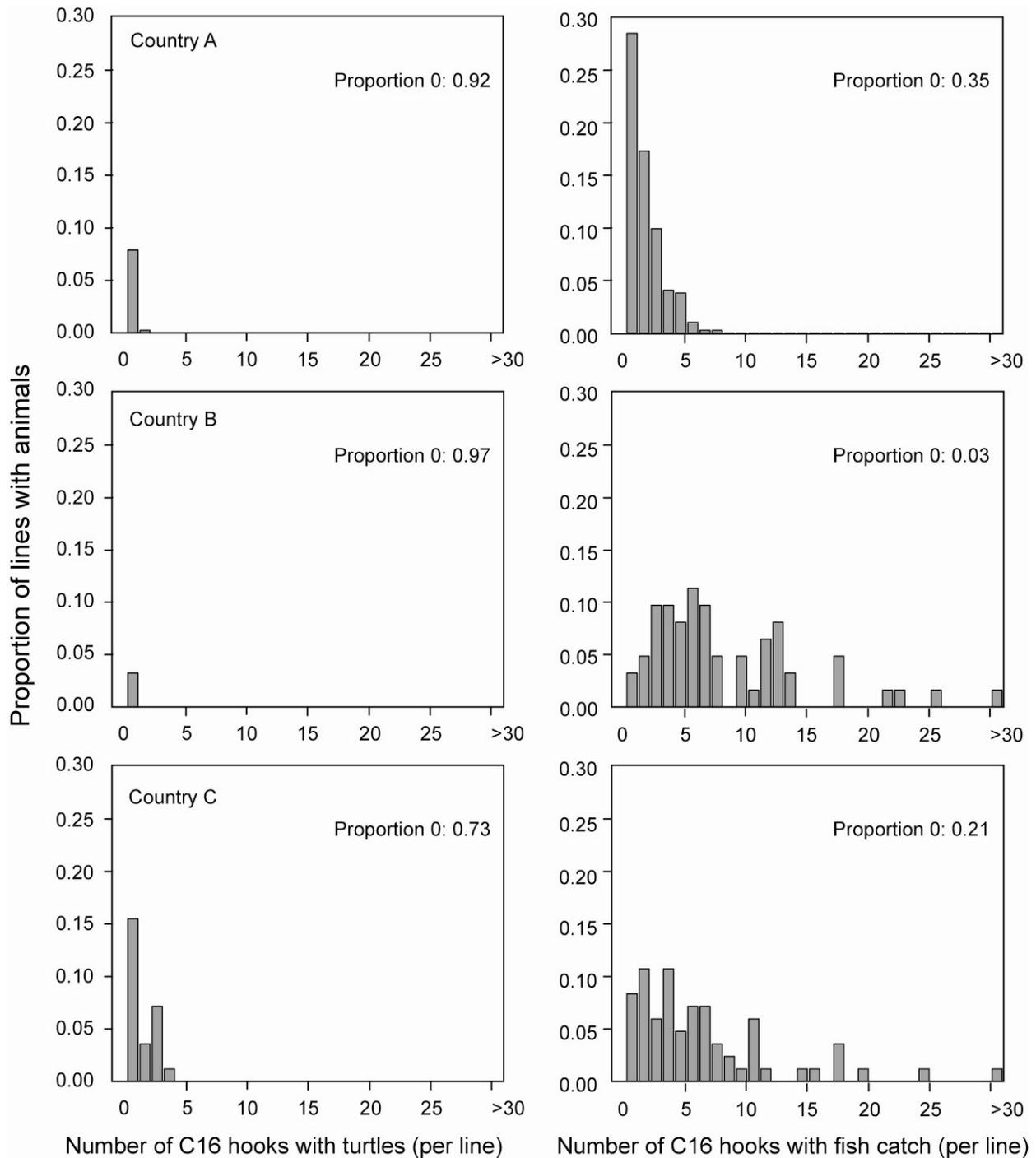
FIGURE 7. Frequency distributions of numbers of turtles and numbers of marketable fishes on C16 hooks per longline, by country. The proportion of longlines with no turtles (fish) on C16 hooks is show for each country.
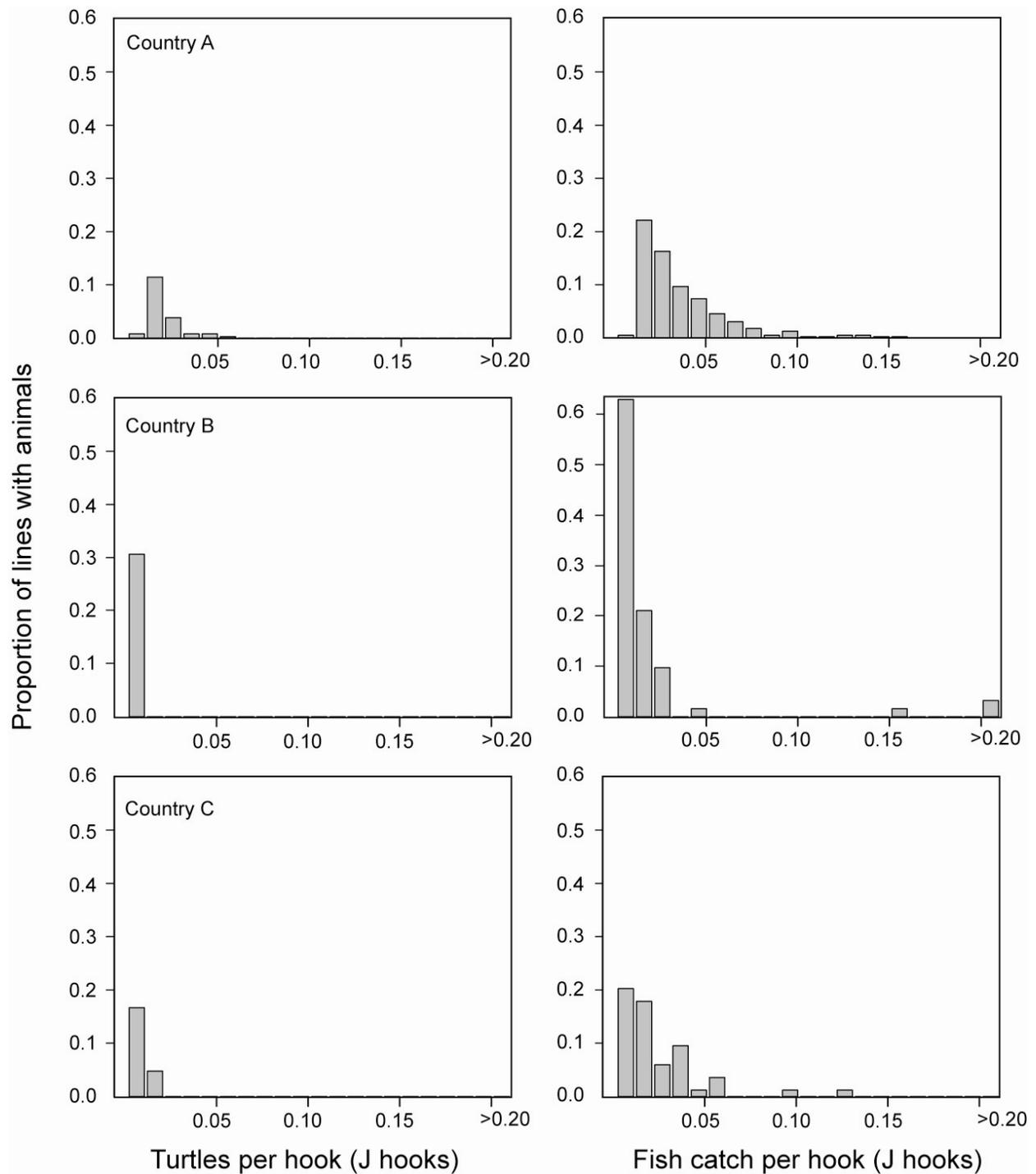
FIGURE 8. Frequency distribution of numbers of animal per J hook, for lines that had animals on J hooks (turtles: left-hand side; marketable fishes: right-hand side), by country. That is, shown are the numbers of animals per line of Figure 6, divided by the number of J hooks per line.
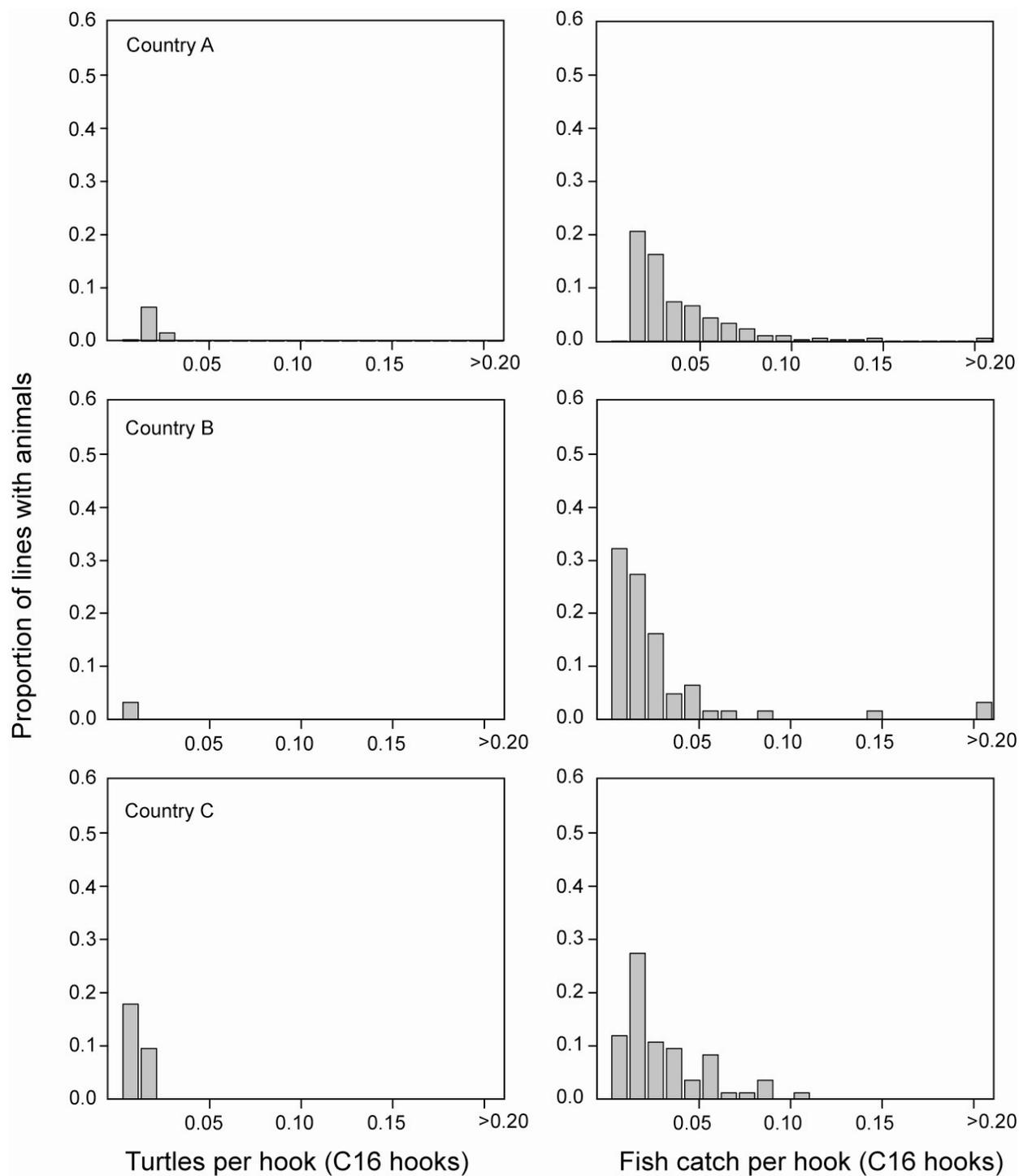
FIGURE 9. Frequency distribution of numbers of animal per C16 hook, for lines that had animals on C16 hooks (turtles: left-hand side; marketable fishes: right-hand side), by country. That is, shown are the numbers of animals per line of Figure 7, divided by the number of C16 hooks per line.
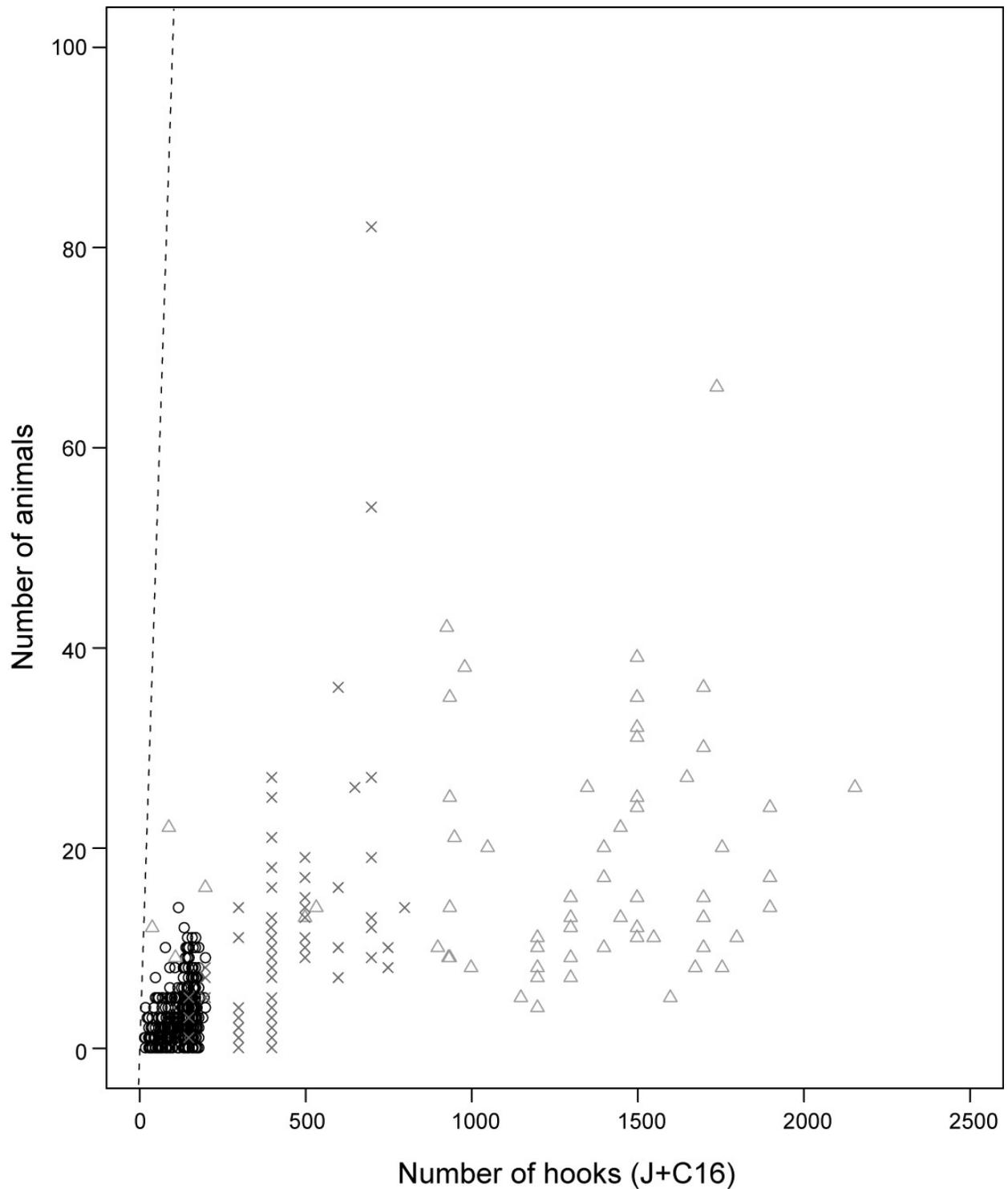
FIGURE 10. Number of hooks on the longline (J + C16) *versus* number of animals hooked (turtles + marketable fishes + other fishes), by country (black circles: Country A; light gray triangles: Country B; dark gray crosses: Country C). The dashed red line is the one-to-one line.
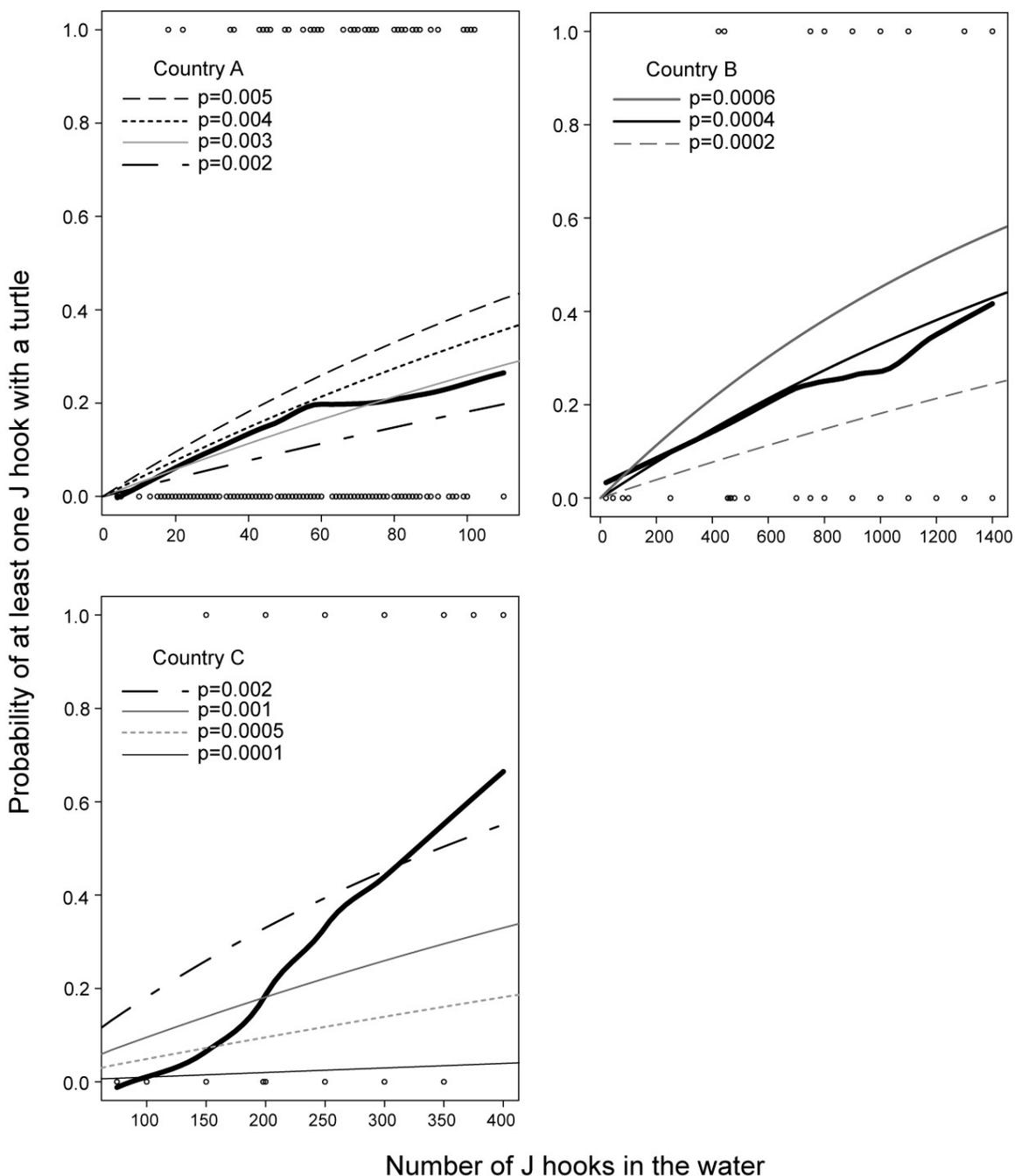
FIGURE 11. Locally weighted regression smooths of turtle presence/absence of data on J hooks (thick black line; span = 0.9, degree = 1), by country. Circles indicate the presence/absence points (one point per line; presence = one or more J hooks with a turtle, absence = no J hooks with a turtle). The lines, other than thick black, are curves showing the probability of at least one J hook per line with a turtle, as a function of the number of J hooks on the line. Curves were generated from hypothetical probabilities (p) that an individual J hook caught a turtle using the following binomial formula for n J hooks per line: probability (at least one J hook per line with a turtle) = $1-[(1-p)^n]$.
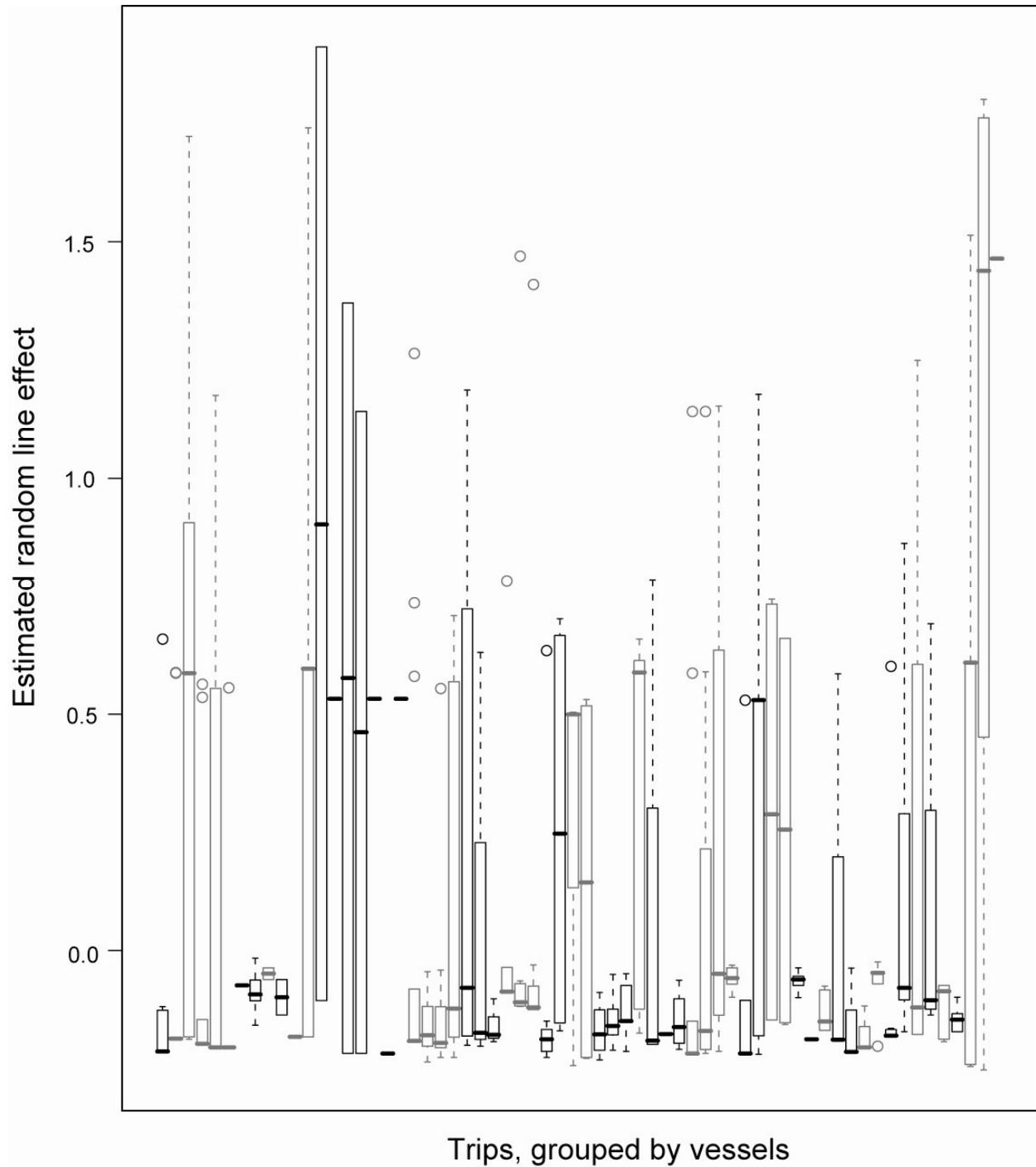
FIGURE 12. Boxplot of the estimated line effects (random effects) per trip for the logistic mixed-effects model fitted to turtle data of Country A. Each individual box/whisker represents one trip. Trips are grouped by vessel according to color, with each group of black or gray boxes representing a different vessel.
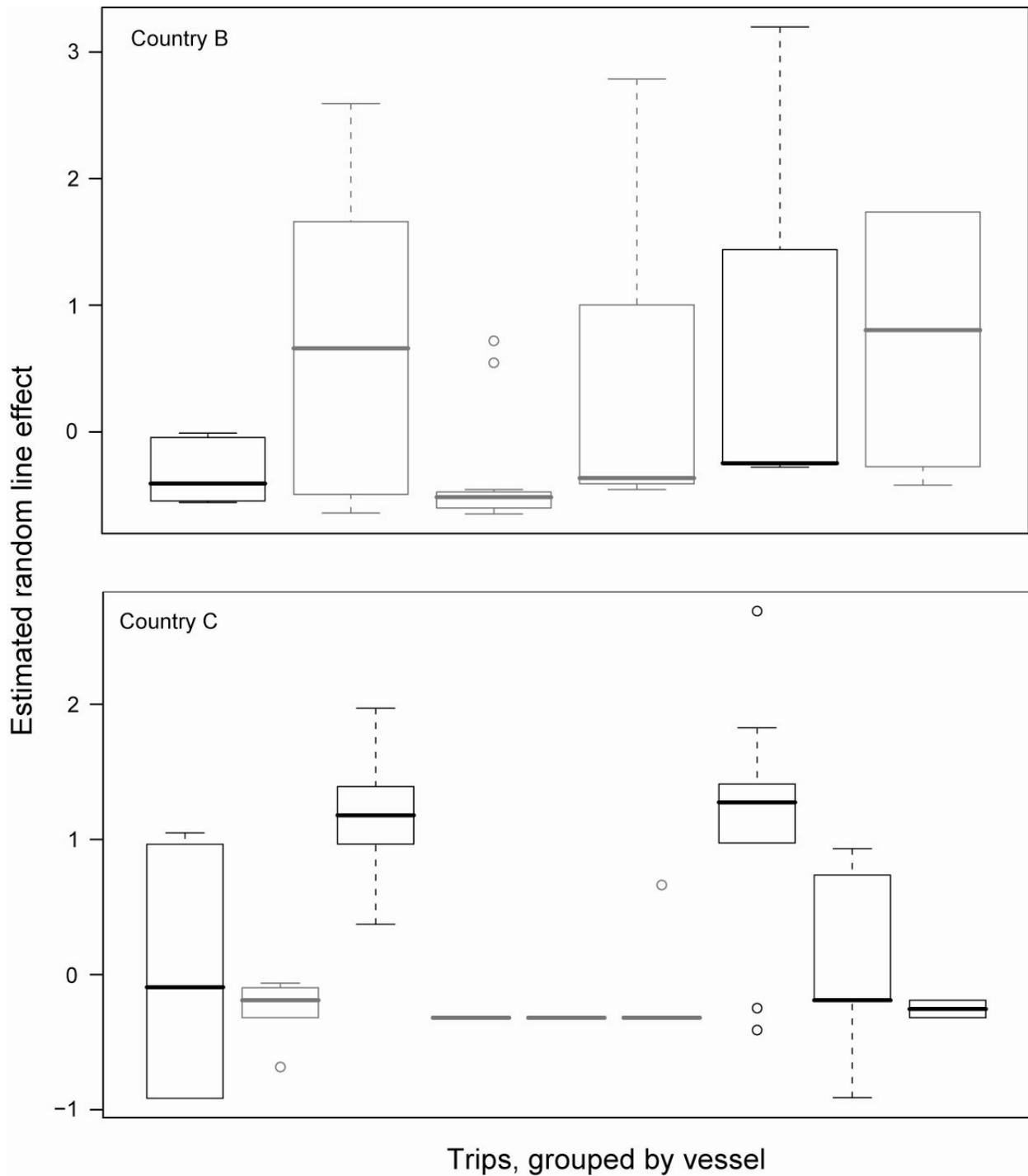
FIGURE 13. Boxplot of the estimated line effects (random effects) per trip for the logistic mixed-effects model fitted to turtle data of Countries B-C. Each individual box/whisker represents one trip. Trips are grouped by vessel according to color, with each group of black or gray boxes representing a different vessel.
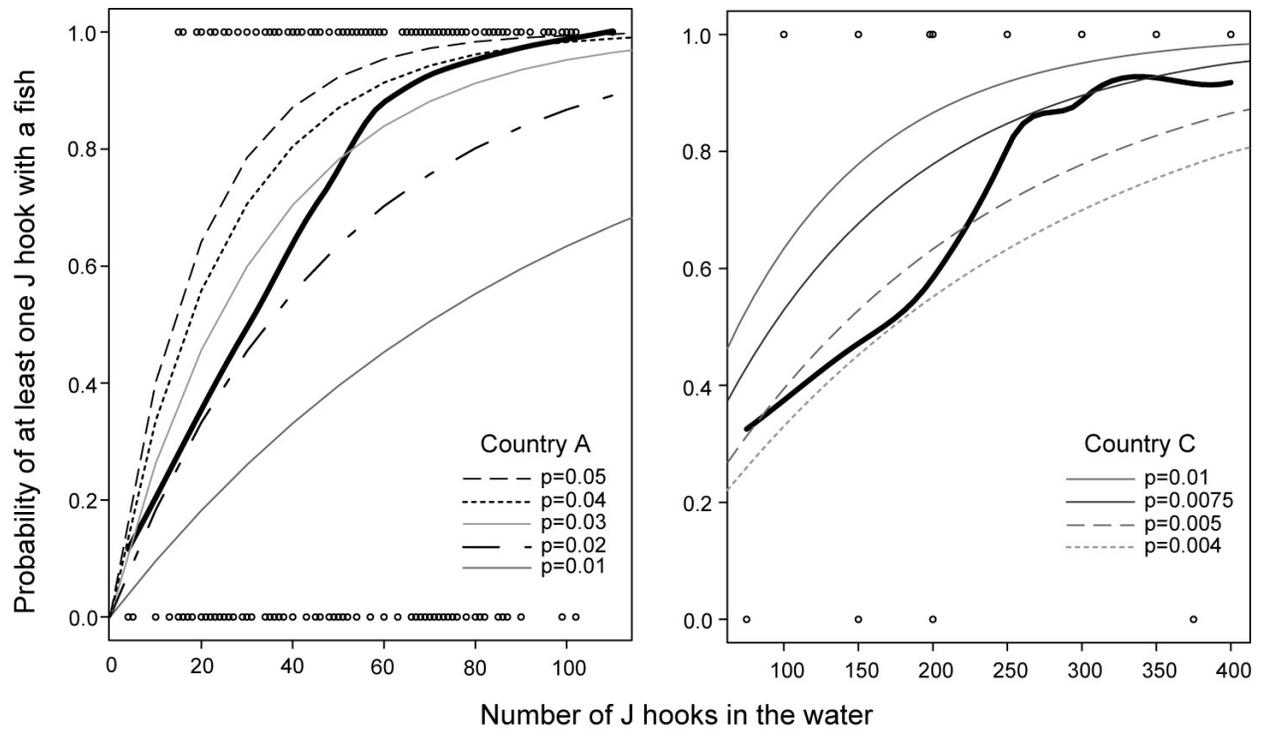
FIGURE 14. Locally weighted regression smooths of marketable fish presence/absence of data on J hooks (thick black line; span = 0.9, degree = 1), by country (all longlines sampled for Country B has at least one marketable fish). Circles indicate the presence/absence points (one point per line; presence = one or more J hooks with a fish, absence = no J hooks with a fish). The lines, other than thick black, are curves showing the probability of at least one J hook per line with a fish, as a function of the number of J hooks on the line. Curves were generated from hypothetical probabilities (p) that an individual J hook caught a fish using the following binomial formula for n J hooks per line: probability (at least one J hook per line with a fish) = $1-[(1-p)^n]$.

TABLE 1. a) Numbers of sampled hooks, sets, trips and vessels, and numbers of hooked turtles and marketable fishes, by country; b) frequency of sets per trip and trips per vessel, by country. Marketable fishes include species of tunas, billfishes and sharks.

(a)

	Number of hooks	Number of sets	Number of trips	Number of vessels	Number of turtles	Number of marketable fishes
Country A	49,362	393	64	28	118	1,119
Country B	80,441	62	6	4	49	1,056
Country C	35,798	84	9	5	76	755

(b)

	Number of sets per trip															
	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16
Country																
A	4	8	5	9	7	3	6	4	6	3	2	5	1			1
B				1					1		2		1	1		
C						3				2	2	2				
	Number of trips per vessel															
A	9	10	5	2	1		1									
B	3		1													
C	3		2													

TABLE 2. Number of longline sets per month, by country.

	April	May	June	July	Aug.	Sept.	Oct.	Nov.
Country A	51	66	92	52	60	45	26	1
Country B				11	18	17	12	4
Country C		1	23	36	2		6	16

TABLE 3. Frequency of bait type (per line) for the first two of three bait types, by country. (Observers could record up to four bait types per longline set.)

(a) Country A

		Second bait type						
		None	Squid	Tuna	Shark	Fish	Sardine	Other/unknown
First bait type	Squid	301		63		14		
	Tuna	6						
	Shark							
	Fish	9						
	Sardine							
	Other/unknown							

(b) Country B

		Second bait type						
		None	Squid	Tuna	Shark	Fish	Sardine	Other/unknown
First bait type	Squid							
	Tuna							
	Shark							
	Fish							
	Sardine	30		4				22
	Other/unknown							6

(c) Country C

		Second bait type						
		None	Squid	Tuna	Shark	Fish	Sardine	Other/unknown
First bait type	Squid	15		2	3		18	1
	Tuna	8				11		
	Shark							
	Fish							
	Sardine	10	4	12				
	Other/unknown							

TABLE 4. Results of mixed-effects logistic model for turtle hookings. The model fitted separately to the data of each country was: logit(probability of turtle on a hook) = constant + hook type effect + β·total hooks in the water + line effect. The hook type effect was a fixed effect (0 = J, 1 = C16), the line effect was a random effect. The models were fitted with the *lmer* function in the *lme4* package (Bates, 2007) of the statistical software *R* (R Development Core Team, 2007), using the Laplace approximation. Also shown (in parentheses) are estimates of the scale parameter and the variance of the line effects, obtained from *lmer* using penalized quasi-likelihood (PQL). Relatively similar estimates of fixed effects were obtained with both methods. Attempts to fit the mixed-effects models with *glmmML* (Broström, 2007) (also of *R*) using Gaussian quadrature failed, possibly because of difficulties estimating the variance of the line effect (simulated data with small values for the variance of the line effect produced similar *glmmML* error messages).

COUNTRY A

AIC	BIC	logLik	Deviance
445.9	464.5	-218.9	437.9

Random effects:

	Variance	
Line	0.9932	(PQL variance estimate: 17.665)
Estimated scale (compare to 1):	0.7744889	(PQL scale estimate: 0.4387959)

Fixed effects:

	Estimate	Std. error	z value	Pr(>|z|)
Intercept	-6.350501	0.509104	-12.474	< 2e-16
Hook type	-0.911403	0.232435	-3.921	8.81e-05
Total hooks	0.001459	0.003459	0.422	0.673

COUNTRY B

AIC	BIC	logLik	Deviance
111.1	122.4	-51.57	103.1

Random effects:

	Variance	
Line	2.6875	(PQL variance estimate: 3.5117)
Estimated scale (compare to 1):	0.8413038	(PQL scale estimate: 0.7733437)

Fixed effects:

	Estimate	Std. error	z value	Pr(>|z|)
Intercept	-8.6803510	1.4434688	-6.014	1.82e-09
Hook type	-2.5743245	0.7710514	-3.339	0.000842
Total hooks	0.0003701	0.0009737	0.380	0.703856

COUNTRY C

AIC	BIC	logLik	Deviance
168.3	180.8	-80.14	160.3

Random effects:

	Variance	
Line	1.5134	(PQL variance estimate: 6.7325)
Estimated scale (compare to 1):	0.8473257	(PQL scale estimate: 0.6037494)

Fixed effects:

	Estimate	Std. error	z value	Pr(>|z|)
Intercept	-8.737049	0.774960	-11.274	<2e-16
Hook type	0.131364	0.242826	0.541	0.5885
Total hooks	0.003601	0.001464	2.460	0.0139

TABLE 5. Results of mixed-effects logistic model for marketable fish hookings. The model fitted was: logit(probability of fish on a hook) = constant + hook type effect + β·total hooks in the water + line effect. The hook type effect was a fixed effect (0 = J, 1 = C16), the line effect was a random effect. The models were fitted with *lmer* in the *lme4* package (Bates, 2007) of the statistical software *R* (R Development Core Team, 2007) , using the Laplace approximation. Also shown (in parentheses) are estimates of the scale parameter obtained from *lmer* using penalized quasi-likelihood (PQL), and the estimates of the variance of the line effects based on *lmer* using PQL and *glmmML* (Broström, 2007) (also of *R*) using Gaussian quadrature (GQH). Relatively similar estimates of fixed effects were obtained with all three methods.

COUNTRY A

AIC	BIC	logLik	Deviance
26278	26296	-13135	26270

Random effects:

	Variance	
Line	1.3333	(PQL variance estimate: 0.38325; GQH variance estimate: 0.329)
Estimated scale (compare to 1):	3187955	(PQL scale estimate: 0.939615)

Fixed effects:

	Estimate	Std. error	z value	Pr(>|z|)
Intercept	-3.378094	0.223275	-15.130	<2e-16
Hook type	-0.024735	0.067656	-0.366	0.715
Total hooks	-0.002671	0.001646	-1.623	0.105

COUNTRY B

AIC	BIC	logLik	Deviance
5510	5521	-2751	5502

Random effects:

	Variance	
Line	1.3333	(PQL variance estimate: 0.070937; GQH variance estimate: 0.443)
Estimated scale (compare to 1):	740676174	(PQL scale estimate: 1.887632)

Fixed effects:

	Estimate	Std. error	z value	Pr(>|z|)
Intercept	-3.2592516	0.7220521	-4.514	6.37e-06
Hook type	0.6465375	0.1099963	5.878	4.16e-09
Total hooks	-0.0009486	0.0005185	-1.830	0.0673

COUNTRY C

AIC	BIC	logLik	Deviance
7053	7066	-3523	7045

Random effects:

	Variance	
Line	1.3333	(PQL variance estimate: 0.36275; GQH variance estimate: 1.109)
Estimated scale (compare to 1):	239443975	(PQL scale estimate: 1.345933)

Fixed effects:

	Estimate	Std. error	z value	Pr(>|z|)
Intercept	-4.868576	0.533933	-9.118	< 2e-16
Hook type	0.454781	0.145993	3.115	0.00184
Total hooks	0.001558	0.001207	1.290	0.19689

The IATTC's responsibilities are met with two programs, the Tuna-Billfish Program and the Tuna-Dolphin Program. The principal responsibilities of the Tuna-Billfish Program are (1) to study the biology of the tunas and related species of the eastern Pacific Ocean to estimate the effects that fishing and natural factors have on their abundance, (2) to recommend appropriate conservation measures so that the stocks of fish can be maintained at levels that will afford maximum sustainable catches, and (3) to collect information on compliance with Commission resolutions. The principal responsibilities of the Tuna-Dolphin Program are (1) to monitor the abundance of dolphins and their mortality incidental to purse-seine fishing in the eastern Pacific Ocean, (2) to study the causes of mortality of dolphins during fishing operations and promote the use of fishing techniques and equipment that minimize these mortalities, (3) to study the effects of different modes of fishing on the various fish and other animals of the pelagic ecosystem, and (4) to provide a Secretariat for the International Dolphin Conservation Program.

An important part of the work of the IATTC is the prompt publication and wide distribution of its research results. The Commission publishes its results in its Bulletin, Special Report, and Data Report series, all of which are issued on an irregular basis, and its Stock Assessment Reports and Fishery Status Reports, which are published annually.

The Commission also publishes Annual Reports and Quarterly Reports, which include policy actions of the Commission, information on the fishery, and reviews of the year's or quarter's work carried out by the staff. The Annual Reports also contain financial statements and a roster of the IATTC staff.

Additional information on the IATTC's publications can be found in its web site.

La CIAT cumple sus obligaciones mediante dos programas, el Programa Atún-Picudo y el Programa Atún-Delfín. Las responsabilidades principales del primero son (1) estudiar la biología de los atunes y especies afines en el Océano Pacífico oriental a fin de determinar los efectos de la pesca y los factores naturales sobre su abundancia, (2) recomendar medidas apropiadas de conservación para permitir mantener los stocks de peces a niveles que brinden las capturas máximas sostenibles, (3) reunir información sobre el cumplimiento de las resoluciones de la Comisión. Las responsabilidades principales del segundo son (1) dar seguimiento a la abundancia de los delfines y la mortalidad de los mismos incidental a la pesca con red de cerco en el Océano Pacífico oriental, (2) estudiar las causas de la mortalidad de delfines durante las operaciones de pesca y fomentar el uso de técnicas y aparejo de pesca que reduzcan dicha mortalidad al mínimo, (3) estudiar los efectos de distintas mortalidades de pesca sobre los varios peces y otros animales del ecosistema pelágico, (4) proporcionar la Secretaría para el Programa Internacional para la Conservación de los Delfines.

La pronta publicación y amplia distribución de los resultados de investigación forman un aspecto importante de las labores de la Comisión, la cual publica los resultados en su serie de Boletines, Informes Especiales, e Informes de Datos, publicados a intervalos irregulares, y sus Informes de Evaluación de Stocks e Informes de la Situación de la Pesquería, publicados anualmente.

La Comisión publica también Informes Anuales e Informes Trimestrales; éstos incluyen información sobre las labores de la Comisión, la pesquería, y las investigaciones realizadas en el año o trimestre correspondiente. Los Informes Anuales incluyen también un resumen financiero y una lista del personal de la CIAT.

En el sitio de internet de la CIAT se presenta información adicional sobre estas publicaciones.