Recommendations for photo-identification methods used in capture-recapture models with cetaceans

Kim Urian
Andrew Read
Nicholas School of the Environment
Duke University
Beaufort, NC 28516 USA

Antoinette Gorgone
NOAA Fisheries Service
Southeast Fisheries Science Center
101 Pivers Island Road
Beaufort, NC 28516 USA

Brian Balmer
Randall S. Wells
Sarasota Dolphin Research Program
Chicago Zoological Society
c/o Mote Marine Laboratory
1600 Ken Thompson Parkway
Sarasota, FL 34236 USA

Per Berggren
School of Marine Science and Technology
Newcastle University
Newcastle upon Tyne NE1 7RU
U.K.

John Durban
Tomoharu Eguchi
Marine Mammal & Turtle Division, Southwest Fisheries Science Center
National Marine Fisheries Service, National Oceanic and Atmospheric Administration
8901 La Jolla Shores Drive
La Jolla, CA 92037 USA

William Rayment
Marine Science Department
University of Otago
P.O. Box 56
Dunedin
New Zealand

Phil Hammond
Sea Mammal Research Unit
Gatty Marine Laboratory
University of St. Andrews
St. Andrews KY16 8LB
Scotland

Corresponding author: Kim Urian
Correspondence address: Duke University Marine Laboratory
135 Duke Marine Lab Road
Beaufort, NC, 28516 USA
Telephone: (252) 504-7516
Fax: (252) 504-7648
E-mail: kim.urian@gmail.com

Abstract

Capture-recapture methods are frequently employed to estimate abundance of cetaceans using photographic techniques and a variety of statistical models. However, there are many unresolved issues regarding the selection and manipulation of images that can potentially impose bias on resulting estimates. To examine the potential impact of these issues we circulated a test data set of dorsal fin images from bottlenose dolphins to several independent research groups. Photo-identification methods were generally similar, but the selection, scoring, and matching of images varied greatly amongst groups. Based on these results we make the following recommendations. Researchers should: (1) determine the degree of marking, or level of distinctiveness, and use images of sufficient quality to recognize animals of that level of distinctiveness; (2) ensure that markings are sufficiently distinct to eliminate the potential for "twins" to occur; (3) stratify data sets by distinctiveness and generate a series of abundance estimates to investigate the influence of including animals of varying degrees of markings; and (4) strive to examine and incorporate variability among analysts into capture-recapture estimation. In this paper we summarize these potential sources of bias and provide recommendations for best practices for using natural markings in a capture-recapture framework.

Key words: capture-recapture, mark-recapture, photo-identification, abundance, population size estimates

Introduction

Natural markings have long been used to identify individual cetaceans. Initially, researchers used these markings to follow the movements of individually distinctive animals. For example, repeated sightings of a bottlenose dolphin (*Tursiops truncatus*) with a disfigured dorsal fin provided inferences into its home range (Caldwell 1955). Other researchers followed the movements of an individual humpback whale (*Megaptera novaeangliae*) with a distinctive dorsal fin and pigmentation pattern on the underside of its flukes (Schevill and Backus 1960). Photographs enhanced the ability of researchers to use natural markings (*i.e.*, photo-identification) to identify individual cetaceans. The approach was then extended to longitudinal photo-identification studies of cetacean populations beginning with killer whales (*Orcinus orca*) in the 1970s (Bigg 1982) and soon expanded to other species, including humpback whales (Katona and Whitehead 1981) and right whales (*Eubalaena glacialis*) (Payne 1986), bottlenose dolphins (Würsig and Würsig 1977, Wells and Scott 1999) and spinner dolphins (*Stenella longirostris*) (Norris *et al.* 1994). Depending on the species, various features are used to identify individuals, including: notch patterns in fluke edges, nicks and notches in the trailing edges of dorsal fins, the shape of dorsal ridges, pigmentation patterns, or callosity patterns and scars. Eventually, photo-identification methods were developed to obtain quantitative estimates of population parameters, such as abundance and survival (Hammond 1986). In 1988 the International Whaling Commission (IWC) held a workshop to review and standardize photographic techniques, sampling protocols, and analytical methods. The results of the workshop were published (Hammond *et al*. 1990).

In the late 1990s, the advent of affordable and durable digital cameras drastically changed photo-identification methods in both the field and laboratory (Markowitz *et al.* 2003, Mazzoil *et al*. 2004). For the first time, researchers were able to take large numbers of images without stopping to change film, manually focus, or modify camera settings. And, importantly, it was possible to review images in the field to determine which individuals had been captured with images of sufficient quality. In the laboratory, the use of digital photography eliminated the cost and time involved in developing film (Markowitz *et al*. 2003) and allowed manipulation of images to resolve fine features that might not have been visible using traditional techniques.

There have been parallel advances in statistical analyses over the past two decades, facilitating the development and fitting of a wider array of capture-recapture models to more realistically describe the processes underlying individual capture (Pollock *et al.* 1990, Pollock 2000). A large and increasing number of researchers are using photo-identification methods to derive estimates of abundance for cetaceans using statistical models implemented in computer programs such as MARK, POPAN, or CAPTURE (Otis *et al.* 1978, Arnason and Schwarz 1999, White and Burnham 1999) or using custom models tailored towards specific cetacean applications (*e.g*., Corkery *et al.* 2008; Durban *et al*. 2010; Conn *et al*. 2011; Fearnbach *et al*. 2012). However, there are many unresolved issues regarding data selection to match model assumptions that can potentially impose biases on resulting estimates. Here we argue that these issues and their effects on estimates of abundance are both important and under-appreciated. The 1988 IWC Workshop (Hammond *et al.* 1990) began a discourse regarding these issues and we take up the discussion here and develop a series of "best practices" to standardize laboratory methods and, we hope, improve future studies using photo-identification to estimate the abundance of cetacean populations.

The primary objective of this paper is to determine how best to select photo-identification data sets for use in capture-recapture analyses, with a particular focus on identifying potential sources of bias arising from practices used in the field and laboratory and on the development of methods to minimize these biases.

Capture-recapture statistical models used to estimate abundance require photo-identification data to conform to a specific set of assumptions in order to provide adequate model fit (Hammond 2009, Hammond 2010). With careful attention to experimental design, image selection, data analysis, and model choice, researchers can minimize the potential bias associated with violating these assumptions.

Three of the primary assumptions are related to the accuracy of the data themselves (the marked animals): 1) marks are unique, 2) read without error, and 3) do not change or are not lost. Two further assumptions of conventional capture-recapture models are related to the behavior of the animals and/or researchers and determine how representative the data are of the sample population: 4) capture probability is unaffected by marking and 5) is equal among individuals within a sampling occasion. A final assumption is relevant only to the analysis of closed populations, that is, 6) no births, deaths, permanent immigration, or permanent emigration occur between sampling occasions.

These assumptions and their effects on estimates of population size of terrestrial and marine mammals have been extensively reviewed elsewhere (Carothers 1973, Otis *et al.* 1978, Seber 1982, Begon 1983, Hammond 1986, Wilson *et al.* 1999, Chao 2001, Read *et al.* 2003, Amstrup *et al*. 2005). Here we address two other sets of concerns: best practices in the laboratory for evaluating image quality and distinctiveness so that marks are recorded correctly, and whether subsequent selection of data for analysis is representative so that bias is minimized in the resulting estimates.

To ensure that marks are read without error, it is first critical for a marked animal to be recognized with (near) 100% certainty if recaptured in an image of acceptable quality. Explicit in this definition is an interaction between image quality and the distinctiveness of features used to identify an individual. For example, the most distinct individuals may be identifiable in poor quality images, but individuals with subtle features may be recognizable only in higher quality images. Herein lies a critical analytical question: how best to balance accuracy (minimizing the violations of assumptions) with precision (largely a function of sample size) in estimation?

The goal of the present paper, therefore, is to provide recommendations for best practices in the selection of images and data used in photographic capture-recapture studies used to estimate the abundance of cetaceans.

Examination of variation in methods used across laboratories and researchers

We circulated a test data set of dorsal fin images of bottlenose dolphins to researchers who had considerable experience with photo-identification of this and other species. Each researcher provided the results of their photo-identification efforts and responded to questions regarding selection of images from the data set for capture-recapture analyses.

Experimental design

Each researcher evaluated and matched images independently, which allowed us to compare error rates and to estimate the magnitude of bias resulting from different data selection methods. All results were submitted anonymously. The test data set of images represented two separate dolphin encounters, each including a known number of individual animals. Each encounter comprised 50 images chosen by two experienced analysts from a catalog of known individuals. The images represented a range of quality and distinctiveness, contained in a 3x3 matrix of excellent, good, and poor quality images and well marked, moderately marked, and 'clean' (fins with no markings). Each participant applied their laboratory's photo-identification methods for the purpose of capture-recapture analysis. We also asked each researcher to provide a description of the criteria used to evaluate images and select marked individuals. Participants identified matches within each encounter and between the two encounters. This allowed us to assess the effects of image quality and levels of distinctiveness on recapture rates. We specifically asked participants to provide the following information from their evaluation of the data sets:

1. The number of images that were of sufficient image quality to be used in a capture-recapture analysis;
2. The number of images that were "unmarked" or insufficiently distinct to be used;
3. The number of unique dolphins in each encounter;
4. The number of matches within each encounter;
5. The number of matches between the two encounters.

Results

Eighteen participants from 12 research groups conducted our photo-identification experiment; some respondents were from the same laboratory, but submitted their results independently (Table 1). Thirteen participants provided their protocols for selection of images for photo-identification. In selecting images, every respondent assessed the following four sets of features: (1) focus/clarity/sharpness; (2) contrast/lighting/exposure; (3) angle of the dorsal fin to the photographer; and (4) whether the entire fin was visible.

Some research groups assigned quantitative values to each of these criteria to generate an image quality score; others assigned an overall quantitative or qualitative grade of image quality. Respondents used various features to identify and match individuals; all used permanent features with some additionally relying on temporary marks and lesions (Table 2). Generally, most evaluated the range of distinctiveness of an animal's features using a categorical scoring system: very distinct (high), moderately distinct (average), or not distinct (clean).

Overall, there was a surprisingly high degree of variation among responders in the results of the experiment (Fig. 1-5). For example, when participants were asked to report the number of unique distinctive individuals within an encounter (out of a total of 50 images), the responses ranged from 17 to 47 (Fig. 3). The parameters with the highest inter-individual variance were the evaluation of distinctiveness (CV=113% and 77% in encounters 1 and 2, respectively) and the number of matches within (CV=77% and 52%) and between encounters (CV=49%). These results underscored the need to review the criteria used for selecting photo-identification images to be used for capture-recapture analysis.

We calculated an abundance estimate for each data set using the Chapman modification of the Lincoln-Peterson estimator (Fig. 6). Not surprisingly, there was a high degree of variation in the resulting point estimates, although the confidence intervals did overlap, indicating that there were no statistical differences among estimates. It should be noted that this result was most likely due to a lack of power caused by the small number of recaptures made in the two sampling periods. Regardless, there was an unsettling degree of variation among researchers in the evaluation of image quality, distinctiveness, images selected, and matches. Participants from the same institution generally had similar results, suggesting that most variation was due to the different methods used by each laboratory. There was no apparent effect of the degree of experience with photo-identification. Some researchers selected images of relatively low quality to match, whereas others were much more selective and restricted their data set to high quality images of distinctive individuals.

We suggest that future studies using natural markings should address this potential source of heterogeneity by assigning scores of distinctiveness and image quality, and by exploring the potential effects of variation in these parameters on estimates of abundance. The exercise demonstrated that researchers in our field exhibit considerable variation in the methods used to select images and data for capture-recapture analyses with bottlenose dolphins. With the results of this exercise in mind, we focused discussion on practical issues of image and data selection for photographic capture-recapture analysis.

Photographic Quality, Individual Distinctiveness and Matching Criteria **(Assumptions 1, 2 & 3)**

To address how photo-identification images and marks are evaluated and matching criteria are used to select the sample of marked animals, and how bias may be introduced during this process, we addressed the following questions, which pertain to assumptions (1), (2), and (3) above. Which images should be retained and which should be discarded - should every image be evaluated in some quantitative fashion? Is it appropriate to manipulate or edit images? Is the scoring process replicable? Should there be some minimum quality standard for image selection? Is there a minimum threshold to consider an animal 'marked'? And, how can one ensure that all potential matches are identified?

Photographic Quality

Analysts are typically concerned with the presence (1) or absence (0) of individuals in the capture histories used to estimate abundance, but this process begins with selection of the images for inclusion in the analysis. The inclusion of poor quality images increases the risk of making incorrect matches or missing them altogether (Hammond 1986). False positive matches (recording two different animals as the same individual) introduce avoidable error, cause estimates of abundance to be negatively biased, and are difficult to detect in long-lived species, especially over a long time series (Gunnlaugsson and Sigurjonsson 1990, Yoshizaki *et al.* 2009). False negative matches (recording one animal as two or more by missing a match because of poor image quality or when individuals acquire new features) create "ghost histories" of individuals and result in positively biased estimates (Yoshizaki *et al.* 2009). It is well documented that error rates increase with decreasing photographic quality (Stevick *et al.* 2001, Friday *et al.* 2008, Frasier *et al.* 2009, Barlow *et al.* 2011). It is essential to define and implement a threshold for photographic quality in capture-recapture studies because it is assumed that every individual is recognized or identified correctly; if the effects of alternative thresholds are explored in analysis (see below) these should also be defined.

To understand which practices are used in our field, we reviewed 34 publications from 1999 to 2011 that employed photo-identification images to estimate abundance of cetaceans (see Appendix S3). The vast majority (*n*=31) assessed photographic quality; 21 of these studies employed a quality scale and nine used a binary scale (one was unclear on the criteria used). Most authors regarded focus or clarity as the most critical element of a good quality image (Table 3). Proper exposure and lighting and/or contrast were also considered to be important components of a good quality image, although with photo-management software it is possible to enhance these elements in a digital image. However, none of these papers mentioned digital manipulation or enhancement other than the cropping of images. Other features used to assess whether a photograph is of acceptable quality pertain more to the subject than the image: the angle of the subject to the photographer; whether all potentially distinguishing features of the animal are visible in the frame and not obscured by waves, water, adjacent animals, or barnacles; and distance to the subject or the size of the subject relative to the frame.

Among the 34 publications (Appendix S3), two of the most widely cited lists of criteria for scoring image quality were those of Wilson *et al.* (1999) for studies of bottlenose dolphins in the Moray Firth, Scotland and Urian *et al.* (1999) for bottlenose dolphins in the western North Atlantic. Wilson *et al.* (1999) graded images on a scale of 1 to 3 and used only grade 3 images that were well lit, in focus, free from spray and with fins parallel to the photographer with the dolphin's flank exposed. Urian *et al*. (1999) (updated in Urian *et al.* in press) used a more complex grading scheme, with five criteria scored independently for focus, angle, contrast, proportion of fin in frame, and full or partial fin in frame. These scores are weighted depending on their contribution to the overall quality of the images and the sum of the scores is used to describe overall photographic quality. If an image is deficient in any one of the criteria, the image is rejected. This system has worked well in trying to minimize subjectivity within and between laboratories, but is relatively time-consuming.

All types of analyses, including those of social association and home range, are essentially capture-recapture in which photographic sighting histories of individuals are accumulated over time, so that, regardless of the research question, low quality images should not be included in such analyses. The threshold of image quality should be clearly defined and described in any study. There is an understandable desire to standardize the evaluation of image quality across studies, but, in reality, the criteria used will vary from study to study and site to site. However, there is a clear need for researchers to be explicit about the quality and distinctiveness criteria used and to standardize reporting of these methods in the literature.

In some cases it is possible to apply simulation models to better capture some potential sources of bias in photographic data sets. If there is variation in methods amongst contributing researchers or laboratories, for example, it is possible to incorporate these differences as uncertainty in the resulting abundance estimate (*e.g.*, Barlow *et al.* 2011).

Transition from slides to digital media

In many long-term photo-identification studies, older records were derived from images on color slide film or black and white negatives. The transition to digital media introduced another potential source of bias. For example, Rayment (unpublished data) noted that scanned slides in digital catalogs are generally of lower quality than original digital images. Urian (unpublished data) found that the behavior of individual photographers changed between capture-recapture studies of bottlenose dolphins employing slide film in 2000 and digital photography in 2006. Photographers took fewer images of very distinctive dolphins when they were able to review digital images in the field, but took more images of very distinctive animals with slide film, perhaps to ensure that they 'captured' these individuals. There were also fewer poor quality images and more average quality digital images because of the ability to zoom in on features without losing resolution. A great advantage of digital technology is that higher resolution images can increase capture probability, thus increasing precision. An arguably equally important advantage is that increasing capture probability will tend to decrease any heterogeneity and thus also potentially decrease bias (if heterogeneity is otherwise unaccounted for) (Hammond 2010).

Manipulation of images – the use of photo editing tools to enhance photographs

We recommend that researchers report the file format of images used and note whether images are cropped or enhanced (*e.g.*, manipulating the brightness or contrast of an image). Some researchers have expressed concern with the use of JPEG image format because of issues associated with artifacts of JPEG compression from the RAW image file format, which may result in a loss of information (*e.g.*, Mizrock 2007). If the loss of image quality from conversion from RAW format to a JPEG compromises the matching process, then the marks being considered are probably too subtle. We recommend that the RAW image file be archived so that no features are lost or altered during manipulation of the original image.

Individual Distinctiveness

The use of natural markings differs from traditional capture-recapture studies, in which animals are physically caught using traps and are marked by researchers with unique tags. Therefore, the use of natural markings in a capture-recapture framework relies on the use of features that are distinct enough to eliminate the potential for "twins" to occur in the population. As noted above, false positive and false negative matches may be introduced not only by including images of poor quality in the matching process, but also by including animals with subtle or temporary mark types. Natural markings must be distinct enough to be reliably captured (and recaptured) in an image that meets the defined quality threshold. Herein is the problematic issue of the interplay between the quality of an image for photo-identification and the distinctiveness of the individual for matching; if there is a threshold for distinctiveness, should this threshold depend on image quality (Agler 1992; Friday *et al.* 2000, 2008; Read *et al.* 2003)?

The threshold used for distinctiveness often depends on the population being studied, such that subtle or temporary markings may be used with small populations in a limited range and within a short time period, whereas only very well-marked animals should be used with large populations that range across extensive areas and/or over a long time period. This issue is strongly related to capture probability; individuals from small local populations have higher probability of being captured and thus more subtle marks may be included. The choice of marks should be related to not only the study species, but also to the frequency of sampling periods and overall duration of the study, so it is valuable to have some knowledge of the range and relative size of the population being studied, as well as the intended frequency and overall time span of sampling.

In an ideal world, image quality and mark distinctiveness would be independent but, as was apparent in our exercise, in practice different standards or thresholds are applied. Some of our participants attempted to increase capture probability by including well-marked fins in poor quality images, but this will introduce, or increase, heterogeneity in capture probabilities. In addition, our respondents used a range of qualitative and quantitative descriptions to evaluate distinctiveness.

Some laboratories use separate analysts to score image quality and individual distinctiveness to help minimize any interplay between the two scores. The interplay between image quality and distinctiveness seems to be an inherent issue in the selection of photo-identification images (Friday *et al.* 2008).

One approach for selecting the criteria for considering whether an animal is categorized as "marked" or "captured" is to determine the degree of marking, or what level of distinctiveness will be used, and then decide on the image quality threshold necessary to recognize animals based on that level of distinctiveness. For example, if only very distinctive animals are included in the analysis, then lower photographic quality criteria may be used, but if subtle features are used to identify individuals, then only very high quality images should be included.

Matching criteria

To verify a match, most researchers require confirmation from an additional experienced researcher, and some laboratories require at least three judges to confirm a match. Instituting systematic protocols for the matching process minimizes errors in assigning false positive, but does not address the issue of missing matches (false negatives). And, although matches are typically confirmed by other researchers, few studies report how analysts check for unmatched individuals. It is possible to reduce matching error rate by using multiple analysts to search for potential matches or to include a measure of certainty or confidence associated with each match. If a match is very difficult to confirm or reject, the quality of the image or distinctiveness of the animal is likely to be insufficient. Additionally, if consensus is not reached among analysts, the potential match should be rejected. Hence, protocols are inherently averse to false positives, thereby increasing false negatives (Stevick *et al*. 2001).

An alternative to eliminating these data, and reducing statistical power, is to use variability in the assignments among individual analysts to generate estimates of the probability of a match. If data from multiple analysts can be built into an appropriate observation model for the identifications, then a state-space approach (*e.g*., Royle 2008) could be used to incorporate this key uncertainty into inference from a capture-recapture process model. We expect the development of such an approach in the near future.

Errors in matching can occur as a result of many issues inherent in the photo-identification process. This error rate may be a function of fatigue and catalog size; it is very time consuming to search manually through large digital catalogs. The number of comparisons can be reduced by subdividing catalogs into mark types; several software applications allow images to be organized based on features or quality such as *FinBase* (Adams *et al*. 2006) and computer-assisted matching programs, such as *Darwin*, *Finscan,* or *Fluke Matcher* (Wilkin *et al*. 1998, Hillman *et al*. 2003, Kniest *et al.* 2010), also assist in this regard. Computer matching programs can ease fatigue associated with working with large catalogs and help to minimize subjectivity in the matching process, although the analyst still makes the final decision regarding a match.

Recommendations

Image quality should be assessed prior to the matching process in capture-recapture studies and the criteria and thresholds used should be reported not only in the literature, but explored for their impacts on the final estimates. Researchers studying different populations will use practices best suited for their study species, but it is necessary to report these practices clearly. It is desirable to incorporate the effects of variation in grading images and matching into capture-recapture models. If it is possible to estimate the error rate (see Stevick *et al.* 2001, Barlow *et al.* 2011), then the population estimate can be adjusted accordingly; however, if this is not possible then only high quality images should be included in the analyses, (at the cost of a limited sample size), which will minimize bias but decrease precision. Variation clearly exists amongst analysts in this regard; such variability is not inherently bad, but it should be estimated and incorporated in the analysis. Variability among analysts should be examined and incorporated into observation models when using capture-recapture techniques to estimate abundance, but this will require development of new statistical procedures.

Researchers may employ simple or complex grading schemes to evaluate image quality. We recommend applying a simple grading system for large populations and data sets. Researchers should not feel compelled to adhere to any specific set of criteria, but they should report their methods clearly, preferably with examples. The specific criteria used will depend on the species and features used for individual identification. For example, focus and angle are critical for using notch patterns to identify individuals and contrast is not as important. On the other hand, it is essential to have images with good contrast when using pigmentation patterns to identify animals. Therefore, the criteria used do not need to be standardized across studies, but should be evaluated, reported, and replicable.

We recommend that thresholds of photographic quality (see above) should be determined by how well-marked an animal should be for capture-recapture studies. One approach is to set the mark level first, then set the threshold of image quality to ensure that animals with such markings will be recaptured in any image of this quality. Therefore, subtle features may be included if the image quality threshold is high. Potential bias through individual heterogeneity is introduced when including animals with very few markings, which may not be evident in images of lesser quality. When a data set is restricted to excellent quality images, the capture probability of individuals included in a capture-recapture analysis is reduced. If the image quality threshold is relaxed, and more individuals are included in the analysis, potential heterogeneity bias is introduced if less distinctive animals cannot be reliably identified in subsequent pictures of equal quality. A significant source of variation in photographic capture-recapture studies is due to re-scaling estimates of the marked population to arrive at an estimate of the total population (Durban *et al*. 2010; Eguchi *in press*, and see below). Not all animals have reliable marks and thus are not distinguishable, but these individuals need to be included in the estimate of total population size. We recommend that researchers stratify their data sets by distinctiveness ratings and generate a series of abundance estimates to investigate the influence of including animals of varying distinctiveness.

Permanence of Marking and Mark Evolution **(Assumptions 2 & 3)**

Relatively few studies have addressed the issue of how marks change over time with cetaceans in a quantitative manner. This relates to the issue of evolving marks as it pertains to assumptions (2) and (3), specifically the following questions: what is the rate of change of markings over time and how can this rate be estimated? Does this rate vary by species or population and how might evolution of marks affect estimates of abundance? To address this question we examined the results of several long-term studies that evaluated mark evolution.

Sperm whales (*Physeter macrocephalus)*, which are identified by markings along the trailing edge of the flukes, had a 1.3% probability of mark change each year (Dufault and Whitehead 1995). Wilson *et al.* (1999) assessed mark permanence for bottlenose dolphins in the Moray Firth, Scotland over a three-year period. Nicks and notches on the dorsal fin were relatively stable, but scratches and skin disorders faded or disappeared over the course of the study. Gowans and Whitehead (2001) conducted a nine-year study on mark permanence of northern bottlenose whales (*Hyperoodon ampullatus)*. This study identified back indentations, mottled patches, or dorsal fin notches as the most appropriate long-term markings for individual identification, with no loss of these marks and up to a 2% gain rate per year. Aschettino *et al.* (2011) showed that mark changes in melon-headed whales (*Peponocephala electra*) in Hawai'i occurred once every 9.2-13.8 yr. False killer whales (*Pseudorca crassidens*) marks were changed once every 6.9 to 8.8 yr in Hawaii (Baird *et al*. 2008). Auger-Méthé and Whitehead (2007) calculated the rates of acquisition for each mark type in a photo-identification study of long-finned pilot whales (*Globicephala melas)*. Dorsal fin markings were determined to be the most permanent mark type, but only one-third of the animals had markings that were distinctive enough to be used for long-term identification, suggesting that additional mark types such as scarring and saddle patches might help to increase the number of identifiable individuals in a population.

Overall, the results of these studies indicate that permanent notches of the dorsal fin or flukes and persistent pigmentation patterns (*e.g*., blue whales (*Balaenoptera musculus*); Ramp *et al.* 2006) are the most appropriate mark types for long-term identification. However, each study population experiences different ecological circumstances (including anthropogenic influences and degree of predation pressure) that may lead to marks, so acquisition rates will vary from one population to the next.

The community of bottlenose dolphins in Sarasota Bay, Florida has been studied for over 40 yr and approximately 96% of the dolphins are identifiable (Wells 2003, Wells 2013). This community, therefore, provides a model case study to assess mark acquisition rates. Seventy-seven dolphin calves were monitored using photo-identification methods to identify rates of mark acquisition from 2004-2011 (calves were added to the study throughout this time period, so all calves born in 2004 were followed, in addition to calves born in subsequent years). At the end of the seven-year study, each individual was grouped into one of four mark acquisition phases:

Not distinctive (DN): no information content in pattern, markings, and leading or trailing edge features.

Marginally distinctive (DM): very little information content in pattern, markings, and leading or trailing edge features.

Moderately distinctive (D2): two features or one major feature on dorsal fin.

Very distinctive (D1): multiple major features on dorsal fin.

The mean number of days for an individual to move from *Not Distinctive* to any of the other three mark acquisition phases was determined. Of the 77 dolphins monitored, 57% remained *Not Distinctive*, 23% were *Marginally Distinctive*, 16% were *Moderately Distinctive*, and 4% were *Very Distinctive* at the end of the seven-year study. The mean number of days for an individual to become *Marginally Distinctive*, *Moderately Distinctive*, and *Very Distinctive* was 477 ± 347 SD, 752 ± 480 SD, and 613 ± 582 SD, respectively. Thus, there was a high level of individual variation in mark acquisition for bottlenose dolphins in Sarasota Bay. Current research in Sarasota is focused on examining the ontogeny of fin features over time and quantifying significant changes in dorsal fin markings that could result in misidentification of a given individual. The long-term research program in Sarasota Bay provides an excellent opportunity to assess fin changes over time and to measure differences in markings associated with age-sex demographics. This analysis should be compared to other long-term studies to determine the differences in mark acquisition rates among populations.

Recommendations

We conclude that the rate of mark acquisition is not likely to be an issue with small populations studied over short time periods. However, when researchers estimate abundance for large, open populations, and particularly when with survey effort is conducted over longer time intervals, they should make an effort to estimate mark acquisition rates. Notches on the dorsal fin and flukes are long-term, if not permanent; survey effort over long time periods may increase the likelihood of committing identification errors as marks are acquired over time. Future research is necessary to link mark acquisition rates to an appropriate survey methodology that limits errors in photo-identification. In particular, researchers should estimate the rate of mark change or measure the duration of marks when using temporary marks, such as skin lesions (Wilson *et al.* 1997, Wilson *et al*. 1999). As a practical matter, researchers should endeavor to use markings that change as little as possible, monitor mark evolution, and estimate the rate of mark loss or change.

***Behavior of Unmarked Animals* (Assumptions 4 & 5)**

The assumptions of conventional models that the capture probability is unaffected by the “marking” or photographing (assumption 4) and that catchability is homogeneous (assumption 5) may also be violated. Most researchers assume that the behavior of the marked animals they capture in photographic images is representative of the population, but few studies have tested this assumption. This issue becomes particularly important as the proportion of marked individuals in a sample decreases. Are distinctive individuals really representative of the entire population? How should this assumption be tested? Two potential sampling effects may result in a violation of this assumption: an 'animal' effect and an 'observer' effect. We consider both types of effect below.

The primary issue that may contribute toward the 'animal' sampling effect occurs when animals are distinctive, but are not encountered or available to be photographed. The influence of the platform used to approach animals for photographic capture may have an effect on sampling. For example, some animals may be more timid around survey vessels, whereas other animals may be attracted to them. This kind of behavioral response may contribute to the situation in which animals are individually identifiable but not captured - is the animal sensitive to the sampling method and how is this potential bias assessed? Sampling methods may be adjusted to address individuals that avoid boats by employing quiet vessels or alternative platforms to determine whether the vessel is influencing the behavior of the study animal. For animals that are evasive, another option is to increase the focal length of the camera lens to photograph animals from a greater distance. By applying alternative methods suited for the study animal, the avoidance behavior of some animals can be mitigated to some degree.

This 'animal' effect is likely to vary among species and among populations due to local factors, such as the presence of other boats, the occurrence of predators, and/or habitat type. If marks are obtained from anthropogenic impacts (*e*.g., boat strike), it is possible that marked animals may be more wary of boats and thus less available to be photographed. To ensure that all animals (marked and unmarked) are photographed and that the behavior of unmarked animals is accounted for, it is best to sample animals as uniformly as possible; photographs should be taken of all animals, regardless of how well marked the individual is or whether an individual has already been photographed (Eguchi 2003). Also, increasing capture probability by increasing the study area for animals with large ranges and intensifying sampling effort will help to detect more individuals and decrease this bias.

For some cetacean species, specifically those identified by notch patterns on fins or fluke edges, individuals become marked as a function of age (see above). Most calves, for example, typically do not have identifying marks, and are not normally included in capture-recapture analyses; younger animals may be included at an earlier age only when using excellent quality images to identify small or subtle features. Researchers should report whether they include calves in their sample, (and clearly define the category "calf"), as calves are usually closely associated with their mothers and thus are not mixed at random in the population (Rosel *et al.* 2011). In many species, males acquire marks earlier in life than females (Tolley *et al*. 1995, Wilson 1997), which may introduce a sex bias in estimates, although this may vary from population to population, and species to species.

There are several possible ways to test for differences in the behavior of marked and unmarked animals. Tags can be applied to marked and unmarked animals to compare behavioral responses to survey vessels. However, the application of tags may change or influence the behavior of the animal, confounding individual variation in behavior. Behavioral responses to tagging may be mitigated by tagging the animal remotely instead of capturing the animal to apply the tag, allowing sufficient time between tagging and data collection (Elwen *et al.* 2006), and ensuring that both marked and unmarked animals are tagged. A noninvasive method to compare the behavior or catchability of marked and unmarked animals is to use marks not typically used in capture-recapture analyses. Auger-Méthé and Whitehead (2007) used this approach for long-finned pilot whales in Nova Scotia, Canada. They used 15 mark types, such as scrapes, saddle patches, eye blazes, scars, tooth rakes to determine whether these temporary marks could improve identification. The study showed that the proportion of the population that was identifiable did not differ from the rest of the population in its susceptibility to factors causing marks, such as predation, and was representative of the whole population. The potential for this source of bias should be evaluated in other species.

There are two potential sources of 'observer' effect: (1) field sampling may vary among photographers and (2) the criteria used to determine which animals are marked or unmarked in the laboratory may be subjective, resulting in misidentifications. Some photographers may be more skilled at capturing animals with an unbiased approach in the field. It would be useful to examine the process of photographic capture in the field and determine how this may influence the resulting photo-identification images. It would be particularly interesting to examine the effects of group size, behavior, and survey conditions on the quality and number of images obtained. Despite the use of criteria for the selection of images and the evaluation of distinctiveness, subjectivity may be introduced if more than one observer is involved in the identification and capture-recapture analysis, as was clear from our photo-identification exercise.

There is also a potential interaction between the animal effect and the observer effect. Inexperienced photographers may under-represent classes/individuals that are more difficult to photograph (*e.g.,* calves), and thus there may be fewer of such animals within their samples (or in the extreme case, some classes/individuals will be effectively unavailable). However, this could be investigated by generating estimates using different data samples, for example exploring the effect of restricting the data set to photographs from experienced photographers.

Recommendations

By applying alternative methods (*e.g.,* different survey platforms), the avoidance behavior of some animals can be mitigated to some degree. To address the issues of animals that are not observed or photographed and animals that are observed or photographed, but are not distinctive or marked, it is important that researchers attempt to photograph all individuals in an encounter, marked or unmarked. Complete, unbiased photographic coverage of a group is recommended, but if that is not possible, then they should be sure to take photographs of a random sample of individuals in the encounter.

Estimating Proportion of Marked Individuals in the Population

Another potential source of variation and bias arises from ways in which the proportion of marked animals is estimated and used to scale the estimate of abundance to include animals that lack marks. We address the following questions and provide a new method for estimating this proportion. How is the proportion of marked individuals in an encounter estimated? How are unmarked individuals accounted for in the estimate?

In many species of cetaceans, most of the population is naturally marked. For example, right whale callosity patterns (Payne 1986), blue whale pigmentation (Sears *et al.* 1990) and humpback whale flukes (Katona *et al.* 1979) are sufficiently different such that most individuals can be uniquely identified. However, the proportion marked is much lower in some species, such as Atlantic white-sided dolphins (*Lagenorhynchus acutus*), which typically possess nondistinctive dorsal fins (Weinrich *et al.* 2001). As noted above, unmarked animals may include calves and juveniles that have not yet developed distinctive marks. In situations where the proportion marked is less than 100%, an estimate of the proportion of marked animals is required in order to estimate the total abundance from the estimated number of marked animals.

When group sizes are small, the proportion of unmarked individuals can be determined in the field (*e.g.,* Williams *et al.* 1993). In other species and areas, the proportion of marked and unmarked individuals needs to be estimated to generate an estimate of abundance. This can be accomplished by analysis of good quality photographs.

For example, the population size (Ñ) of bottlenose dolphins in Doubtful Sound, New Zealand was estimated by the number of marked individuals in the population $\hat{N}$ and data on the proportion of marked individuals in the population.

$$\tilde{N} = \frac{\hat{N}}{1-Q} \text{ and } var\,(\tilde{N}) = \tilde{N}^2 \left(\frac{var\,(\hat{N})}{\hat{N}^2} + \frac{1-P}{nP}\right),$$

where $\hat{N}$ is the estimated abundance of marked individuals, *Q* is the proportion of photographs containing unidentified individuals (*P*= 1 - *Q*; proportion of photographs containing identified individuals) and *n* is the total number of photographs from which *P* was computed (Williams *et al.* 1993). However, this variance term does not include sampling error related to the estimated proportion of marked individuals in the population and therefore underestimates the total variance. To include this, the term (1-*P*)/*nP* should be replaced with.

$$\frac{var(\hat{P})}{\hat{P}^2}$$

These authors attempted to photograph all individuals present, whether they were identifiable or not. It may be difficult to use this method with large groups, because it is not possible to determine whether or not all the animals in the group were photographed. Wilson *et al.* (1999) also used this approach and estimated $\tilde{N}$ from the proportion of individuals encountered by using subtle skin markings to identify all individuals using high quality photographs.

More specific analytical approaches to this problem are currently in development. For example, Eguchi (in press) proposes a sampling and analytical process that can estimate the proportion of identifiable individuals in a population from photo-identification data. The proposed statistical models require a simple random photographic sampling of animals, where the photographic captures are treated as sampling with replacement within each group. The total number of images, including those that cannot be identified, and the number of images that contain identifiable individuals are used to make inferences about the proportion of identifiable individuals. When multiple groups are sampled, the population level proportion of identifiable individuals is estimated from the group estimates. Further, the number of images of each individual within each group is used to make inference about the group size. Combined with capture -recapture models and appropriate sampling protocols, abundance estimates of the total population and their uncertainty can be obtained.

Choosing appropriate mark-recapture models: matching the sampling design to model choices and assumptions

Using photographic documentation of natural markings is an unconventional application of capture-recapture methods, so we need to think unconventionally about how to analyze the data. Choices made during data selection may induce heterogeneous capture probabilities. For example, if identifications of well-marked individuals are used from lower quality photographs that are not usable for all individuals, this will result in biased estimates using conventional mark-recapture models that assume equal capture probabilities (Otis *et al.* 1978). However, heterogeneity is even more likely, and in reality unavoidable, due to the challenges of sampling mobile individuals in the marine environment. Rather than controlling the capture process, for example through the use of trapping grids, cetacean researchers are generally faced with the problem of sampling individuals with heterogeneous ranging patterns and behavioral responses to the survey vessel, with the effective coverage of photographic samples varying over time due to both changes in survey conditions and animal behavior. These sources of variability simply cannot be adequately controlled in the capture process, and models that allow for both temporal and individual variation in capture probability are typically required (*e.g.*, Wilson *et al.* 1999, Read *et al*. 2003).

More recently, advances in statistical models and computing also allow the fitting of models that describe more "realistically complex" capture processes. For example, mixture models can be used to describe clustered heterogeneity (Whitehead and Wimmer 2005, Durban *et al.* 2010) that may result from animals having similar capture probabilities within relatively stable social groupings, with greater variance among clusters. A further example is the use of hierarchical models to describe either positive or negative covariance between repeat surveys in terms of which individuals they captured; this may occur when certain surveys are more or less likely to capture certain individuals because they are unevenly distributed in either time, space or both (*e.g.*, Durban *et al.* 2005, Durban *et al.* 2010). Such dependencies between survey samples can arise particularly when using opportunistic photographic samples, rather than data collected solely for the purpose of photographic capture-recapture sampling, and these modern capture-recapture models offer the ability to relax the assumption of independent or random sampling.

Consideration of the spatial context of sampling is also very important, because the ranges of individual cetaceans often extend beyond small study areas (Durban *et al.* 2005). This mobility can result in heterogeneity in ranging patterns (*e.g*., Lusseau *et al*. 2006), while temporary emigration beyond the study area (Whitehead 1990, Durban *et al.* 2000*a*) and the presence of "transient" individuals among local or "resident" populations (Conn *et al.* 2011) creates uncertainty over population definition. When estimation of abundance is the research focus, temporary emigration serves to decrease capture probability (Kendall and Nichols 2002) that may change as a function of time (Hammond 1990). When the area is consistently used by at least a subset of individuals, it may be possible to model this structured heterogeneity with mixture models to classify and monitor a distinct local population cluster (Conn *et al.* 2011, Fearnbach *et al.* 2012). In this case, it is important to be explicit and consistent about the spatial extent of sampling for consistent population definition.

Many of these recent developments in capture-recapture modeling have been aided by advances in statistical computation. There is increasing use of the program MARK (White and Burnham, 1999) for application of a suite of mark-recapture models to cetacean data sets, and WinBUGS (Lunn *et al.* 2000) has enabled researchers to more easily fit Bayesian hierarchical models using Markov chain Monte Carlo (MCMC) sampling methods where analytic solutions are intractable. Bayesian inference based on full probability distributions is increasingly advocated as appropriate for quantifying and communicating uncertainty in ecological data analysis (Durban *et al*. 2000*b*, Wade 2000).

This utility extends to model selection, allowing inference to be based on a weighted average of candidate models simply by sampling across a mixture of competing models in the same MCMC fitting procedure (*e.g*., Durban *et al.* 2005, King *et al*. 2010), thus incorporating model selection uncertainty into the final probability distribution for abundance. This is important in unconventional situations when it has not been possible to control the capture process to fit one particular model, but it is a poor substitute for careful sample design that controls and maximizes capture probabilities to allow more precise inference. Once the best model(s) has been selected, it remains important to check the adequacy of model fit, but this is a component of inference that is often overlooked. Posterior predictive checks offer a very flexible approach for assessing model fit within a Bayesian framework: by predicting data from the model to compare to the real data this approach allows for the checking of overall model fit (*e.g.*, Durban *et al*. 2010) in addition to specific structural aspects of a model such as the differential fit to the capture histories of individuals (Fearnbach *et al*. 2012).

Capture-recapture analysis: the importance of good practice in the field and laboratory

The main goal in capture-recapture studies is to minimize bias and maximize precision; typically a compromise exists between these two desiderata. As software programs facilitate the ease with which an increasing array of capture-recapture models can be applied to photographic data, field researchers need to be increasingly vigilant in their choice of data acquisition and selection methods to ensure robust inference. Although recent advances in analytical methods can help overcome some of the unavoidable sources of heterogeneity, this does not mean that researchers can ignore the potential for bias. Instead, we encourage them to try to evaluate the bias-precision tradeoffs associated with data collection and processing.

In summary, we recommend that researchers using photo-identification methods to estimate abundance of cetaceans should address potential sources of heterogeneity by assigning scores of distinctiveness and image quality and explore the potential effects of variation in these parameters on abundance estimates and on the selection and fit of capture-recapture models. The results of our photo-identification exercise demonstrated that researchers in our field exhibit considerable variation in the methods used to select images and data for capture-recapture analyses, and we underscore a previous recommendation that variability among analysts be incorporated into observation and capture-recapture models (Barlow *et al.* 2011).

We recommend that image quality be assessed prior to the matching process in capture-recapture studies and that relevant criteria and thresholds used should be reported. This is particularly important because of recent advances in digital media which have allowed researchers to obtain large numbers of high resolution images that can be easily manipulated and enhanced. Researchers should stratify their data sets by distinctiveness ratings and investigate the influence on abundance estimates of including animals of varying distinctiveness. As noted above, researchers should also endeavor to use markings that change as little as possible, monitor mark evolution, and estimate the rate of mark loss or change, particularly for studies that span long time periods. The criteria used by researchers for photographic capture-recapture analysis do not need to be standardized across species, but should be evaluated and reported in the literature. It is good practice to first decide on the mark(s) or features that are deemed to be distinctive in each case study, and then decide on the level of image quality necessary to reliably document these marks.

In the field, researchers should strive to photograph all individuals in an encounter, whether they are marked or unmarked, or at a minimum, to photograph a representative sample of individuals present. This will help to minimize the introduction of bias caused by animals that are "trap happy" or particularly well-marked or "trap shy" or less well-marked. Analyses should investigate possible "photographer" effects by stratifying data by the experience level of the photographer and investigating the sensitivity of abundance estimates to data choices.

Heterogeneity is inherent in photo-identification data, some of which can be minimized in the sampling design, in the field, during the analytical process and, finally, in model selection. There is now a wide array of mark-recapture modeling tools available, ranging from conventional models that can be implemented using standard software to hierarchical models that can be tailored to specific applications. Model selection uncertainty should be quantified where possible, especially when photo-identification data have not been collected by design to suit a specific capture-recapture model. Where data allow, models should be fitted that describe the capture process as realistically as possible, and the adequacy of model fit should always be examined.

The tools of photographic capture-recapture have changed markedly since the IWC workshop was held twenty-five years ago, but the underlying applications of data obtained by these tools remain unchanged. We hope that the recommendations outlined in this paper will allow researchers to use these tools to minimize sources of bias and variation in estimates of abundance and other population parameters.

ACKNOWLEDGEMENTS

We thank all the participants in the photo-identification exercise, and our colleagues too numerous to list here, who provided recommendations and guidance for the evolution of this manuscript. We thank Keith Mullin, Patty Rosel and Lori Schwacke who provided insight and suggestions that led to the genesis of this paper. Danielle Waples and Reny Tyson recorded our discussions and we thank them for their careful edits of this manuscript. Heather Foley and Zach Swaim helped design and test the photo-identification experiment and Dave Johnston provided input on the examination of the results. The efforts of the staff, students, and volunteers of the Sarasota Dolphin Research Program over the decades of photographic surveys is much appreciated, especially the work of Jason Allen toward providing the data for the case study in this paper. W.R. would like to thank Steve Dawson, Trudi Webster, Liz Slooten and Marta Guerra (University of Otago) for discussions regarding photo-identification capture-recapture.

LITERATURE CITED

Adams, J., T. Speakman, E. Zolman and L. H. Schwacke. 2006. Automating image matching, cataloging, and analysis for photo-identification research. Aquatic Mammals 32:374-384.

Agler, B. A. 1992. Testing the reliability of photographic identification of individual fin whales (*Balaenoptera physalus*). Report of the International Whaling Commission 42:731-737.

Amstrup, S. C., T. L. McDonald and B. F. J. Manly. 2005. Handbook of capture-recapture analysis. Princeton University Press, Princeton, NJ.

Arnason, A.N. and C.J. Schwarz. 1999. POPAN-5: Using POPAN-5 to analyze banding data. Bird Study 46:S157-S168.

Aschettino, J.M., R.W. Baird, D.J. McSweeney, D.L. Webster, G.S. Schorr, J.L. Huggins, K.K. Martien, S.D. Mahaffy, and K.L. West. 2011. Population structure of melon-headed whales (*Peponocephala electra*) in the Hawaiian Archipelago: evidence of multiple populations based on photo-identification. Marine Mammal Science doi: 10.1111/j.1748-7692.2011.00517.x

Auger-Méthé, M., and H. Whitehead. 2007. The use of natural markings in studies of long-finned pilot whale (*Globicephala melas*). Marine Mammal Science 23:77–93.

Baird, R.W., D.L. Webster, S.D. Mahaffy, D.J. McSweeney, G.S. Schorr, and A.D. Ligon. 2008. Site fidelity and association patterns in a deep-water dolphin: rough-toothed dolphins (*Steno bredanensis*) in the Hawaiian Archipelago. Marine Mammal Science 24:535-553.

Barlow, J., J. Calambokidis, E. A. Falcone, *et al*. 2011. Humpback whale abundance in the North Pacific estimated by photographic capture-recapture with bias correction from simulation studies. Marine Mammal Science 27:793-818.

Begon, M. 1983. Abuses of mathematical techniques in ecology: applications of Jolly's capture-recapture method. Oikos 40:55-158.

Bigg, M. 1982. An assessment of killer whale (*Orcinus orca*) stocks of Vancouver Island, British Columbia. Report of the International Whaling Commission 32:655-666.

Caldwell, D. K. 1955. Evidence of home range of an Atlantic bottlenose dolphin. Journal of Mammalogy 36:304-305.

Carothers, A. D. 1973. The effects of unequal catchability on Jolly-Seber estimates. Biometrics 29:79-100.

Chao, A. 2001. An overview of closed capture-recapture models. Journal of Agricultural, Biological, and Environmental Statistics 6:158-175.

Conn, P. B., A. M. Gorgone, A. R. Jugovich, B. L. Byrd and L. J. Hansen. 2011. Accounting for transients when estimating abundance: A case study of bottlenose dolphins in Choctawhatchee Bay, Florida. Journal of Wildlife Management 75:569-579.

Corkrey, R., S. Brooks, D. Lusseau, K. Parsons, J. W. Durban, P. S. Hammond and P. M. Thompson. 2008. A Bayesian capture–recapture population model with simultaneous estimation of heterogeneity. Journal of the American Statistical Association 103:948–960.

Dufault, S. and H. Whitehead. 1995. An assessment of changes with time in the marking patterns used for photo-identification of individual sperm whales, *Physeter macrocephalus*. Marine Mammal Science 11:335-343.

Durban, J. W., K. M. Parsons, D. E. Claridge and K. C. Balcomb. 2000*a*. Quantifying dolphin occupancy patterns. Marine Mammal Science 16:825-828.

Durban, J. W., P. M. Thompson, D. A. Elston and X. Lambin. 2000*b*. A role for Bayesian inference in cetacean population assessment and management decisions. Journal for Cetacean Management and Conservation 2:117-123.

Durban, J. W., D. A. Elston, D. K. Ellifrit, E. Dickson, P. S. Hammond, and P. M. Thompson. 2005. Multi-site mark–recapture for cetaceans: population estimates with Bayesian model averaging. Marine Mammal Science 21:80–92.

Durban, J. W., D. Ellifrit, M. Dahlheim, *et al*. 2010. Photographic mark-recapture analysis of clustered mammal-eating killer whales around the Aleutian Islands and Gulf of Alaska. Marine Biology 157:1591-1604.

Eguchi, T. 2003. A hierarchical Bayes approach to capture-recapture abundance estimation. Ph.D. Dissertation. Montana State University. 201 pp.

Eguchi, T. In press. Estimating the proportion of identifiable individuals and group sizes in photographic identification studies. Marine Mammal Science.

Elwen, S., M. A. Meÿer, P. B. Best, P. G. H. Kotze, M. Thornton and S. Swanson. 2006. Range and movements of female Heaviside’s dolphins (*Cephalorhynchus heavisidii*), as determined by satellite-linked telemetry. Journal of Mammalogy 87:866–877.

Fearnbach, H., J. W. Durban, K. M. Parsons and D. E. Claridge. 2012. Photographic mark-recapture analysis of local dynamics within an open population of dolphins. Ecological Applications 22:1689–1700.

Frasier, T. R., P. K. Hamilton, M. W. Brown, S. D. Kraus and B. N. White. 2009. Sources and rates of errors in methods of individual identification in a comprehensive longterm wildlife study: The North Atlantic right whale (*Eubalaena glacialis*). Journal of Mammalogy 90:1246–1255.

Friday, N., T. D. Smith, P. T. Stevick and J. Allen. 2000. Measurement of photographic quality and individual distinctiveness for the photographic identification of humpback whales, *Megaptera novaeangliae*. Marine Mammal Science 16:355-374.

Friday, N., T. D. Smith, P. T. Stevick, J. Allen and T. Fernald. 2008. Balancing bias and precision in capture-recapture estimates of abundance. Marine Mammal Science 24:253-275.

Gowans, S. and H. Whitehead. 2001. Photographic identification of northern bottlenose whales (*Hyperoodon ampullatus*): sources of heterogeneity from natural marks. Marine Mammal Science 17:76-93.

Gunnlaugsson, T., and J. Sigurjonsson. 1990. A note on the problem of false positives in the use of natural markings for abundance estimation. Report of the International Whaling Commission (Special Issue 12:143-145.

Hammond, P. S. 1986. Estimating the size of naturally marked whale populations using capture-recapture techniques. Report of the International Whaling Commission (Special Issue 8):253−282.

Hammond, P.S. 1990. Heterogeneity in the Gulf of Maine? Estimating humpback whale population size when capture probabilities are not equal. Report of the International Whaling Commission (Special Issue 12):135-139.

Hammond, P. S., S. A. Mizroch and G. P. Donovan. 1990. Individual recognition of cetaceans: use of photo-indentification and other techniques to estimate population parameters. Report of the International Whaling Commission (Special Issue 12). Cambridge. UK.

Hammond, P. S. 2009. Mark-recapture. Pages 705-709 *in* W.F. Perrin, B. Würsig and J.G.M. Thewissen eds. Encyclopedia of marine mammals, second edition. Elsevier, Canada.

Hammond, P. S. 2010. Estimating the abundance of marine mammals. Pages 42-67 *in* I. L. Boyd, W. D. Bowen, S. J. Iverson eds. Marine mammal ecology and conservation: A handbook of techniques. Oxford University Press NY.

Hillman, G. R., B. Würsig, G. A. Gailey, N. Kehtarnavaz, *et al*. 2003. Computer-assisted photoidentification of individual marine vertebrates: A multi-species system. Aquatic Mammals 29:117–123.

Katona, S., B. Baxter, O. Brazier, S. D. Kraus, J. Perkins and H. Whitehead. 1979. Identification of humpback whales by fluke photographs. Behavior of marine animals: Current perspectives in research, Vol. 3: Cetaceans. Plenum Press, New York, London.

Katona, S. K., and H. P. Whitehead. 1981. Identifying humpback whales using their natural markings. Polar Record 20:439-444.

Kendall, W. L., and J. D. Nichols. 2002. Estimating state-transition probabilities for unobservable states using capture–recapture/resighting data. Ecology 83:3276–3284.

King, R., B.J.T. Morgan, O. Gimenez and S.P. Brooks. 2010. Bayesian analysis for population ecology. Chapman & Hall/CRC Press, Boca Raton, FL.

Kniest, E., D. Burns and P. Harrison. 2010. *Fluke Matcher*: a computer-aided matching system for humpback whale (*Megaptera novaeangliae*) flukes. Marine Mammal Science 26:744-756.

Lunn, D. J., A. Thomas, N. Best and D. Spiegelhalter. 2000. WinBUGS—a Bayesian modelling framework: concepts, structure and extensibility. Statistics and Computing 10:325–337.

Lusseau, D., B. Wilson, P. S. Hammond, *et al*. 2006. Quantifying the influence of sociality on population structure in bottlenose dolphins. Journal of Animal Ecology 75:14–24.

Markowitz, T. M., A. D. Harlin and B. Würsig. 2003. Digital photography improves efficiency of individual dolphin identification. Marine Mammal Science 19:217-223.

Mazzoil, M., S. D. McCulloch, R. H. Defran and M. E. Murdoch. 2004. Use of digital photography and analysis of dorsal fins for photo-identification of bottlenose dolphins. Aquatic Mammals 30:209-219.

Mizroch, S. 2007. NMML Digital photo protocol. http://www.afsc.noaa.gov/nmml/pdf/NMMLDigitalPhotoProtocol.pdf

Norris, K. S., B. Würsig, R. S. Wells and M. Würsig. 1994. The Hawaiian spinner dolphin. University of California Press, Berkeley, CA.

Otis, D. L., K. P. Burnham, G. C. White and D. R. Anderson. 1978. Statistical inference from capture data on closed animal populations. Wildlife Monographs 62:1-135.

Payne, R. 1986. Long term behavioral studies of the southern right whale (*Eubalaena australis*). Report of the International Whaling Commission (Special Issue 10):161-167.

Pollock, K. H., J. D. Nichols, C. Brownie and J. E. Hines. 1990. Statistical inference for capture-recapture experiments. Wildlife Monographs 107.

Pollock, K. H. 2000. Capture-recapture models. Journal of the American Statistical Association 95:293-296.

Ramp, C., M. Berube, W. Hagen and R. Sears. 2006. Survival of adult blue whales *Balaenoptera musculus*, in the Gulf of St. Lawrence, Canada. Marine Ecology Progress Series 319:287-295.

Read, A. J., K. W. Urian, B. Wilson and D. M. Waples. 2003. Abundance of bottlenose dolphins in the bays, sounds and estuaries of North Carolina, USA. Marine Mammal Science 19:59-73.

Rosel, P. E., K. D. Mullin, L. Garrison, *et al.* 2011. Photo-identification Capture-Mark-Recapture Techniques for Estimating Abundance of Bay, Sound and Estuary Populations of Bottlenose Dolphins along the U.S. East Coast and Gulf of Mexico: A Workshop Report. NOAA Technical Memorandum NMFS-SEFSC-621. 30 pp.

Royle, J.A. 2008. Modeling individual effects in the Cormack-Jolly-Seber model: a state-space formulation. Biometrics 64:364-370.

Schevill,W. E. and R. H. Backus. 1960. Daily patrol of a Megaptera. Journal of Mammalogy 41:279-281.

Sears, R., M. J. Williamson, F. W. Wenzel, M. Bérubé, D. Gendron and P. Jones. 1990. Photographic identification of the blue whale (*Balaenoptera musculus*) in the Gulf of St Lawrence, Canada. Report of the International Whaling Commission (Special Issue 12):335–342.

Seber, G. A. F. 1982. Estimation of animal abundance and related parameters. Second edition. Macmillan, New York, NY.

Stevick, P. T., P. J. Palsbøll, T. D. Smith, M. V. Bravington and P. S. Hammond. 2001. Errors in identification using natural markings: rates, sources and effects on capture-recapture estimates of abundance. Canadian Journal of Fisheries and Aquatic Sciences 58:1861-70.

Tolley, K. A., A. J. Read, R. S. Wells, K. W. Urian, M. D. Scott, A. B. Irvine and A. A. Hohn. 1995. Sexual dimorphism in wild bottlenose dolphins (*Tursiops truncatus*) from Sarasota, Florida. Journal of Mammalogy 76:1190-1198.

Urian, K. W., A. A. Hohn and L. J. Hansen. 1999. Status of the photo-identification catalog of coastal bottlenose dolphins of the western North Atlantic: report of a workshop of catalog contributors. NOAA Technical Memorandum NMFS-SEFSC-425. 22 pp.

Urian, K. W., D. M. Waples, R.B. Tyson, L. E.W. Hodge, and A. J. Read. 2014. Abundance of bottlenose dolphins (*Tursiops truncatus*) in estuarine and near-shore waters of North Carolina, USA. Journal of the North Carolina Academy of Science.

Wade, P. R. 2000. Bayesian methods in conservation biology. Conservation Biology 14:1308–1316.

Weinrich, M. T., C. R. Belt and D. Morin. 2001. Behavior and ecology of the Atlantic white-sided dolphin (*Lagenorhynchus acutus*) in coastal New England waters. Marine Mammal Science 17:231-248.

Wells, R. S., and M. D. Scott. 1999. Bottlenose dolphin *Tursiops truncatus* (Montagu, 1821). Pages 137-182 *in* S. H. Ridgway and S. R. Harrison, eds. Handbook of marine mammals Volume 6: The second book of dolphins and the porpoises. Academic Press, San Diego, CA.

Wells, R. S. 2003. Dolphin social complexity: Lessons from long-term study and life history. Pages 32-56 *in* F. B. M. de Waal and P. L. Tyack, eds. Animal social complexity: Intelligence, culture, andiIndividualized societies. Harvard University Press, Cambridge, MA.

Wells, R.S. 2013. Social structure and life history of common bottlenose dolphins near Sarasota Bay, Florida: Insights from four decades and five generations. Pp. 149-172 *In:* J. Yamagiwa and L. Karczmarski (eds.), Primates and cetaceans: Field research and conservation of complex mammalian societies, Primatology Monographs, Tokyo, Japan: Springer. DOI 10.1007/978-4-431-54523-1_8.

White, G.C. and K.P. Burnham. 1999. Program MARK: survival estimation from populations of marked animals.

Bird Study 46:120. doi:10.1080/00063659909477239.

Whitehead, H., and T. Wimmer. 2005. Heterogeneity and the mark–recapture assessment of the Scotian Shelf population of northern bottlenose whales (*Hyperoodon ampullatus*).Canadian Journal of Fisheries and Aquatic Sciences 62:2573–2585.

Whitehead, H. 1990. Mark–recapture estimates with emigration and re-immigration. Biometrics 46:473–479.

Wilkin, D. J., K. R. Debure and Z.W. Roberts. 1998. Query by sketch in DARWIN: digital analysis to recognize whale images on a network. Pages 41-48 *in* M. M. Yeung, B. L. Yeo, and C. A. Bouman eds. Storage and retrieval for image and video databases VII, Proceedings of the International Society of Optical Engineering SPIE vol. 3656, Bellingham, WA.

Williams, J. A., S. M. Dawson and E. Slooten. 1993. The abundance and distribution of bottlenosed dolphins (*Tursiops truncatus*) in Doubtful Sound, New Zealand. Canadian Journal of Zoology 71:2080-2088.

Wilson, B., P. M. Thompson and P. S. Hammond. 1997. Skin lesions and physical deformities in bottlenose dolphins in the Moray Firth: population prevalence and age-sex differences. Ambio 26:243-247.

Wilson, B., P. S. Hammond and P. M. Thompson. 1999. Estimating size and assessing trends in a coastal bottlenose dolphin population. Ecological Applications 9:288-300.

Würsig B and M. Würsig. 1977. The photographic determination of group size, composition and stability of coastal porpoises (*Tursiops truncatus*). Science 198:755-756.

Yoshizaki, J., K. H. Pollack, C. Brownie and R. A. Webster. 2009. Modeling misidentification errors in capture-recapture studies using photographic identification of evolving marks. Ecology 90:3-9.

Supporting Information

The following supporting information is available for this article online:

Appendix S1. List of publications describing photographic quality used for literature review.

Table 1. List of research groups that participated in the photo-identification exercise.

No.	Research organization	No. individuals
1	Cascadia Research (USA)	1
2	Dolphin Biology and Conservation (Italy)	2
3	Duke University Marine Lab (USA)	2
4	Eckerd College (USA)	1
5	Harbor Branch Oceanographic Institute (USA)	4
6	Murdoch University Cetacean Research Unit (Australia)	1
7	Pacific Islands Fisheries Science Center (USA)	1
8	National Ocean Service/NOAA Charleston (USA)	1
9	University of Otago Marine Mammal Research Group (New Zealand)	1
10	Sarasota Dolphin Research Program/Chicago Zoological Society (USA)	2
11	Sea Mammal Research Unit (Scotland)	1
12	University of Aberdeen (Scotland)	1
		18

Table 2. Summary of image selection and grading criteria used by participants for the photo-identification exercise; thirteen of the eighteen participating individuals provided their image selection criteria. Letters represent different researchers. Open circles indicate not evaluated and filled circles identify individuals which evaluated photo quality or distinctiveness features.

Photo Quality	A	C	D	E	F	G	H	I	L	N	P	R	S
quantitative	○	○	●	●	○	●	○	○	○	●	●	●	○
overall score	○	●	○	○	●	○	●	○	●	○	○	○	●
qualitative	●	○	○	○	○	○	○	●	○	○	○	○	○
Distinctiveness													
Permanent fin features	●	●	●	●	●	●	●	●	●	●	●	●	●
Scars, lesions	●	○	●	○	○	○	●	●	○	●	○	○	○
Temporary markings	●	○	○	○	○	○	●	●	○	○	○	○	○

Table 3. Summary of criteria assessed in evaluating photographs for photo-identification for capture-recapture studies from literature review of 34 publications (listed in Appendix S3).

Criteria evaluated	Proportion reported
Focus/clarity	84%
Angle	77%
Exposure	42%
Lighting/contrast	26%
Proportion visible	29%
Size in frame	42%
No criteria	13%

Table 4. Summary of timing for transitions of individuals from *Not Distinctive* to any of the other three mark acquisition phases.

	Not Distinctive (DN)	Marginally Distinctive (DM)	Moderately Distinctive (D2)	Very Distinctive (D1)
Number of dolphins in distinctiveness category at end of study period (%)	44 (57%)	18 (23%)	12 (16%)	3 (4%)
Number of dolphins transitioning to or through distinctiveness category	n/a	28	14	3
Mean number of days to reach distinctiveness category	>689	477	752	613
S.D.	n/a	347	480	582
Range (days)	37-2,699	0-1,320	0-1,775	214-1,281

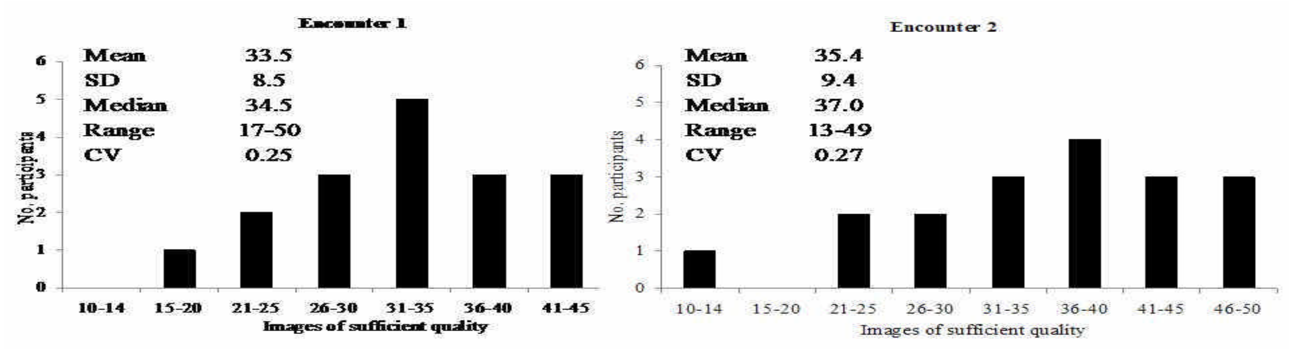

Figure 1. Summary of responses of participants in photo-identification experiment to the question, “How many images in each encounter (Encounter 1 and Encounter 2) were of sufficient image quality for photo-identification”? Note: each encounter included 50 images.

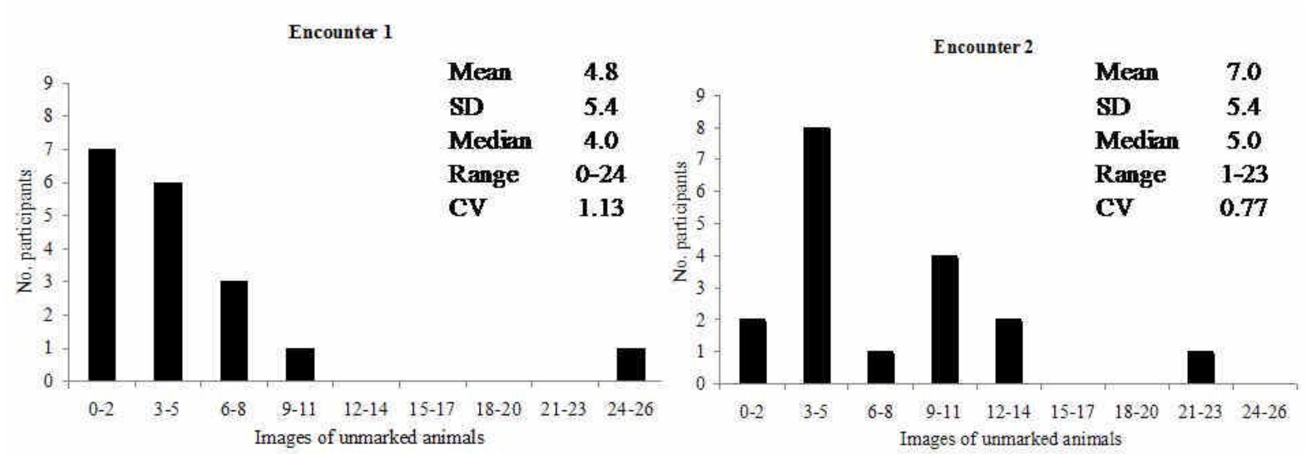

Figure 2. Summary of responses of participants in photo-identification experiment to the question, “What is the number of images that you considered to be “unmarked” or of insufficient distinctiveness for capture-recapture analysis in Encounter 1 and Encounter 2”? Note: each encounter included 50 images.

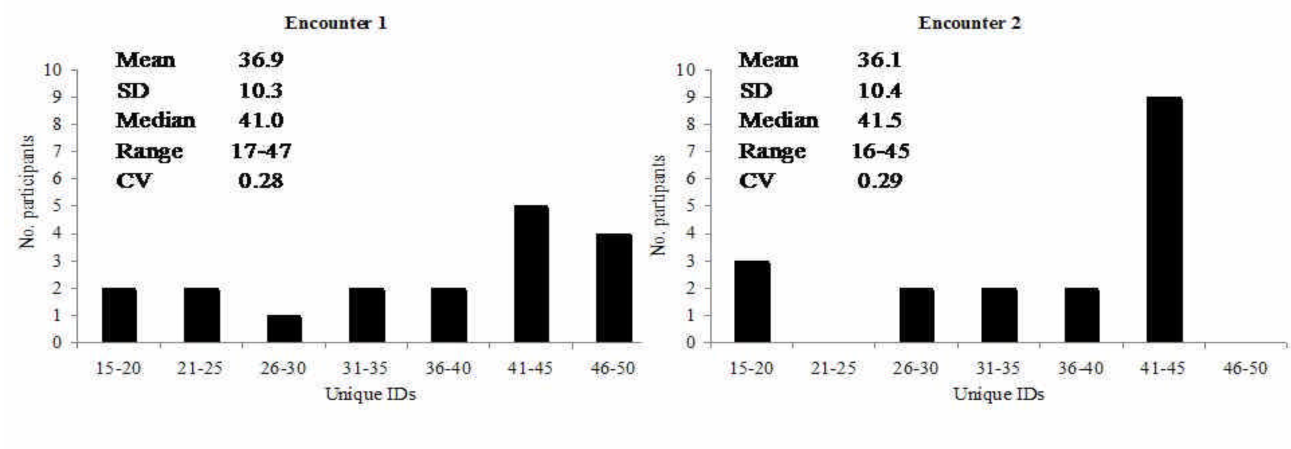

Figure 3. Summary of responses of participants in photo-identification experiment to the question, “What is the number of unique individuals within each encounter”? Note: each encounter included 50 images.

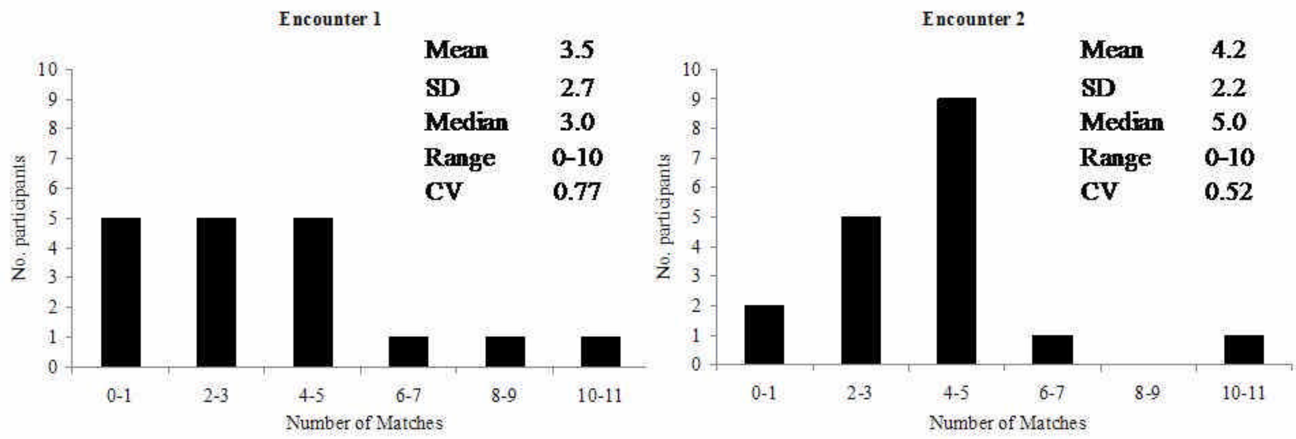

Figure 4. Summary of responses of participants in photo-identification experiment to the question, “What is the number of matched individuals within each encounter”? Note: each encounter included 50 images.

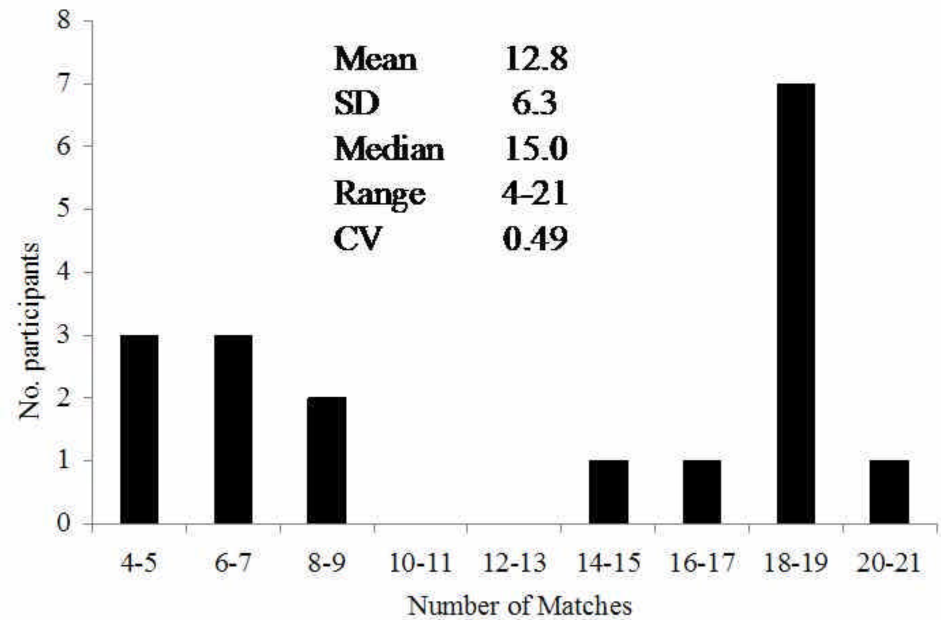

Figure 5. Summary of responses of participants in photo-identification experiment to the question, “What is the number of matched individuals between each encounter”?

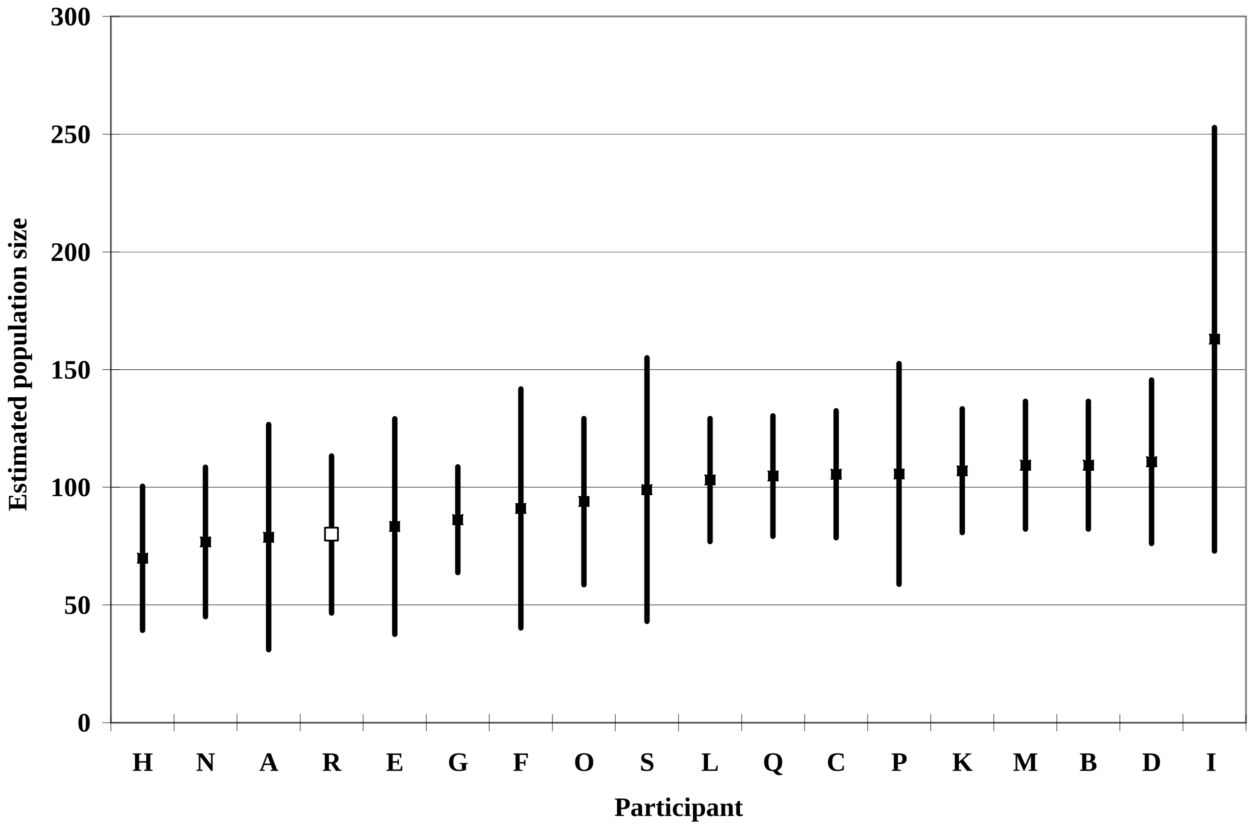

Figure 6. Results of Chapman modification of the Lincoln-Peterson model applied to test data set results, showing point estimate and 95% CI, sorted from the minimum estimate to the maximum. Letters represent participants in the photo-identification exercise. The open box representing data point "R" is the best estimate given the known number of individuals included in the exercise.