Abundance of harbor porpoise (*Phocoena phocoena*) in three Alaskan regions, corrected for observer errors due to perception bias and species misidentification, and corrected for animals submerged from view

Roderick C. Hobbs (contact author)
Janice M. Waite

Email address for contact author: Rod.Hobbs@noaa.gov

National Marine Mammal Laboratory
Alaska Fisheries Science Center
National Marine Fisheries Service
7600 Sand Point Way N.E.
Seattle, Washington 98115

Abstract—Estimating the abundance of cetaceans from aerial survey data requires careful attention to survey design and analysis. Once an aerial observer perceives a marine mammal or group of marine mammals, he or she has only a few seconds to identify and enumerate the individuals sighted, as well as to determine the distance to the sighting and record this information. In line-transect survey analyses, it is assumed that the observer has correctly identified and enumerated the group or individual. We describe methods used to test this assumption and how survey data should be adjusted to account for observer errors. Harbor porpoises (*Phocoena phocoena*) were censused during aerial surveys in the summer of 1997 in Southeast Alaska (9844 km survey effort), in the summer oft 1998 in the Gulf of Alaska (10,127 km), and in the summer of 1999 in the Bering Sea (7849 km). Sightings of harbor porpoise during a beluga whale (*Phocoena phocoena*) survey in 1998 (1355 km) provided data on harbor porpoise abundance in Cook Inlet for the Gulf of Alaska stock. Sightings by primary observers at side windows were compared to an independent observer at a belly window to estimate the probability of misidentification, underestimation of group size, and the probability that porpoise on the surface at the trackline were missed (perception bias, $g(0)$). There were 129, 96, and 201 sightings of harbor porpoises in the three stock areas, respectively. Both $g(0)$ and effective strip width (the realized width of the survey track) depended on survey year, and $g(0)$ also depended on the visibility reported by observers. Harbor porpoise abundance in 1997–99 was estimated at 11,146 animals for the Southeast Alaska stock, 31,046 animals for the Gulf of Alaska stock, and 48,515 animals for the Bering Sea stock.

Manuscript submitted 2 March 2009.
Manuscript accepted 4 March 2010.
Fish. Bull. 108:251–267 (2010).

The views and opinions expressed or implied in this article are those of the author (or authors) and do not necessarily reflect the position of the National Marine Fisheries Service, NOAA.

Accurate estimation of abundance of cetaceans from survey data requires careful attention to both survey design and analysis (Buckland et al., 2001). Aerial surveys of cetaceans depend on rapid discovery, recognition, and recording of sightings of individuals and groups of animals by observers, in addition to accounting for animals that were missed because the observers did not notice them (perception bias) or because the animals were below the surface (availability bias) (Buckland et al., 2001). Although an experienced trained observer is efficient at recognition of a species and recording data, it is necessary to include methods that can measure error rates of observers and account for them in the estimation of abundance. We present here the results of a series of aerial surveys designed to estimate the abundance of harbor porpoise (*Phocoena phocoena*) in Alaskan waters.

When, during an aerial line-transect survey, an object or group of objects is encountered, an aerial observer has only a few seconds to complete several tasks: 1) perceive the objects, 2) identify the objects, 3) enumerate the objects, and 4) determine the distance of the objects from the trackline. Items 1 and 4 are the major concern for the estimation of perception bias and for line-transect survey analysis, and it is generally assumed that the observer completes items 2 and 3 correctly or indicates uncertainty correctly (e.g., species code "unidentified porpoise" indicates uncertainty between Dall's porpoise [*Phocoenoides dalli*] and harbor porpoise). We develop methods to test the assumptions of correct species identification and enumeration and apply them to the analysis of line-transect survey data and the estimation of abundance.

From 1991 to 1993, the National Oceanic and Atmospheric Administration (NOAA) conducted aerial surveys in three regions of the Alaskan coast: 1) Cook Inlet and Bristol Bay in 1991; 2) in the waters around Kodiak Island and south of the Alaska Peninsula in 1992; and 3) in the offshore waters of Southeast Alaska from Dixon Entrance to Prince William Sound in 1993. The inside waters of Southeast Alaska were surveyed in each of these years by NOAA crews aboard the NOAA RV *John N. Cobb*. The abundance estimates for these regions were combined to produce an abundance estimate for the Alaska stock of harbor porpoise (Dahlheim et al., 2000). Since then, the Alaska stock has been split into three stocks: Southeast Alaska (SEA), Gulf of Alaska (GOA), and the Bering Sea (BS) stocks (Fig. 1). The 1991–93 abundance estimate was subdivided to correspond with the new stock boundaries (Hill and DeMaster, 1998). To maintain up-to-date stock assessments, abundance

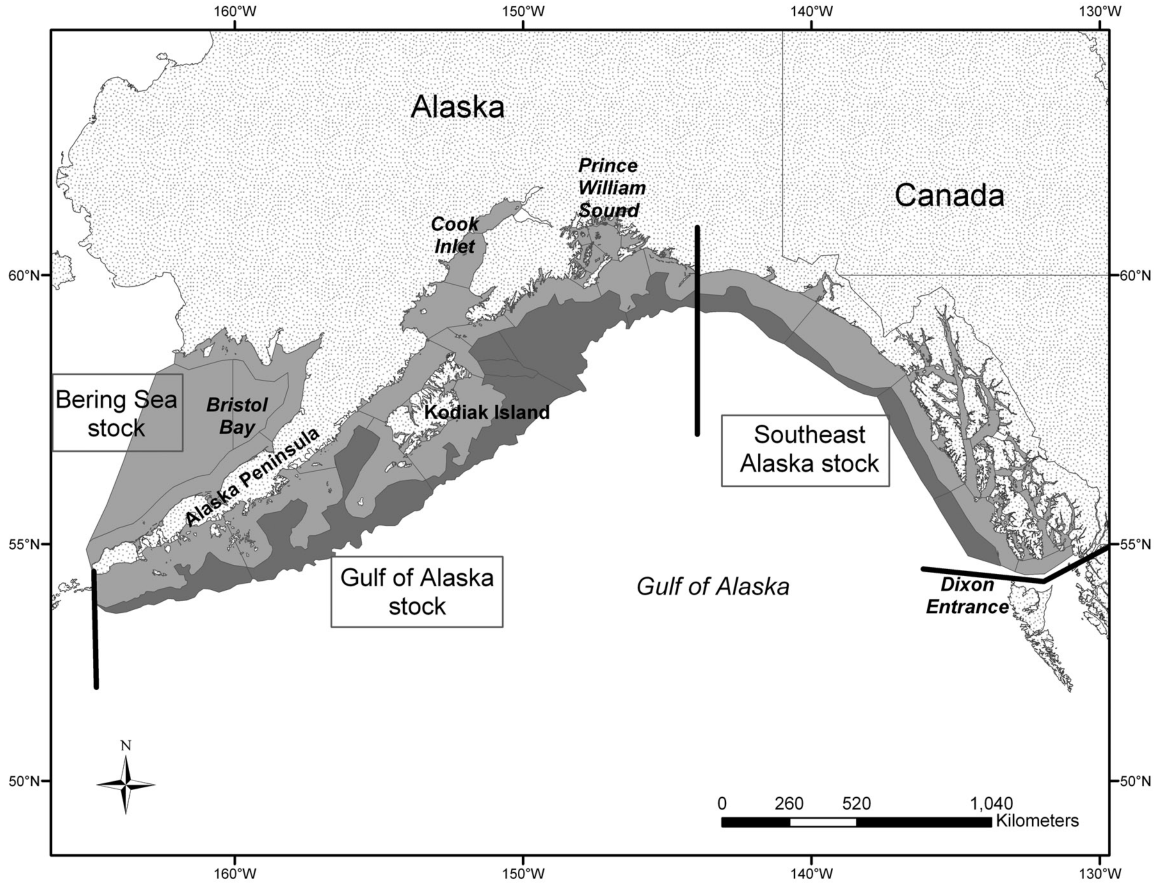

Figure 1

The three harbor porpoise (*Phocoena phocoena*) stock regions in Alaska (Southeast Alaska, Gulf of Alaska, and Bering Sea). The gray shaded areas represent the areas surveyed in 1997– 99, subdivided into areas based on geographical features and depth zones. Dark gray offshore areas were surveyed at one third the effort level per square km than lighter gray nearshore areas. Black lines represent boundaries between stocks.

estimates are required to be based on data not more than 8 years old (Wade and Angliss, 1997). To meet this requirement, abundance surveys of the harbor porpoise stocks in Alaskan waters were conducted from 1997 through 1999.

An important consideration when conducting multiyear surveys is that animals may move from one survey area to another among years and therefore may be counted more than once. Little is known about the year-to-year changes in the distribution of harbor porpoises in Alaska. For two studies of harbor porpoise (*Phocoena phocoena*) on either side of the Atlantic, in the Danish Belt seas (Teilmann et al.[1]) and in the Gulf of Maine (Read and Westgate, 1997), it was concluded that porpoises follow similar movement patterns from year to year and typical ranges of up to 200 km. This finding indicates that a net movement in response to interannual variation in habitat could occur over a range of 100 km. Although this could result in a bias in the estimation of abundance for each stock, depending on the year of the survey, it is not likely that this is a significant occurrence. Each stock region comprises 800 to 1200 km of shoreline so at most approximately 10% to 15% of the region is potentially subject to a net shift in distribution from or into the adjacent stock. Also, the stock boundaries have been chosen to correspond with areas of low harbor porpoise density and therefore there are few animals available to make a shift. These two arguments suggest that if a net shift does occur, it affects at most a small percentage of the population.

This study had three objectives:

1 to present the results of an aerial survey of three harbor porpoise stocks in Alaskan waters during the summers of 1997, 1998, and 1999;

[1] Teilmann, J., R. Dietz, F. Larsen, G. Desportes, and B. Geertsen. 2003. Seasonal migrations and population structure of harbour porpoises (*Phocoena phocoena*) in the North Sea and inner Danish waters based on satellite telemetry. Abstract in proceedings of annual meeting of the European Cetacean Society, Tenerife, Spain.

2 to produce a correction factor for perception bias that was specific to harbor porpoise in Alaskan waters and to the surveys presented here and to develop methods to test observer performance, by using data collected during these surveys; and

3 to generate abundance estimates for the three stocks of harbor porpoise in Alaska during 1997–99.

Materials and methods

Survey design

Aerial surveys were conducted during June and July beginning with the SEA stock in 1997, proceeding westward through the GOA stock in 1998, and on to the BS stock in 1999. Each study region was divided into areas; 70 areas in Southeast Alaska, 39 areas in the Gulf of Alaska, and 4 areas in the Bering Sea (primarily in Bristol Bay) based on geographical features for inside waters, such as straits and inlets, and two depth zones for offshore waters (Fig. 1). Southeast Alaska was divided into more areas because of its complicated system of waterways, whereas Bristol Bay has relatively homogenous features and therefore was divided into fewer survey areas. Survey effort was stratified by area in Southeast Alaska based on harbor porpoise encounter rates calculated from sightings made in previous surveys (Dahlheim et al., 1993, 1994[2]). The survey transect design each year varied depending on the body of water. In general, the transects in offshore waters were stratified by depth and distance from shore, after an alternating two short and one long sawtooth transect pattern, so that survey effort in the nearshore strata was about three times that in the offshore strata. The 1991–93 surveys were designed with fixed distances of 28 km offshore for the short and 74 km offshore for the long sawtooth tracklines. Our surveys were designed to include the area surveyed in 1991–1993 but also to cover the continental shelf if it extended beyond the original survey. Each set of sawtooths had two criteria and the further offshore of the criteria determined the length of the line. Specifically, in 1997, the short transects in the sawtooth transect pattern extended to a distance of 31 km offshore or the 183-m (100 fm) depth contour, and the long transects extended 74 km or to the 1829-m (1000 fm) depth contour, whichever was farthest from shore. In the 1998 GOA survey, the shelf fell much more gradually in places and funding limited the total survey time. Therefore, the nearshore strata transects were reduced to a distance of 28 km or to the 91-m (50 fm) depth contour, whichever was farthest from shore, whereas the long transects followed the same criteria as in 1997. Because the entire Bering Sea survey region is shallower than 183 m (100 fm), it was covered equally with short transects out to 30 km along the shore and with parallel north–south lines through the center approximately 18 km apart. Smaller bays and inlets were treated separately and stratified by the width of the mouth of the bay or inlet. A subset was chosen to approximate the survey effort by area for the other survey regions, and selection was made on the basis of convenience (i.e., bays and inlets close to the end of survey effort lines were chosen).

Line-transect surveys were flown at an altitude of 152.5 m and a speed of 185 km/h in a DeHavilland Twin Otter aircraft. Survey areas were chosen each day to complete coverage of contiguous areas during weather with winds below 15 knots and at a ceiling above 1000 ft (305 m). Survey lines were broken off and other tracklines with better conditions were sought if the Beaufort sea state exceeded 3 or if visibility dropped to poor for a significant period (at the discretion of the team leader). A primary observer (also referred to as a "side observer") was stationed at the left and right bubble windows of the plane; these positions allowed them to see water directly below the plane. To collect additional sightings and data to estimate perception bias for this study, an independent observer was stationed at a belly window located in the floor at the back of the plane (this observer is also referred to as the "belly observer"). This window provided a circular field of view 100 m (30°) to either side of the trackline and 200 m along the trackline. Five observers rotated in 40-minute shifts through five positions: the right and left bubble windows (primary observers), the belly window (independent observer), a computer station, and a rest position. A headset system was used by the primary observers and computer operator to communicate openly, and the independent observer was isolated and used a string attached to the arm or ankle of the computer operator to indicate a sighting and a notepad to relay information. A simple short hand was developed so that the belly observers would not need to take their eyes off of the trackline.

Survey data were recorded directly to a laptop computer in the airplane using a Turbo PASCAL (vers. 5.0, Borland Software Corp., Austin, TX) language-based software customized for the survey. The software included a proprietary routine (Survey, vers. 3.2, Cascadia Research, Olympia, WA) which read the text output of a global positioning system (GPS) unit connected directly to the serial port of the computer. The date, time, and position of the aircraft were automatically entered into the survey data every minute or whenever other data were entered by the recorder. At the start of each transect, waypoint numbers, observer positions, and environmental conditions were recorded. Environmental conditions included percent cloud cover, Beaufort sea state, visibility (a subjective rating of sighting conditions by each observer at the following levels (excellent, good, fair, poor, and unacceptable), and glare (none, minor,

[2] Dahlheim, M., A. York, J. Waite, and R. Towell. 1993. Abundance and distribution of harbor porpoise (*Phocoena phocoena*) in Southeast Alaska and Western Gulf of Alaska, 1992. 1992 Annual report to the Marine Mammal Protection Act (MMPA) Assessment Program, 52 p. Office of Protected Resources. NMFS, NOAA, 1335 East-West Highway, Silver Spring, MD.

bad, or reflective) experienced by each observer. Visibility was defined as the observer's subjective assessment of the conditions for the likelihood of seeing a harbor porpoise and the observer's assessment of the effect of glare, sea state, as well as less quantifiable factors such as turbidity, sun angle, unusual weather conditions, and fatigue on the observer's ability to sight a harbor porpoise. The observers reported these environmental data as changes in such data were noticed along a transect. For each sighting, the observer notified the computer operator when the beam line of the plane crossed the animal's location. The primary observers used inclinometers to obtain the vertical angle below the horizontal to convert the perpendicular distance of the animal from the trackline (Lerczak and Hobbs, 1998). To determine the distance of a sighting from the trackline indicated by a center line on the belly window, the window was subdivided with a grease pencil into six 10°-bins (out to 30° to either side of the trackline for an averaged eye height), labeled 1–6 from port to starboard. When alerted to a sighting by the primary or independent observers, the computer operator immediately entered the sighting by using a hot key assigned to an observer (which recorded the observer's initials and which captured the time and position from the GPS unit). The hot key also opened a window for entering species name, vertical angle or angle bin, group size, and any notable animal behavior.

Matching sightings from side and belly windows

Sighting data (time, perpendicular distance, species, and group size) collected on the same transects were compared between side and belly observers. For comparison purposes, left- and right-side sighting angles were converted to corresponding belly observer bin number. Sightings were considered matches (same group seen by both observers) if they 1) occurred within 5 seconds of each other; 2) were not greater than one 10° bin difference; and 3) met other conditions such as a species of similar size or of hierarchical relation (e.g., harbor porpoise matched to unidentified small cetacean) and similar group size. Matched sightings were used 1) to estimate an empirical average angle for each belly window bin, based on the angles measured from the side windows; 2) to identify circumstances resulting in unreliable species identifications (see *Errors in species identification* in Appendix I); 3) to estimate bias in group-size estimates by the belly observer; 4) to estimate perception bias and $g(0)$ (here $g(0)$ accounts only for the consequences of perception bias; correction for availability bias is treated separately as described below); and 5) to eliminate duplicate sightings from the distance analysis.

Correction for bias in group-size estimates determined by belly observers

Initial inspection of the data when both the side and belly observers reported a sighting indicated that the group size estimate of the belly observer was occasionally less than that provided by the side observer—a result of the restricted visual field and limited observation time for the belly observer. For each of these pairs, the count by the side observer was divided by the count by the belly observers. These ratios were then grouped by belly observer group size and averaged to estimate a correction for each group size reported by the belly observer. The standard error for each correction factor was estimated by the usual formula. The correction was applied to all group sizes from belly sightings included in the average estimate of group size.

Distance smearing

Angle rounding occurred in both the side observer data and the belly observer data. In the case of the side observer data, peaks in frequency occurred on multiples of 5°. The rounding of angles often occurred after the sighting was out of the field of view and the observer estimated the angle from a remembered location. The accuracy for these remembered locations may not have been any better than 5°, and therefore created a tendency for observers to use a close 5° increment number rather than one of the marks in between. Belly observer data were assigned to a bin and were thus automatically rounded. To remove these effects, side sightings were dithered uniformly over 13 m (2.5° on either side of the reported angle) and belly sightings were dithered uniformly over 26 m (5° on either side of the reported angle, the center of sighting bins). The dithering distance was chosen empirically as the minimum distance necessary to remove the rounding effect. The dithering was repeated several times and the cumulative distribution of the sightings by distance to the trackline was examined. An instance of the dithering which gave a visually smooth distribution was retained and used as the data set for further analysis to estimate the sighting distribution.

Estimation of perception bias and $g(0)$

All three years of data were combined to estimate perception bias from comparisons of the primary and the independent observer sightings. Logistic regression with a generalized linear model (the GLM function in S-PLUS, Lucent Technologies, Murray Hill, NJ) and an offset algorithm for comparison of paired sightings (Buckland et al., 1993) was used to estimate the perception bias of the side and belly observers on the trackline. Review of the ratio of matched to unmatched sighting for the belly and side observers by 25-m bins indicated that the two inner bins were consistent with each other (0–25 m and 25–50 m), whereas the outer bin (50–75 m) differed. Consequently, perception bias was estimated by using only sightings within 50 m of the trackline (approximately 20° at the standard survey altitude or bins 2 through 5 in the belly window). Sightings beyond this cutoff distance were excluded. Possible covariates in the logistic regression were visibility, sea state, cloud cover, glare, group size, observer, and survey year. Covariates were initially tested individually to identify functional forms or groupings that could reduce the number of parameters necessary to represent them. The discrete covariates (visibility, sea state, glare, group size, observer, and survey year) and the continuous covariate cloud cover (grouped into five categories: 0–20%, 20–40%, 40–60%, 60–80%, 80–100%), were examined individually as categorical factors. The coefficients from the categorical analysis were then charted against their hierarchical ranks. Where the coefficients appeared to fit a simple functional form of the hierarchical ranks (line, square root, natural logarithm, exponential) or could be grouped to reduce the number of parameters, the analysis was repeated with this alternative. The parameters for the function or grouping were estimated by using the regression described above and compared to the result of the categorical factor by using Akaike's information criterion (AIC). The function or grouping was used in the subsequent analysis if it improved the AIC. From this preliminary analysis, visibility and sea state were found to have a nearly linear effect and were treated as linear functions by using the hierarchical number as the value and by setting the best condition to one; group size was also considered to be linear. Cloud cover, glare, observer, and survey year were considered as categorical data. Significant covariates were then combined in the GLM model and removed in a stepwise manner until the AIC had been minimized. The perception bias for each observer position and each transect segment was estimated from the final model and combined to estimate $g(0)$ for each transect segment (see Appendix II for details).

The program DISTANCE, vers. 3.5 (Thomas et al., 1998) allowed only a global $g(0)$ and thus did not accommodate $g(0)$ to be estimated for each transect segment from environmental and observer covariates. It was possible to circumvent this limitation by adjusting the length of each trackline to allow an estimate of density in the vicinity of each trackline because $g(0)$ and length are multiplied together to estimate density. The estimates of $g(0)$ for each transect segment were averaged for all three years weighted by the transect lengths to estimate an average $g(0)$. The length of each transect segment was multiplied by its estimated $g(0)$ divided by the average $g(0)$ to generate an adjusted transect segment length which accounted for the $g(0)$. The adjusted transect segment lengths were then used in DISTANCE in place of the actual lengths and the average $g(0)$ calculated above was used as the global $g(0)$ in DISTANCE. The standard error for the global $g(0)$ was estimated as the weighted average of the standard errors of the $g(0)$ estimates for the individual transect segments.

Estimation of abundance

The line-transect analysis program DISTANCE (vers. 3.5) was used to estimate the observed density of harbor porpoise in each surveyed region. Two sighting probability curves were estimated so that for transect segments with usable belly observer effort data, sightings from the side and belly observers could be combined and duplicates removed or, when no belly observer data were available, sightings from the side observers only could be used. To identify significant effects of possible covariates for estimated strip width (presence or absence of a belly observer, survey year, individual observers, visibility levels, glare types, percent cloud cover, and sea state), each factor was considered separately as a covariate and the one with the lowest value for the AIC was retained. This process was repeated with the remaining possible covariates in an additive manner until further addition of covariates did not lower the AIC. Distances were pooled into 50-m bins to allow application of the estimate of perception bias. Densities were estimated for the individual areas with usable survey data. Unsurveyed areas such as the small bays and inlets were assigned the average densities from the surveyed areas of that stratum. These densities were then averaged, weighted by the area of each survey region, to estimate an average observed density and abundance for each stock. Variances were calculated as in Buckland et al. (2001). The correction factor for availability bias is the inverse of the estimate of availability from Laake et al. (1997) (2.96=(1/0.338), CV=0.18). This factor was applied as a multiplier to the observed abundance estimates to produce the abundance estimate for each stock.

Incorporation of other survey data

The vast area comprising Alaska waters made it impossible to survey all areas where harbor porpoises occur. Harbor porpoise sighting data were available from a concurrent NMFS beluga whale line-transect survey in Cook Inlet. For this survey, an Aero Commander aircraft with bubble windows was used; however, the windows were smaller than those of the Twin Otter aircraft, and the observers could not see directly below the plane. Survey methods were similar, except that the beluga whale survey was conducted at an altitude of 244 m and the primary focus was beluga whales. The search effort, therefore, was not concentrated as close to the trackline as it would have been if the survey had been designed to survey harbor porpoise. NMFS National Marine Mammal Laboratory has conducted these beluga whale surveys each year since 1993. We estimated abundance for harbor porpoise in Cook Inlet using the 1998 survey data, a strip width estimated from all beluga surveys (1993 to 1999), and the correction for availability bias from Laake et al. (1997). Perception bias could not be estimated for this survey. This abundance estimate was added to the abundance estimate from the GOA survey to produce a combined estimate for that stock.

Minimum abundance estimate

A minimum abundance estimate, N_{min}, defined in Wade and Angliss (1997) as the lower 20^{th} percentile of the lognormal error distribution, is used in management

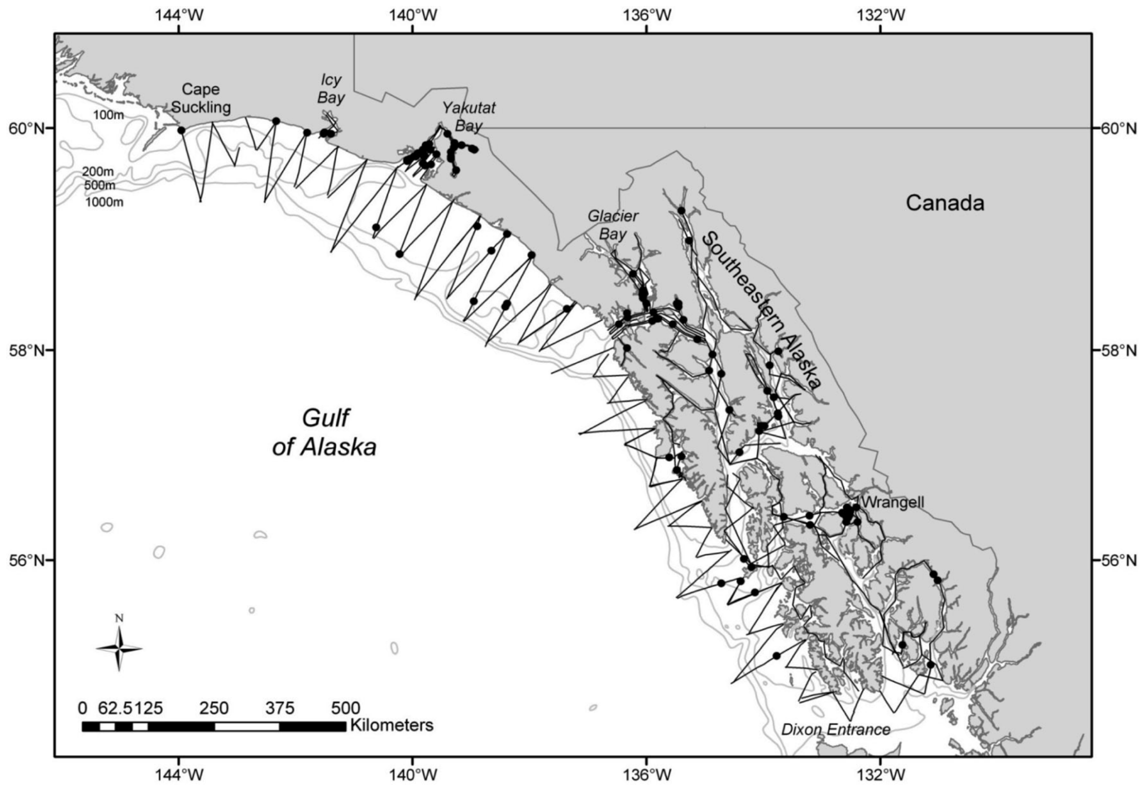

Figure 2

Completed survey transects and sightings (circles) of harbor porpoise (*Phocoena phocoena*) during the 1997 aerial survey in the Southeast Alaska stock region.

decisions by NMFS. This quantity was calculated for each stock from the completed abundance estimates as

$$N_{\min} = Ne^{-0.842\left[\sqrt{\ln(1+CV(N)^2)}\right]},$$

where N = estimated abundance; and
$CV(N)$ = the estimated coefficient of variation of N.

Results

The 1997 line-transect aerial survey was conducted from 27 May to 7 June and 10–28 July 1997 in the inside waters of southeastern Alaska, Yakutat Bay, Icy Bay, and in offshore waters from Dixon Entrance to Cape Suckling (Fig. 2). Necessary repairs on the survey plane resulted in an unplanned month-long break in the survey, and adverse weather prevented a second survey of offshore waters. A total of 9844 km were surveyed. The 1998 survey was conducted from 27 May to 28 July 1998 in Prince William Sound, the western Gulf of Alaska (from Cape Suckling to the west side of Kodiak Island), and Shelikof Strait (Fig. 3). Gaps in the survey effort occurred on account of inclement weather, primarily off the Kenai Peninsula and the southern side of the Alaska Peninsula west of Kodiak Island. A total of 9486 km were surveyed. The 1999 survey was conducted 11 June to 4 July in Bristol Bay and associated bays. In addition, an area south of the Alaska Peninsula west of Chignik Bay was surveyed that was not completed in 1998 (Figs. 3 and 4). A total of 8490 km were surveyed. The 1999 data for the Gulf of Alaska was included with the 1998 data to estimate abundance of harbor porpoise for the Gulf of Alaska.

Sightings of harbor porpoise for each region (Figs. 2–4) were more common in nearshore areas, but occurred throughout the depth range surveyed during all three surveys. High densities of harbor porpoise were found in Yakutat Bay and near Wrangell (Fig. 2), between Prince William Sound and Cape Suckling, on the southeast side of Kodiak Island, southwest of Chignik Bay (Fig. 3), and in a few small bays on the northern side of Bristol Bay (Fig. 4).

Corrections for species misidentification and for under-counting group sizes by belly observers

A cursory examination of the discrepancies in species identification between side and belly observers from the 1997 and 1998 seasons indicated that discrepancies occurred primarily when inexperienced observers were at the belly window and had fewer than 10 days of

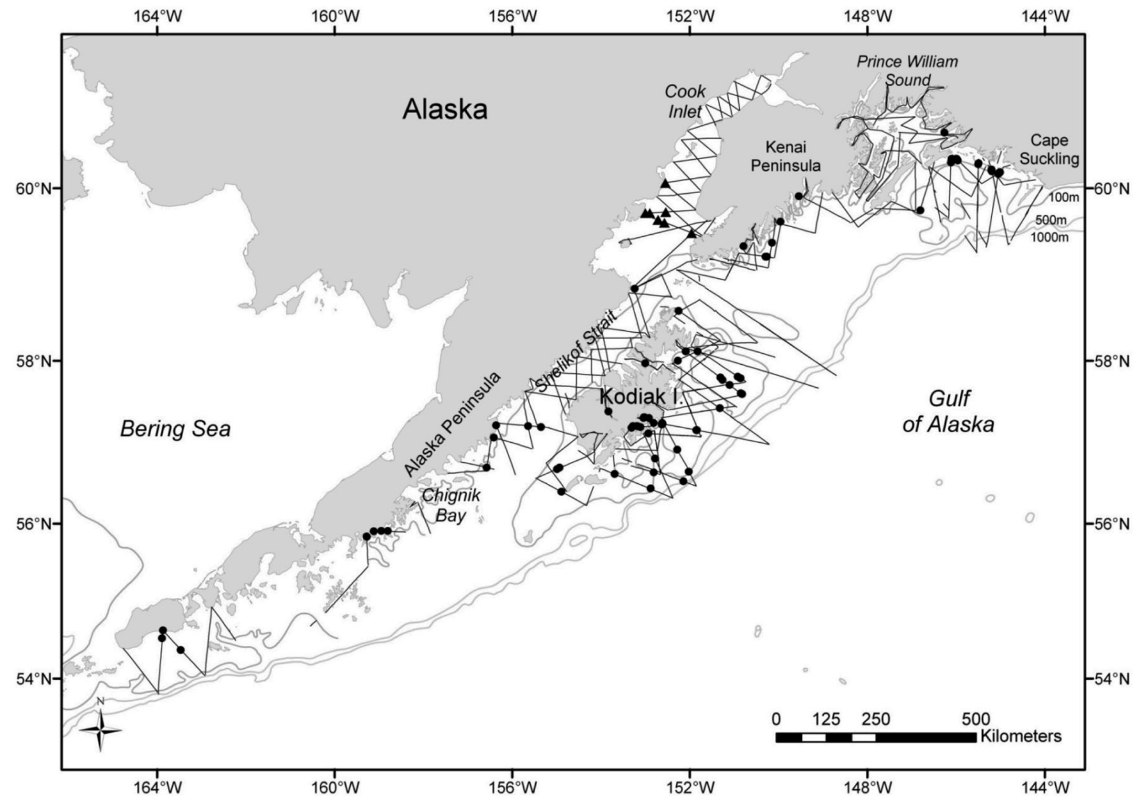

Figure 3

Completed survey transects and sightings (circles) of harbor porpoise (*Phocoena phocoena*) during the 1998 aerial survey and the 1998 beluga whale aerial survey (triangles) in the Gulf of Alaska stock region. Also shown are transects and sightings of harbor porpoise made during the 1999 aerial survey south of the Alaska Peninsula and west of Chignik Bay.

survey experience. Based on this *ad hoc* assessment of discrepancies, an experienced observer was defined as one who had 10 or more days of observation experience on the survey. Species discrepancies involving harbor porpoise were those that led to a misidentification of harbor porpoise as either a Dall's porpoise or harbor seal (*Phoca vitulina*).

Within the data from the first two years of the survey, there were 68 species identifications of harbor porpoise, Dall's porpoise, and harbor seal from paired observers. Of these 68 identifications, 52 were determined by paired experienced observers in the side and belly (one discrepancy), 12 were determined by an experienced side observer paired with an inexperienced belly observer (4 discrepancies), and 4 were determined by an inexperienced side observer paired with an experienced belly observer (no discrepancies). No correlation between discrepancies and environmental conditions was found. To verify the observation that the inexperienced observers in the belly position had a higher than average misidentification rate and determine if the rate was unacceptable, four possible models were compared and AIC was used to identify the most parsimonious model. The models were 1) side observers and experienced and inexperienced belly observers were all different (four parameters); 2) experienced and inexperienced side observers and experienced belly observers were equivalent and inexperienced belly observers were different (two parameters); 3) experienced and inexperienced observers were different but side and belly were equivalent (two parameters); and 4) all observers were equivalent (one parameter). For model 1, probabilities of a correct identification were >0.99 for both experienced and inexperienced side observers, 0.98 for experienced belly observers, and 0.67 for inexperienced belly observers with an AIC of 13.0. For model 2, side observers with experienced belly observers had a probability of 0.98, and inexperienced belly observers had a probability of 0.67 with an AIC of 9.1. Model 3 resulted in a probability of 0.99 for experienced observers and >0.76 for inexperienced observers with an AIC of 11.6. Model 4, a probability for all observers, was 0.96 with an AIC of 17.6. The most parsimonious model (lowest AIC) was model 2, indicating that an inexperienced observer in the belly position had a low reliability for species identification. Consequently, observation effort and sightings by inexperienced observers in the belly during their first 10 survey days were treated as practice and were not included in the subsequent analysis. Although it would be possible to estimate a $g(0)$ that accounted

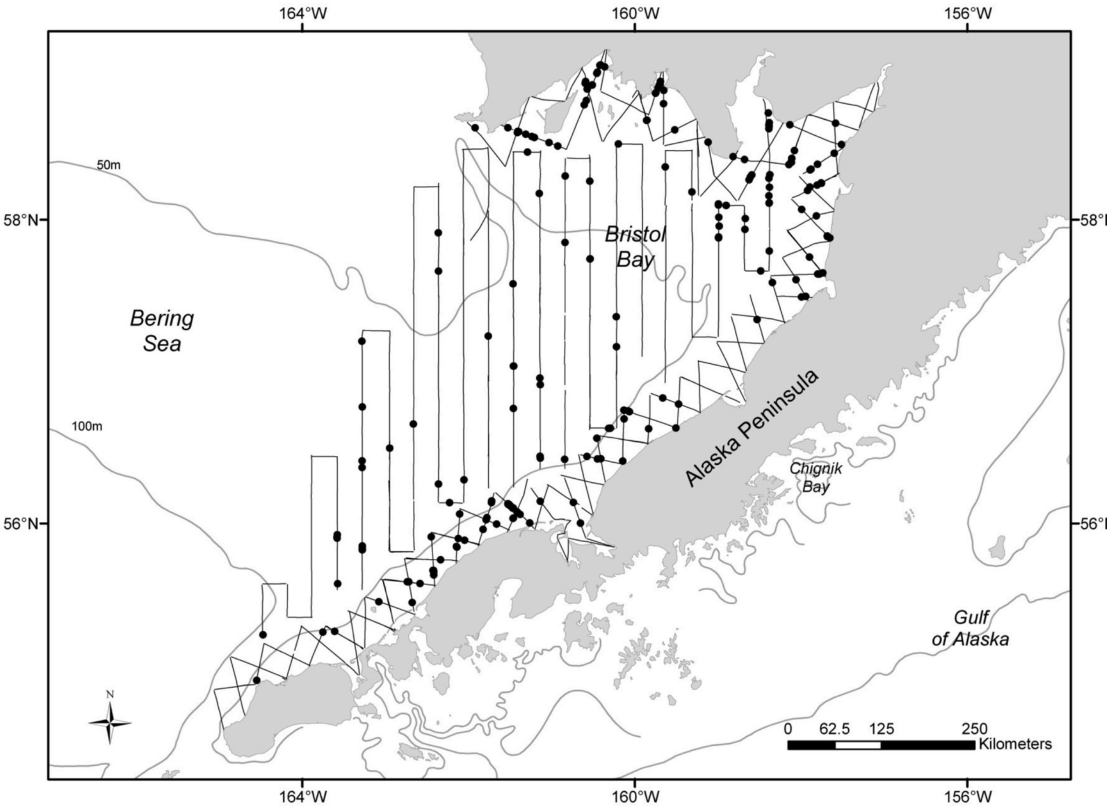

Figure 4

Completed survey transects and sightings (circles) of harbor porpoise (*Phocoena phocoena*) during the 1999 aerial survey in the Bering Sea stock region. Transect and sightings on the south side of the Alaska Peninsula are shown in Figure 3 as part of the Gulf of Alaska survey.

for the inexperienced observers in the belly, we did not have the estimate of the density for the other species necessary to complete this calculation.

Group size was typically underestimated by the belly observer in comparison to the side observers. In sightings by both the belly and side observers, separate corrections were calculated for group sizes reported as one individual and groups reported as two individuals by the belly observer. Of 30 groups reported as one harbor porpoise by the belly observer, 25 were reported as a group size of one by the side observer and 5 were reported as a group size of 2, yielding a multiplicative correction of 1.167 (CV=0.059) groups of size one seen only by the belly observer. Likewise, of 12 groups reported as two harbor porpoise by the belly observer, 11 were reported as a group size of 2 by the side observer and one was reported as a group size of 3, yielding a multiplicative correction of 1.042 (CV=0.080) for groups of 2 observed from the belly of the aircraft . The group-size estimate from the side observer was used when a group was reported by both the side and belly observers. The correction was only applied to group size when the group was seen only by the belly observer. A correction for total animals was estimated as the sum of the group sizes with the belly-window–derived groups sizes corrected, divided by the sum of the group sizes with the belly-window–derived group sizes uncorrected. As a result, a multiplier of 1.018 (CV=0.006) was applied to the abundance estimates. It was necessary to apply a general correction rather than correct individual group sizes to avoid problems with the $g(0)$ estimate and DISTANCE analysis arising from non-integer group sizes.

Estimation of perception bias and $g(0)$

Comparisons between sightings by the belly observer and the side observers indicated that each missed a small but significant fraction of the near-surface animals on the trackline. A total of 129 potential matches between experienced belly observers and independent side observers within 50 m of the trackline were examined for perception bias and $g(0)$ estimation. Although several of the potential covariates were significant by themselves, only visibility as a continuous variable remained in the stepwise elimination. A significant difference in estimated strip width between survey year 1997 and the years 1998 and 1999 was identified in the distance analysis and, therefore, year was included as a covariate as well. Although year was not a significant coefficient, there was a significant turnover in personnel from the

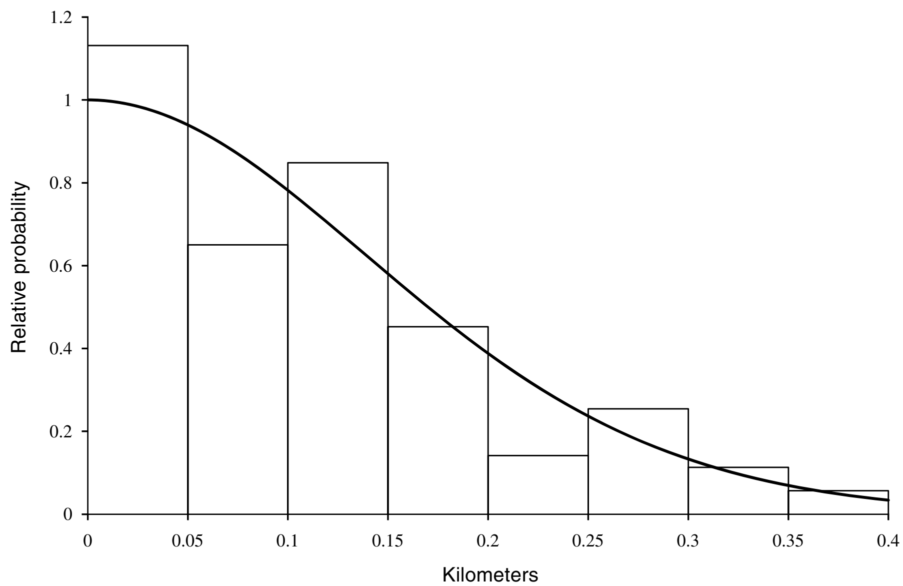

Figure 5

Relative probability of detection of harbor porpoise (*Phocoena phocoena*) by distance from the survey trackline (km) determined with a half-normal model for the 1997 sighting data of harbor porpoise in the Southeast Alaska stock region. The histogram shows the distribution of sightings in 0.05-km groups.

first to the second and third years, and 1997 was the first year of a belly observer for the survey leaders. The belly observer had a significantly higher probability of sighting an available group than the side observers for any particular sighting, and the belly observers routinely reported better visibility than did the side observers. This difference in perception bias was accounted for by the difference in reported visibility. The logistic regression coefficients were as follows: constant=1.187 ±0.542, (t=2.19); for year=0 for 1997 and 1 for 1998, 1999, coefficient=0.296 ±0.290, (t=1.02); for visibility=1 (excellent), 2 (good), 3 (fair), 4 (poor), 5 (unacceptable) as a continuous variable, coefficient=–0.502 ±0.217, (t=–2.31). The model for probability of sighting of an available group for a single observer is then

$$P\left(sighting \middle| year, visibility\right) = \frac{e^{1.187+0.296\,year-0.502\,visibility}}{1+e^{1.187+0.296\,year-0.502\,visibility}}.$$

Heterogeneity in probability of sighting a harbor porpoise resulted in a decrease of a difference of roughly 0.11 for each reduction in visibility and an increase by roughly 0.06 from 1997 to 1998–99 (Table 1). The average observed $g(0)$ values (perception bias only) for the SEA stock, the GOA stock, and BS stock of harbor porpoise were 0.641 ±0.069, 0.729 ±0.048, and 0.748 ±0.046, respectively, and yielded average perception bias correction factors of 1.560 (CV=0.108), 1.372 (CV=0.066), and 1.337 (CV=0.062), respectively (Table 2).

Estimated strip width for observations

Variation in effective strip width (ESW) occurred for the configuration of observers and visibility as reported by the observers. Few sightings occurred beyond 400 m and therefore this distance was chosen as the truncation point for distance from the trackline and sightings beyond this distance were not included in the analysis. Effort was separated into effort with and without a belly observer. ESW without a belly observer was the ESW on one side of the plane covered by a side observer. ESW with a belly observer represented the effort of one side observer and half of the belly observer because the belly observer's field of view was divided by the trackline (note that duplicate sightings were removed such that where sightings were reported by both the side and belly observer, only the side sighting record was used). The ESW for 1997 was significantly different from that of 1998 and 1999; therefore they were treated separately (Table 1). In the 1997 data, significant variation in ESW was related to visibility level and in the 1998 and 1999 data, to the presence or absence of the belly observer. The best fit for the detection function was a half-normal curve for the 1997 data set and a half-normal curve with a one term cosine correction for the 1998–99 data set (Figs. 5 and 6, Table 2). The ESW of the survey team in 1997 decreased by roughly 20% per step change in visibility. When this decrease in ESW was combined with approximately a 12% decrease in $g(0)$ with each step in visibility, the product (Table 1) indicated an

Table 1

Estimated $g(0)$ (probability of detecting an animal at the surface on the trackline (perception bias) and effective strip width (ESW) in km for harbor porpoise in Alaska determined from surveys from 1997 through 1999. Data were obtained from individual observers and teams of observers, all of whom reported the same visibility code. The product of these two values, $g(0)$ and ESW, is a measure of the relative effectiveness of the observer team under different conditions. Single observer=single observer at either the right, left, or belly window of aircraft. Team of observers=a team of observers at the right, left, and belly window of the aircraft.

		Single observer				Team of observers			
		1997		1998–99		1997		1998–99	
Visibility		Value	SE	Value	SE	Value	SE	Value	SE
1 (excellent)	$g(0)$	0.66	0.08	0.73	0.06	0.89	0.01	0.93	0.01
	ESW	0.252	0.022	0.139	0.008	0.252	0.022	0.118	0.007
	$g(0)$ESW	0.166	0.009	0.101	0.004	0.224	0.002	0.11	0.001
2 (good)	$g(0)$	0.55	0.06	0.62	0.04	0.79	0.02	0.85	0.01
	ESW	0.2	0.017	0.139	0.008	0.2	0.017	0.118	0.007
	$g(0)$ESW	0.11	0.004	0.086	0.002	0.158	0.003	0.1	0.001
3 (fair)	$g(0)$	0.42	0.07	0.49	0.06	0.66	0.04	0.74	0.02
	ESW	0.153	0.013	0.139	0.008	0.153	0.013	0.118	0.007
	$g(0)$ESW	0.064	0.002	0.068	0.002	0.101	0.003	0.087	0.001
4 (poor)	$g(0)$	0.31	0.1	0.37	0.09	0.52	0.11	0.61	0.08
	ESW	0.116	0.01	0.139	0.008	0.116	0.01	0.118	0.007
	$g(0)$ESW	0.036	0.001	0.051	0.002	0.06	0.003	0.072	0.004
5 (unacceptable)	$g(0)$	0.21	0.11	0.26	0.12	0.38	0.19	0.46	0.16

approximately 30% decrease in effective effort with each step in visibility and that survey effort during poor conditions had less than one quarter of the effectiveness of effort during the best conditions (Fig. 7). In the 1998–99 surveys (Fig. 8), the ESW was narrower overall and slightly broader when the belly observer was not present. Although this seems counterintuitive, it is the result of a peak that occurred near the trackline when the belly sightings were included and which made the distribution away from the trackline relatively lower and resulted in the narrower ESW (Table 1, Fig. 8). When ESW and $g(0)$ were multiplied together, the added value of the belly observer was 10% under the best conditions and nearly 50% under poor conditions.

Density and abundance of harbor porpoise

Abundance estimates of harbor porpoise increased from east to west as did estimates of average density by stock (0.10, 0.19, and 0.44 porpoise/km^2, respectively). Average observed harbor porpoise densities (uncorrected for availability or perception biases) for the SEA, the GOA, and the BS stocks were 0.033 groups/km^2 (CV=17.2%), 0.062 (CV=11.9%), and 0.153 (CV=13.2%), respectively. Approximately 5% of the study areas, consisting primarily of inlets and channels, were unsurveyed. Density estimates for these unsurveyed areas were extrapolated from similar surveyed areas in the same general region (Table 3). The correction factor of 2.96 (CV=0.180) (Laake et al., 1997) was applied to each abundance estimate to account for availability bias. The full corrections for visibility bias (correction for perception bias × correction for availability bias) were 4.62 (SEA, CV=21%), 4.06 (GOA, CV=19%), and 3.96 (BS, CV=19%).

For the Cook Inlet survey, the effective strip width (0.280 km, CV=0.281) was based on 44 sightings from the 1993 to1999 surveys. Truncation of the sighting strip by discarding sightings less than 0.1 km from the trackline or greater than 0.6 km from the trackline on each side of the plane was necessary to obtain a good fit of the detection function. The best fit for the detection function for the Cook Inlet data, based on AIC, was a hazard-rate curve with a cosine correction (Fig. 9). The 1998 beluga whale survey in Cook Inlet resulted in eight harbor porpoise sightings along 1355 km of trackline. No data were available to estimate the perception bias for this survey and its format was sufficiently different from the harbor porpoise survey with the result that it was uncertain whether the perception bias correction would be approximately correct. Consequently, only the correction for availability (2.96, CV=0.180) was applied. This results in a rather conservative estimate with a known negative bias which we feel is preferable to one with an unknown bias.

The abundance estimate for the SEA stock of harbor porpoise was 11,146 animals (CV=24.2%; N_{min}=9116,

Table 2

Survey parameters and abundance estimates for harbor porpoise (*Phocoena phocoena*) stocks off Southeast Alaska, the Gulf of Alaska, and the Bering Sea in 1997, 1998, and 1999, respectively. Cook Inlet survey results are taken from a survey for beluga whales which followed a protocol similar to the harbor porpoise surveys. No perception bias correction was available for the 1998 Cook Inlet survey; consequently observed abundance is used in place of total abundance for this area. The Cook Inlet abundance was included in the Gulf of Alaska stock total abundance. Extrapolated areas were small inlets and unsurveyed areas where the density of harbor porpoise was assumed to be the same as similar surveyed areas for the purpose of estimating total abundance. Coefficient of variation (CV) of a statistic is the standard error of the statistic divided by the statistic; confidence intervals are calculated by using a log-normal distribution and $N_{min} = Ne^{-0.842\sqrt{\ln(1+CV(N)^2)}}$.

	Southeast Alaska		Gulf of Alaska		Cook Inlet Survey		Bering Sea	
	Estimate	CV(%)	Estimate	CV(%)	Estimate	CV(%)	Estimate	CV(%)
Study region (km^2)	106,087		158,733		18,948		106,381	
Total trackline (km)	9844		10,127		1355		7849	
No. if sightings	129		88		8		201	
Average correction for perception bias (1/g(0))	1.560	10.8%	1.372	6.6%	1		1.337	6.2%
Effective half strip width (km)	0.182	8.7%	0.122	5.8%	0.280	28.1%	0.122	5.8%
Average group size (no. of individuals)	1.279	4.4%	1.289	2.3%	1.129	5.0%	1.289	2.3%
Corrected average group density (groups/km^2)	0.026	16.6%	0.048	11.7%	0.012	60.5%	0.119	13.0%
Average porpoise density (porpoise/km^2)	0.033	17.2%	0.062	11.9%	0.013	60.7%	0.153	13.2%
Uncorrected abundance in surveyed areas	3505	17.2%	9791	11.9%	249	60.7%	16,289	13.2%
Extrapolated area (km^2)	6539		4722					
Abundance in extrapolated area	261	40.5%	449	60.6%				
Total uncorrected abundance	3766	16.2%	10,489	11.5%			16,289	13.2%
Correction for availability (from Laake, et al., 1997)	2.96	18.0%	2.96	18.0%			2.96	18.0%
Average perception correction ×availability correction	4.62	21.0%	4.06	19.2%	2.96	18.0%	3.96	19.0%
Total abundance (N)	11,146	24.2%	31,046	21.4%			48,215	22.3%
N_{min}	9116		25,987				40,039	
Lower 95% confidence limit	6980		20,520				31,285	
Upper 95% confidence limit	17,788		46,972				74,308	

Table 1). The abundance estimate for the GOA stock, which included the Cook Inlet harbor porpoise abundance estimate, was 31,046 animals (CV=21.4%; N_{min}=25,987, Table 1), and the abundance estimate for the BS stock was estimated as 48,215 animals (CV=22.3%; N_{min}=40,039, Table 1).

Discussion

Habitat type may account for the increase in density of harbor porpoise from east to west—at least for the much higher abundance of the BS stock compared to the other two stocks. The Bering Sea encompasses a vast sea ranging from a large shallow bay (Bristol Bay) extending to a large shelf that descends to the abyssal sea. Our entire survey of this stock region was conducted in Bristol Bay where water depth never exceeds 100 m. In contrast, the shelf area is narrower in the Gulf of Alaska so that the surveys off Southeast Alaska and in the Gulf of Alaska were routinely conducted in waters up to 200 m, and occasionally up to 1800 m. Despite the greater ranges of depths surveyed in Southeast Alaska and the Gulf of Alaska, harbor porpoise were present primarily in waters less than 100 m in depth. Off northern California, higher than expected numbers of harbor porpoise were found between the 20 m to 60 m isobaths and fewer than expected in waters deeper than 60 m (Carretta et al., 2001). Similarly, Barlow (1988) found harbor porpoise primarily distributed in waters shallower than 110 m in depth. In contrast, Raum-Suryan and Harvey (1998) found that harbor porpoise near the northern San Juan Islands, Washington, were present at depths greater than 100 m. Differences in

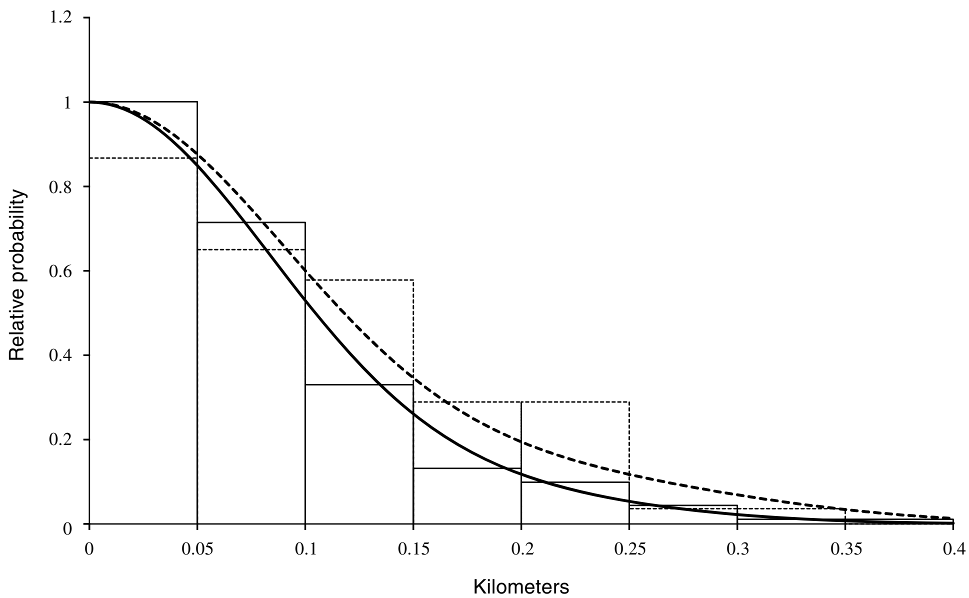

Figure 6

Relative probability of detection of harbor porpoise (*Phocoena phocoena*) by distance from the survey trackline (km) determined with a half-normal model with cosine adjustment for the 1998 and 1999 harbor porpoise sighting data. The solid line represents the probability of a sighting when both side observers and a belly observer were present. The dashed line represents the probability of a sighting when only side observers were present (without a belly observer). The histogram shows the distribution of sightings in 0.05-km groups.

harbor porpoise occurrence by depth may account for the higher density estimated for the Bering Sea stock; however, the survey comprised only a portion of the entire stock.

The SEA stock abundance estimate is not significantly different from the 1991–93 abundance estimate. The abundance estimates for the GOA stock (31,046) and the BS stock (48,215) are significantly higher than the 1991–93 abundance estimates (8497 and 10,946, respectively) (t-test, natural log of means, $P<0.01$). It should be noted that the GOA stock abundance estimate may be biased low because it includes a survey of Cook Inlet which could not be corrected for perception bias, and the BS stock may have been underestimated as described in the previous paragraph. However, differences in survey design with the earlier surveys confound direct comparison between the abundance estimates. Overall, the area covered in the 1997–99 surveys was larger than that of the 1991–93 surveys and included a wider range of possible harbor porpoise habitat. The 1997–99 surveys were designed to include a sample of bays and inlets within the study region that the earlier surveys did not sample. The 1997–99 survey also included some larger bodies of water, such as Icy Bay and the inside waters of Southeast Alaska, that were not included in the earlier survey and gave more thorough coverage to some areas such as Yakutat Bay and Prince William Sound. The offshore extent of the 1997–99 survey was determined by water depth rather than distance, which extended it farther offshore in the Southeast Alaska and Gulf of Alaska stock regions. The 1999 survey in the Bering Sea stock region covered much of the same area as the 1991 survey but at a higher density of effort. In 1999, the survey area to the south of the western end of the Alaska Peninsula (a survey area that was not completed in 1998) was surveyed. This area was not surveyed in the 1991–93 surveys. The survey design allowed for the inclusion of potential harbor porpoise habitat that was not covered in the previous surveys, especially areas such as Yakutat Bay and Sitkalidak Strait (Kodiak Island). Another difference between the surveys was the use of correction factors. A perception-bias correction was estimated from independent observer data, and therefore only the Laake et al., 1997 correction of 2.96 for availability bias was required to make a combined visibility correction factor of 4.62 for the SEA stock, 4.06 for the GOA stock, and 3.96 for the BS stock; these correction factors are 49%, 31%, and 28%, respectively, larger than the factor of 3.1, used in the 1991–93 surveys (Hill and DeMaster, 1998). The correction factors used in our analyses better reflect conditions encountered during aerial surveys for porpoise in Alaska because they incorporate a direct measure of animals missed by observers during the surveys, as well as the best available estimate of the animals missed while out of view underwater.

It is likely that the shorter sighting time for the observer in the belly window increased the probability that inexperienced observers misidentified species of similar size during observations. Observers in the belly position of the aircraft during this survey had approximately 2–4 seconds to perceive, identify, and enumerate a group of animals. This is about half of the time available to the side observers and leaves little time for the observer to double check cues to distinguish among species. Thus, the observers are left with their first impressions which may be mistaken if there is little prior experience in observing and recording individual and groups of harbor porpoises. Laake et al. (1997) found a difference in perception bias between experienced ($g(0)=0.86$) and inexperienced ($g(0)=0.23$) observers in an experiment where the sighted species was known. We concur with Laake et al. (1997) that experienced observers should be positioned at the belly window and a training period should be considered for new aerial observers before their data from the belly position is used to estimate $g(0)$.

This analysis was completed in 2000 with the software that was available (DISTANCE, vers. 3.5), which did not include features to use multiple resights, so that perception bias had to be estimated separately. The current software DISTANCE 5.0 can use multiple resight data to estimate perception bias but does not correct for bias in the estimation of group size or for errors in species identification. Although some of the components of the analysis presented here are now completed automatically within the current software, the analysis of observer performance would have to be completed separately.

Acknowledgments

Funding for this project was provided by Recover Protected Species Program, National Marine Fisheries Service, National Oceanic and Atmospheric Administration. We thank A. Andriolo, R. Angliss, K. Forney, K. Leinbach, J. Lerczak, L. Litzky, M. Merklein, S. Moore, S. Norman, D. Rugh, K. Shelden, V. Vanek, P. Wade, and K. Wynne for participating as observers. We thank K. Shelden for acting as field team leader on several occasions. We thank the pilots of the NOAA Twin Otter (M. Finn, J. Hagan, P. Hall, J. Longenecker, M. Moran, M. Pickett, and T. Strong) for their dedication and excellent handling of the aircraft. D. DeMaster, S. Moore, D. Rugh, J. Laake, and N. Friday provided valuable reviews of this manuscript. This research was conducted under permit no. 782-1438 issued by the National Marine Fisheries Service.

Table 3

Estimates of abundance of harbor porpoise (*Phocoena phocoena*) in unsurveyed inlets and channels for the 1997–99 aerial surveys in Alaska. Density of harbor porpoise was averaged for similar areas that were surveyed and then used to extrapolate density and abundance in the similar unsurveyed areas. These extrapolated abundances were summed by stock and included in the total abundances of the Southeast and Gulf of Alaska stocks.

	Southeast Alaska stock			Gulf of Alaska stock				
	Southeast Alaska	Frederick Sound	Southeast Alaska total	Kenai Peninsula	Kodiak Island	Alaska Peninsula	Prince William Sound	Gulf of Alaska total
Number of comparable areas (e.g., inlets and channels) surveyed	32	2	34	3	9	2	5	19
Area surveyed (km^2)	4792	2564	7356	439	2782	202	4188	7611
Weighted average porpoise density in surveyed regions	0.037	0.080		0.130	0.151	0.092	0.008	
Weighted standard deviation of densities	0.129	0.058		0.071	0.255	0.423	0.020	
Number of unsurveyed areas	326	1	327	15	17	21	9	62
Unsurveyed area (km^2)	6135	404	6539	662	1356	1617	1087	4722
Estimated abundance in unsurveyed areas	229	32	261	86	205	149	8	449
Standard error of abundance for unsurveyed regions	103	24	106	20	143	230	11	272
% Coefficient of variation of abundance for unsurveyed areas	45%	73%	40%	23%	70%	154%	130%	61%

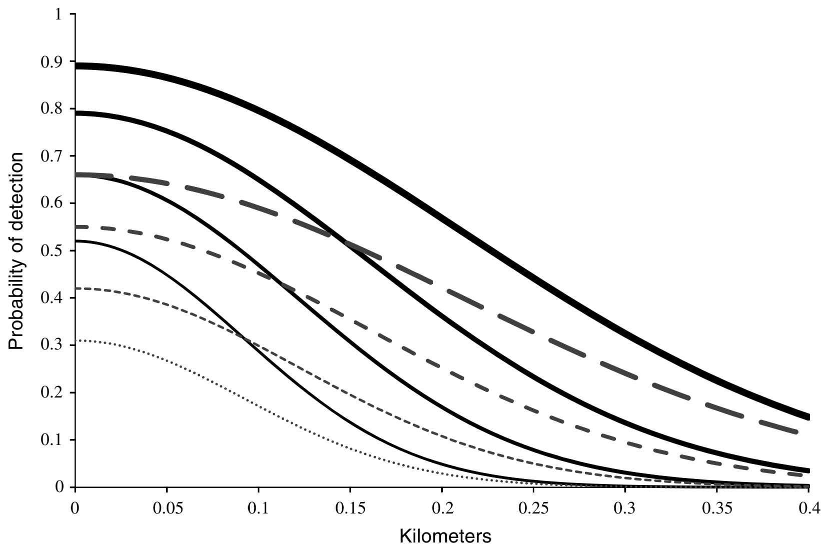

Figure 7

Probability of sighting an available group of harbor porpoise (*Phocoena phocoena*) during the 1997 aerial survey in Southeast Alaska by distance (km) from the survey trackline. Solid lines represent the probability of a sighting when both side observers and a belly observer were present. Thickness of the line (thick to thin) represents visibility codes excellent, good, fair, and poor. Dashed lines represent the probability of a sighting with only side observers (without a belly observer), where long to short dashes represent visibility codes excellent, good, fair, and poor.

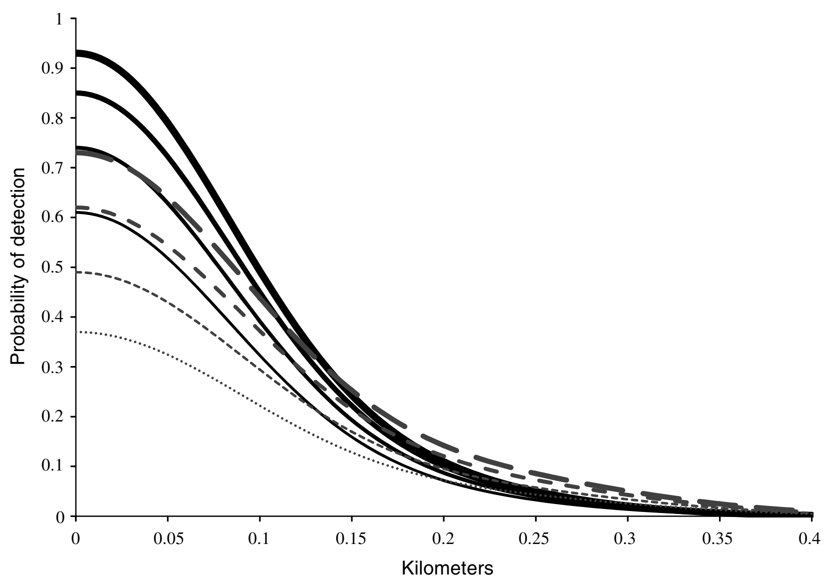

Figure 8

Probability of sighting an available group of harbor porpoise (*Phocoena phocoena*) during the 1998 and 1999 aerial surveys in the Gulf of Alaska and Bering Sea, respectively by distance (km) from the survey trackline. Sighting distributions were sufficiently similar between the two years that a single curve was used for both. Solid lines represent the probability of a sighting when both side observers and a belly observer were present. Thickness of the line (thick to thin) represents visibility codes excellent, good, fair, and poor. Dashed lines represent the probability of a sighting with only side observers (without a belly observer), where long to short dashes represent visibility codes excellent, good, fair, and poor.

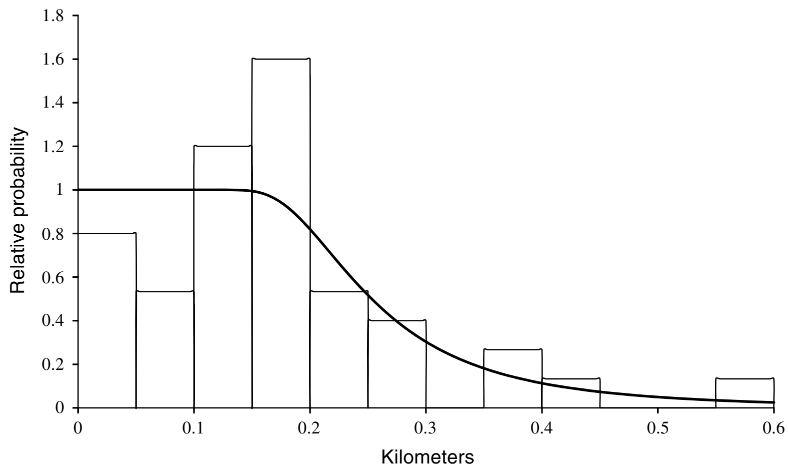

Figure 9

Relative probability of detection by distance from the survey trackline (km) by using a hazard-rate model for the 1998 sighting data for harbor porpoise (*Phocoena phocoena*) in Cook Inlet harbor (relative to 1.0 at zero). The histogram shows the distribution of sightings in 0.05-km groups.

Literature cited

Barlow, J.
1988. Harbor porpoise, *Phocoena phocoena*, abundance estimation for California, Oregon, and Washington: I. Ship surveys. Fish. Bull. 86:417–432.

Buckland, S. T., J. M. Breiwick, K. L.Cattanach, and J. L Laake.
1993. Estimated population size of the California gray whale. Mar. Mamm. Sci. 9:235–249.

Buckland, S. T., D. R. Anderson, K. P. Burnham, J. L. Laake, D. L. Borcher, and L. Thomas.
2001. Introduction to distance sampling: estimating abundance of biological populations, 432 p. Oxford Univ. Press, Oxford, England.

Carretta, J. V., B. L. Taylor, and S. J. Chivers.
2001. Abundance and depth distribution of harbor porpoise (*Phocoena phocoena*) in northern California determined from a 1995 ship survey. Fish. Bull. 99:29–39.

Dahlheim, M., A. York, R. Towell, J. Waite, and J. Breiwick.
2000. Harbor porpoise (*Phocoena phocoena*) abundance in Alaska: Bristol Bay to southeast Alaska, 1991–1993. Mar. Mamm. Sci. 16:28–45.

Hill, P. S., and D. P. DeMaster.
1998. Alaska marine mammal stock assessments, 1998. U.S. Dep. Commer., NOAA Tech. Memo. NMFS-AFSC-97, 166 p.

Laake, J. L., J. Calambokidis, S. D. Osmek, and D. J. Rugh.
1997. Probability of detecting harbor porpoise from aerial surveys: estimating *g*(0). J. Wildl. Manag. 61: 63–75.

Lerczak, J. A., and R. C. Hobbs.
1998. Calculating sighting distances from angular readings during shipboard, aerial, and shore-based marine mammal surveys. Mar. Mamm. Sci. 14:590–598.

Raum-Suryan, K. L., and J. T. Harvey.
1998. Distribution and abundance of and habitat use by harbor porpoise, *Phocoena phocoena*, off the northern San Juan Islands, Washington. Fish. Bull. 96:808–822.

Read, A. J., and A. J. Westgate .
1997. Monitoring the movements of harbour porpoises (*Phocoena phocoena*) with satellite telemetry. Mar. Biol. 130:315–322.

Thomas, L., J. L. Laake, J. F. Derry, S. T. Buckland, D. L. Borcher, D. R. Anderson, K. P. Burnham, S. Strindberg, S. L. Hedley, M. L. Burt, F. F. C. Marques, J. H. Pollard, and R. M. Fewster.
1998. Distance 3.5, release 6. Research unit for wildlife population assessment. Univ. St. Andrews, UK (http://www.ruwpa.st-and.ac.uk/distance/).

Wade, P. R., and R. P. Angliss.
1997. Guidelines for assessing marine mammal stocks: report of the GAMMS workshop April 3–5, 1996, Seattle, WA. U.S. Dep. Commer., NOAA Tech. Memo. NMFS-OPR-12, 93 p.

Appendix I: Errors in species identification

Sightings matched between the side observers and belly observers provided an opportunity to estimate the probabilities that species were misidentified by examining discrepancies in species identification between the matched data. Two types of discrepancies were found: 1) hierarchical discrepancy, where one observer identified the animal (or group) to species while the other observer identified the animal only to genus, family, etc., and 2) mismatched species identification, where two different species were identified. In the case of hierarchical discrepancies, identifications that were not to species could be treated as missed sightings and discarded for the purpose of estimation of species abundance. The mismatched species were of greater concern. There were four possible outcomes when both observers identified the group to the species level: 1) both observers correctly identified the species, 2) one observer was correct and the other was incorrect, 3) both observers were incorrect and disagreed on the species, and 4) both observers were incorrect but agreed on an incorrect identification. From the matched data, types 1 and 4 showed no discrepancy and were thus indistinguishable; types 2 and 3 showed a discrepancy but were also not distinguishable from each other. These discrepancies are assumed to be the result of one of the following: an incorrect identification, an error in reporting by an observer, or a typing error by the data recorder. The errors are assumed to follow a binomial model and to have a generally low probability, so that the likelihood of two errors occurring for the same sighting (outcomes 3 and 4) would be negligible. The following analysis estimates the rates of single errors. Data collected under circumstances with less than 95% reliability for species identification were dropped from the abundance analysis.

The tendency toward errors for species identification can vary by 1) environmental conditions, 2) observer, or 3) recorder. Logistic regression was used to test each of these possible covariates and identify circumstances that were correlated with greater likelihood of discrepancies. A maximum likelihood scheme was then developed to estimate the error rates. Letting p_{ijox} be the probability that an observer o in circumstance x identifies species i as species j, the likelihood (L) that a sighting will be identified as a particular species m is calculated as follows:

$$r_{ai} = R_{ai} \Big/ \sum_i R_{ai},$$

$$L(m,x) = \sum_i r_{ai} p_{imx},$$

where R_{ai} = the actual encounter rate of species i; and
r_{ai} = the fraction of encounters that are species i.

The likelihood of a particular pair of species identifications m and n occurring for a given sighting by one observer in circumstance x and a second observer in circumstance y, is

$$L(m,n,x,y) = \sum_i r_{ai} p_{imx} p_{imy}.$$

In anticipation of a limited data set with a reliability rate greater than 95%, we assumed that outcomes 3 and 4 are rare events compared to outcomes 1 and 2, and therefore we ignored outcomes 3 and 4 in the likelihood model. Second, we assumed that the likelihood of an error is independent of the species involved, and therefore the likelihood is simplified to

$$L(m,n,x,y) \propto \begin{cases} p_x p_y & \text{if } m = n \\ (1-p_x)p_y + (1-p_y)p_x & \text{if } m \neq n \end{cases}.$$

Letting s_{xy} be the number of sighting pairs that occurred under circumstances x and y, and d_{xy} the number of discrepancies in species identification that occurred, the likelihood of a particular set of species identifications was

$$L(S,D,XY) = \prod_{XY} \binom{s_{xy}}{d_{xy}} \left[p_x p_y\right]^{(s_{xy}-d_{xy})} \left[(1-p_x)p_y + (1-p_y)p_x\right]^{d_{xy}},$$

where S = the set of matched sightings;
D = the set of species discrepancies within that set; and
XY = the set of circumstance pairs under which matched sightings were made.

Maximum likelihood solutions were found iteratively for each of the covariate sets identified by the logistic regression as correlated with discrepancies. Observers were stratified into inexperienced (no aerial survey experience before this survey and 10 or fewer days on this survey) and experienced (at least one survey season of experience or more than 10 days on this survey), and environmental factors were considered individually. Likelihoods were compared to identify the most likely model and survey effort under circumstances with less than 95% reliability of correct species identification discarded.

Appendix II: Estimation of $g(0)$ (which accounts for perception bias only)

Perception bias for a single observer, $P(Y)$, was estimated for effort condition vector Y, as the probability that a group of harbor porpoise available to the observer would be perceived and identified to species by an observer from the logistic regression model as

$$P(Y) = \frac{e^{\beta Y}}{1 + e^{\beta Y}},$$

where β = the vector of coefficients estimated in the logistic regression.

Observations from the side observers and belly observer were combined and duplicates removed for density estimation so that the perception bias for the observer team, $P_t(Y_l, Y_b, Y_r)$, (where l, b, and r are the left side, belly, and right side observers, respectively) was the probability that at least one of them would perceive the group. The visual field of the side observers was treated as though their field of view ended at 90° from the horizontal on each side. These observers were in open communication and duplicates were resolved during the survey, so that each observer effectively watched half of the survey trackline. Thus, $g(0)$ for any transect segment with constant environmental conditions was

$$g(0) = 1 - \left[1 - P(Y_b)\right]\left[1 - \frac{P(Y_l) + P(Y_r)}{2}\right].$$

Variance was estimated by the delta method as

$$\mathrm{var}\left(g(0)\right) = \left[D_\beta\left(Y_l, Y_b, Y_r\right)^T \Sigma_\beta D_\beta\left(Y_l, Y_b, Y_r\right)\right],$$

where

$$D_\beta\left(Y_l, Y_b, Y_r\right) = \frac{\partial \Gamma_\beta}{\partial \beta} = \frac{1 - P_b}{2}$$
$$\left[\left(y_b P_b + y_l P_l\right)\left(1 - P_l\right) + \left(y_b P_b + y_r P_r\right)\left(1 - P_r\right)\right],$$

with Γ_B = the variance-covariance matrix for B estimated during the logistic regression;

D_B = the vector of partial derivatives of P_t with respect to the coefficients (B).