

Combing Genetic and Photo-Identification Data to Improve Abundance Estimates for the
North Atlantic Right Whale

By: Kayla Fitzgerald

A Thesis Submitted to Saint Mary's University, Halifax Nova Scotia in Partial Fulfillment of the
Requirements for the Degree of B.Sc. Honours Biology and Minor in Environmental Science.

© Kayla Fitzgerald, 2018

April 23, 2018

Supervisor: Dr. Tim Frasier,

Reader: Dr. Jeremy Lundholm

Table of Contents

Abstract 3

Introduction..... 4

 Abundance Estimates..... 4

 Pedigree-Informed Abundance Estimation..... 5

 Introduction to methods 6

 The North Atlantic Right Whale 9

Methods 13

 The Right Whale Database: 13

 Explanation of Analyses: 16

Results 20

Discussion..... 23

Acknowledgements 26

References..... 27

Combing Genetic and Photo-Identification Data to Improve Abundance Estimates for the North Atlantic Right Whale

By: Kayla Fitzgerald

Abstract

Accurate abundance estimates are important for the management of species. Many abundance estimates are based off of capture-mark-recapture analyses which have several limitations that can reduce the accuracy of abundance estimates. By integrating pedigrees into abundance estimates, the accuracy of these estimates can be improved. In this study, a method created by Creel and Rosneblatt (2013) was adapted and used to estimate abundance of the North Atlantic Right Whale (*Eubalaena glacialis*). Both genetic and sightings data, obtained from the North Atlantic Right Whale Consortium, were used to estimate abundance. This method is compared to the photo-identification method typically used to estimate abundance for this species. The results of this study suggest that the photo-identification method for this species is actually more accurate than other researchers have previously believed.

Introduction

Abundance Estimates

Accurate abundance estimation is crucial for the management and conservation of populations. Abundance estimates are the primary criterion upon which a population's status is based when being considered for listing under endangered species legislation (e.g., the Species At Risk Act in Canada, or the Endangered Species Act in the U.S.), and therefore is key in determining how a population is prioritized with respect to recognition and funding for conservation. Moreover, accurate abundance estimates are required to determine trends over time, such as identifying if a population is increasing or decreasing. Reliably detecting these trends is key for developing new management or conservation plans, as well as when evaluating the effectiveness of existing plans (e.g. Caughley & Gunn, 1996).

Given the importance of abundance estimation, a variety of methods have been used and/or developed. For species where specific individuals cannot be recognized (either due to a lack of distinguishing features, or because population sizes are too large to make this practical) "distance sampling" approaches – such as quadrat or line-transect sampling – are often used (Buckland et al., 2001; Krebs, 1989). For species where individuals can be recognized, based either on naturally occurring physical characteristics, or on artificial "marks", capture-mark-recapture (CMR) methods are often used for abundance estimation (e.g. Amstrup, McDonald, & Manly, 2005; Krebs, 1989).

For CMR abundance estimation, the methods used to identify individuals have changed over time. Originally, most such studies required the placement of artificial "marks" on individuals, such as ear tags for large mammals and leg bands for birds (e.g. Hestbeck, Nichols, & Malecki, 1991; Rice & Harder, 1977). Although these methods are still often used, it is increasingly being recognized that individuals across a wide range of species can be identified based on naturally-occurring characteristics. This allows

mark-recapture studies to be conducted without requiring the physical capture of individuals (instead, just taking photographs of the identifying features) (e.g. Langtimm et al., 2004; Wilson, Hammond, & Thompson, 1999). Even more recently, the increasing ease with which individual-specific genetic profiles can be obtained from small sources of DNA has made it possible (and sometimes more efficient) to use genetic data as the primary means of individual identification, and thus the basis for CMR analyses (Mowat & Strobeck, 2000; Palsboll et al., 1997; Petit & Valiere, 2006).

Pedigree-Informed Abundance Estimation

Many species can be difficult to monitor through typical CMR methods. This is particularly true for species who have a wide distribution which makes it difficult to “capture” an adequate representation of the population. For example, CMR studies require that “tagged” individuals are able to be re-sighted (González-Vicente, Díaz, Mallol, & Goñi, 2012). However, in some instances, these tags may be lost from the tagged individual(s) (González-Vicente et al., 2012). González-Vicente et al. (2012) explored these effects in European spiny lobsters (*Palinurus elephas*) which lose any tag that may have been attached to the original shell when individuals undergo molting. These authors illustrate that the loss of tags result in individuals not being identified as “tagged” which leads to overestimates of abundance. For cases where photo-identification can replace the need of physical tags, CMR studies still require a “recapture” period, which may be difficult for species who have a wide distribution. For example, Gore et al. (2016) used CMR methods and photo-identification to estimate the abundance of a population of basking sharks (*Cetorhinus maximus*). However, as a result of limited recaptures resulting from the wide distribution of the population, accurate abundance estimates were not possible for the population as a whole (Gore, Frey, Ormond, Allan, & Gilkes, 2016). However, when genetic relationships and field data are available, these data can be combined to gain a better understanding of populations (Israel & May, 2010).

Pedigrees are a valuable source of information for monitoring populations, and provide information to aid in understanding population dynamics (e.g., Frasier et al., 2007). Pedigrees use genetics analyses to understand relationships among individuals. Pedigree analyses combine field data and genetic data to assess the genetic relationships among individuals (Frasier, Hamilton, et al., 2007). The use of pedigrees can eliminate some of the issues experienced with typical CMR studies by providing a means to infer the presence of individuals that have not been directly “tagged”. For example, Pearse et al. (2001) conducted a CMR study of painted turtles (*Chrysemys picta*) using genetic fingerprints as individual “marks”. By genetically sampling some individuals within the population, including offspring and one known parent, the second parent could be genetically inferred (Pearse, Eckerman, Janzen, & Avise, 2001). This information was then used to estimate the effective number of breeding individuals in the population (Pearse et al., 2001). Similarly, Frasier et al. (2007) conducted parentage analysis of the endangered North Atlantic right whale (*Eubalaena glacialis*) – a species so well documented that it has conventionally been thought that all individuals have been identified. However, their study found that a potentially large number of North Atlantic right whales had gone undetected by photo-identification-CMR surveys. By using genetic data of female-calf pairs, pedigree analyses allowed them to infer the presence of a substantial number of undetected males that must exist within the population (Frasier, Hamilton, et al., 2007). Where typical CMR surveys left these males undetected, pedigree analyses could include them in the abundance estimates. With the pedigree information, these elusive individuals did not have to be seen to be included in abundance estimates.

Introduction to methods

One recent advancement in the inference of the existence of individuals through pedigrees is that by Creel and Rosenblatt (2013). Previous studies using pedigrees to estimate abundance have been limited to the estimation of the number of breeders. The method developed by Creel and Rosenblatt

allows overall abundance estimates to be obtained through pedigree reconstruction. This method is based on data from three main categories of individuals: (1) those that are directly sampled (N_s); (2) those that are inferred through pedigree analysis (N_{in}); and (3) those that are invisible to the pedigree analysis because they are neither sampled nor inferred (N_{iv}). The sum of these values results in a pedigree-informed estimate of abundance.

$$\hat{N} = N_s + N_{in} + N_{iv} \quad \text{Equation 1}$$

Through field work, the number of individuals directly sampled (N_s) can be directly counted. Through reconstruction of pedigrees from sampled offspring, the existence of one or more non-sampled parents can be inferred, resulting in a count for N_{in} . The crux of this method is to estimate N_{iv} - the number of individuals that are invisible to pedigree reconstruction. These represent two types of individuals: (1) non-sampled non-breeders, which includes adults that weren't sampled and didn't breed, as well as individuals not yet of reproductive age that have not been sampled; and (2) breeders that were neither sampled nor inferred. As shown below, these quantities can be estimated from the data.

To estimate the number of non-sampled non-breeders (N_{nsnb}), we first need to estimate the probability of an individual being sampled. If we make the assumption that sampling patterns are the same among breeders and non-breeders, then we can use data from directly sampled and inferred breeders to estimate the probability of an individual being sampled. Specifically, the probability of being sampled (P_s) should equal the number of sampled breeders (B_s) divided by the total number of breeders (directly sampled, B_s , and inferred, N_{in}).

$$P_s = \frac{B_s}{B_s + N_{in}} \quad \text{Equation 2}$$

Next, we need to estimate the probability of an individual being a breeder (P_b). This can be estimated as the total number of breeders (sampled and inferred) divided by the total number of sampled individuals. Specifically,

$$P_b = \frac{B_s + N_{in}}{N_s} \quad \text{Equation 3}$$

The probability of an individual not being a breeder (P_{nb}) is then calculated as $1 - P_b$.

We can use these data to estimate the number of non-sampled individuals (N_{ns}). Specifically, if we know the number of sampled individuals (N_s), and we have an estimate of the probability of being sampled (P_s), then it holds that $N_s = P_s \cdot X$, where X is the total population size. Solving for X results in (N_s / P_s) . Although solving for X is our ultimate goal, and we could do so here, we can also extend the data further to obtain more information about the population. Therefore, instead of estimating population size at this stage, here we just use these data to estimate number of non-sampled individuals (N_{ns}), which would be $X - N_s$. Combined, this results in:

$$N_{ns} = \frac{N_s}{P_s} - N_s \quad \text{Equation 4}$$

The number of non-sampled non-breeders (N_{nsnb}) can then be calculated by multiplying the number of non-sampled individuals (N_{ns}) by the probability of an individual not breeding (P_{nb}):

$$N_{nsnb} = N_{ns} \times P_{nb} \quad \text{Equation 5}$$

The other category of individuals who are invisible to pedigree reconstruction are non-sampled non-inferred breeders (N_{bnsni}). First, an estimate of the number of non-sampled breeders (N_{bns}) can be obtained by multiplying the estimate of the number of non-sampled individuals (N_{ns}) from above by the probability of breeding (P_b):

$$N_{bns} = N_{ns} \times P_b \quad \text{Equation 6}$$

This includes all non-sampled breeders, and therefore includes those that have already been inferred (N_{in}). Therefore, the number of non-sampled non-inferred breeders (N_{bnsni}) can be obtained by:

$$N_{bnsni} = N_{bns} - N_{in} \quad \text{Equation 7}$$

The estimate for the total number of individuals invisible to pedigree reconstruction (N_{iv}) is then:

$$N_{iv} = N_{nsnb} + N_{bnsni} \quad \text{Equation 8}$$

This can be used, along with N_s and N_{in} , in equation 1, to obtain a pedigree-informed estimate of abundance that includes non-breeders as well as breeders.

The North Atlantic Right Whale

The North Atlantic right whale (*Eubalaena glacialis*) is a highly endangered species (Kraus & Rolland, 2007a). This is primarily the result of a high rate of anthropogenic mortality and a reproductive

rate that is three-times lower than their known potential (Kraus & Rolland, 2007a). Although the population showed a very slow rate of increase (~2%/year) in the past (Kraus & Rolland, 2007b), it is now declining (Pace, Corkeron, & Kraus, 2017). This species inhabits a highly urbanized environment, which results in boat collisions and fishing gear entanglements being a prominent cause of mortality (Kraus & Rolland, 2007b; Pace et al., 2017). While females are capable of giving birth to one calf every three years (Hamilton, Knowlton, & Mar, 2007), intervals between births are frequently more than 8 years, and many females have never reproduced (Cole et al., 2013). A combination of factors are believed to be affecting this low fertility rate, such as: low genetic diversity, oceanographic processes (which influence prey availability and aggregation), disease, stressors from urbanization, and loss of critical habitat (Kraus & Rolland, 2007a; Pace et al., 2017).

The true level of understanding regarding the right whale population is controversial. Some researchers believe that the data available accurately represent the species as a whole, while others are more skeptical (Frasier, Hamilton, et al., 2007; Kraus & Rolland, 2007a). For example, the photo-identification data suggests that the majority of individuals have been identified, because few adults are added to the photo-identification catalog, and most “new” individuals represent known births (Hamilton et al., 2007). Additionally, based on over 40 years of sightings histories, 5 critical habitats have been identified, ranging from the Gulf of Mexico to Iceland (Kraus & Rolland, 2007a). These habitats include: (1) Southeast USA as the only known calving ground; (2) Cape Cod Bay as a winter and spring feeding ground; (3) the Great South Channel as a spring feeding ground; (4) the Bay of Fundy as a summer feeding ground, and as a summer nursery ground for females with lactating calves; and (5) Roseway Basin as a summer feeding area (Kraus & Rolland, 2007a). Each year, the majority of photo-identified individuals are sighted at least once during surveys of these critical habitats. Individuals have been

sighted in other habitats (e.g. Jeffery's ledge), but these areas are not frequently surveyed as a result of funding and surveying limitations (Kraus & Rolland, 2007a).

The researchers that are more sceptical, however, point out that some data suggests that there may be additional individuals, and/or habitats, that have gone unidentified (Frasier, Hamilton, et al., 2007; Kraus & Rolland, 2007a). For example, during the winter months, right whale habitat use patterns are not well understood (Kraus & Rolland, 2007a; Schaeff, Kraus, Brown, & White, 1993). This is a crucial period of the right whale life cycle as it includes both mating and birthing periods (Cole et al., 2013). While new mothers are usually seen with calves off of the coast of Florida, other individuals of the population are not typically sighted during the winter and their location is unknown (Brillant, Vanderlaan, Rangeley, & Taggart, 2015; Brown et al., 2001; Kraus & Rolland, 2007a). The mating ground of this population has yet to be discovered (but some data suggest it is in the Gulf of Maine (Cole et al., 2013). These are crucial aspects of the right whale life cycle that have yet to be discovered (Frasier, Hamilton, et al., 2007; Kraus, Pace III, & Frasier, 2007).

There also exists a group of individuals who are sighted so infrequently they are commonly referred to as "irregular" whales (Kraus & Rolland, 2007a). These individuals can go multiple years without being sighted, suggesting that for every season there are critical habitats yet to be identified (Kraus & Rolland, 2007a). These irregular whales include a number of females that do not take their calves to the common nursery area, the Bay of Fundy (Kraus & Rolland, 2007a). Irregular calving females have been speculated to use the Gulf of Saint Lawrence (Kraus & Rolland, 2007a) and/or Jeffery's ledge (Weinrich, Kenney, & Hamilton, 2000) but this has not been confirmed (Cole et al., 2013). Recent evidence suggests that nursery habitat use is maternally directed in offspring (Cole et al., 2013; Kraus & Rolland, 2007a). This means that all calves born to irregular females are not detected in their first year

(Schaeff et al., 1993). Additionally, these calves will not use the Bay of Fundy with their calves. Thus, identification of an additional nursery ground could significantly increase our knowledge of this species (Frasier, Hamilton, et al., 2007; Kraus & Rolland, 2007b). Irregular whales represent another vital gap in our understanding of this species (Kraus & Rolland, 2007b).

Combined, despite our wealth of knowledge on this species, at any given time of year the location of at least one-third of the individuals is not known (Brown, Kraus, Slay, & Garrison, 2007). The implication is that for every time of year there must exist critical habitats that have yet to be identified. This raises the question of how many whales *only* use these unidentified areas and are therefore *never* photo-identified or included in abundance estimates or other species assessments. Again, there are different thoughts regarding how big of a pool these individuals represent, with some researchers suggesting that the numbers of such individuals are negligible, while others think that they may represent a significant portion of the species.

Management of this highly endangered species requires accurate estimates of abundance and trends over time (Caughley & Gunn, 1996; Kraus & Rolland, 2007a). However, because most survey efforts are focused within 5 main habitats, individuals who do not use these habitats at all, or during survey periods, will continue to go unidentified (Kraus & Rolland, 2007a). Failure of conservation efforts to improve the outlook for this population may be related to misunderstanding of important habitats, population estimates, and undetected population limiting factors due to survey efforts being geographically limited. Thus, researchers may choose to include previously overlooked habitats in population surveys if scientific evidence supports the existence of undetected individuals. If this species is more abundant than what current estimates suggest, then efforts to re-evaluate where surveys are focused could be initiated.

In this study I used both genetic and photo-identification data from the North Atlantic Right Whale Consortium to estimate abundance of the right whale population using a method developed by Creel and Rosenblatt (2013). By using this method, I hoped to obtain more accurate abundance estimates by incorporating individuals that can only be inferred by genetics or have never been documented in the right whale photo-identification catalog. If the population is found to be larger than current estimates have predicted, then this has the potential to alter the way that the species and its habitat use patterns are interpreted and analysed. Additionally, this could change the management systems currently being used to protect this species. A re-evaluation of this species as a result of these data could push other researchers of other species to conduct similar analyses to judge if their knowledge is as accurate as they have assumed.

Methods

The Right Whale Database:

Photo-identification data of North Atlantic right whales have been collected almost year-round, and in many of the known habitats, since 1979 (Hamilton et al., 2007). Photo-identification for this species is primarily based on callosity patterns (Kraus & Rolland, 2007a). Callosities are grey to black patches of thickened skin that may be located on the rostrum, near the blowholes and eyes, and lining the upper lips (Kraus & Rolland, 2007b). These callosity patterns are specific to each individual, but differ among individuals, thus providing a means on which individuals can reliably be identified (Hamilton et al., 2007). Calves cannot be photo-identified until their callosity pattern is developed, which occurs between 7-10 months of age (Hamilton et al., 2007). Therefore, unless genetic samples are obtained,

calves cannot be added to the catalog until this pattern is developed, which in some cases, is after they have been weaned by their mothers (Hamilton et al., 2007; Schaeff et al., 1993).

The photo-identification and genetic data used for this study were obtained from the North Atlantic Right Whale Consortium (NARWC). This is a collaboration of researchers, government, and non-governmental organizations that curates all data collected on this species from 1935 to the present (Kraus & Rolland, 2007b). The primary method used to study this species is photo-identification (Kraus & Rolland, 2007a; Reeves, Read, Lowry, Katona, & Boness, 2007). The protocol for photo-identification is standardized and the method is laid out by Kraus et al. (1986). Individuals are added to the photo-identification database when there is certainty that they have not already been identified and have an accurate and clear picture that can function as a reliable reference for identification in the future (i.e. the majority of the individual's identifiable characteristics are captured clearly) (Hamilton et al., 2007). Individual profiles in the database include the individual's identification number, as well as a detailed drawing showing their identifying features (e.g. callosity patterns, scars, pigmentation patterns, and ridges along the upper lip), which is updated annually. Additionally, individual profiles include: age, sex, genetic sample (if obtained), sightings history, reproductive status, calving history, mortality, and any other relevant information (Kraus & Rolland, 2007b). These data are obtained via ship and/or aerial surveys which are conducted annually (Kraus & Rolland, 2007b). Surveys are conducted in the 5 critical habitats identified for this species (Kraus & Rolland, 2007a). This database has been essential to right whale research and has helped establish a better understanding of this species (Kraus & Rolland, 2007a; Pettis, Pace, Schick, & Hamilton, 2017).

Genetic profiles of each sampled individual are maintained by the NARWC as a data set (Hamilton et al., 2007). The protocols for genetic sampling and analysis were laid out by Brown et al. (1991) and Frasier et al. (2006) respectively. While typical DNA samples are obtained via skin biopsy

samples, genetic information can also be obtained from fecal samples and dead carcasses (Best et al., 2005). Skin biopsies are collected from free-swimming whales with a crossbow and modified bolt (arrow), which contains a hollow tip and a stop collar that also acts as a float. In this way, the bolt hits the whale, penetrates the skin ~1cm, and bounces out with a small piece of skin inside, and floats. The bolt can then be retrieved with a net from the boat. Although this method is slightly invasive, intensive and long-term studies have shown that it does not have any negative short- or long-term effects on the whales, other than an initial startle response, which itself is only seen a small portion of the time (e.g. Best et al., 2005; Brown, Kraus, & Gaskin, 1991; Noren & Mocklin, 2012). This allows skin samples to be obtained from a distance. Biopsy sampling can be useful for obtaining genetic samples from calves that are too young to be photo-identified (Frasier, McLeod, Gillett, Brown, & White, 2007). As of 2009, it was thought that about 75% of photo-identified individuals had also been genetically sampled (Gillett, Frasier, Rolland, & White, 2010). Genetic analyses include analysis of a sex-specific region for gender determination, sequence analysis of the mitochondrial control region, and genotype analysis at 35 microsatellite loci (Frasier, McLeod, et al., 2007). Microsatellite loci are highly variable regions of non-coding DNA that contain repeated segments of 1-6 base pairs (Frasier, McLeod, et al., 2007). These loci are inherited in a Mendelian fashion, where one copy is inherited from the mother, and one from the father (Frasier, McLeod, et al., 2007). Thus, these samples are useful for individual identification as well as assessing parentage and relatedness within the population (Frasier, McLeod, et al., 2007).

Annual sightings data from the NARWC are used on an annual basis to produce an estimate of abundance for this species (Kraus & Rolland, 2007b). Abundance estimates produced from this information are based on individuals which are “presumed alive” which refers to any individuals that have been seen within the past 6 years (Hamilton et al., 2007). This method is used because not all individuals are seen in each year, but 99% of those that are alive are seen within a 6-year time frame. On

the other hand, most whales not seen for over 6 years are never seen again (and are therefore presumably dead) (Hamilton et al., 2007; Pace et al., 2017). It should be noted that as a result of this method, data from earlier years are more robust than those from recent years. This is because individuals have a better chance of being added to abundance estimates as time continues. For example, for the abundance estimate of 1990, individuals included in this estimate have had 28 years to be seen. However, for the abundance estimate of 2016, there are still many individuals that were likely alive in that year, but have yet to be seen, resulting in a downward bias for that year's estimate. Therefore, earlier abundance estimates can be considered more robust compared to more recent abundance estimates (Pace et al., 2017). This can also result in a lag in detection of population trends, as the data can become more robust with time (Pace et al., 2017). This "presumed dead" method therefore provide a coarse view of abundance, but which also results in fairly accurate estimates when compared with those obtained using more robust statistical procedures (e.g. estimates are lower than in reality if the "presumed dead" individuals undergo a longer than typical sightings gap) (Pace et al., 2017) and thus is stated to not be considered as an "absolute" abundance estimate (Frasier, McLeod, et al., 2007; Pettis et al., 2017).

Explanation of Analyses:

The photo-identification and genetic data from 1980-2016 were obtained from the NARWC and imported into a MySQL relational database. To obtain reference abundance estimates based solely on the photo-identification data, the database was queried for all whales that were alive (or presumed to be alive) during each year. The database was additionally queried for the number of individuals known to be sampled or not-sampled, broken down by males, females, and for the total population. Non-sampled individuals result from individuals that are known to be alive and are believed to still be alive (i.e. have been identified and sighted a maximum 6 years prior to this year), but have not yet been sampled.

Next, I estimated abundance based on pedigrees. For each year I obtained all genetically sampled mother-calf pairs from the database. Additionally, I also obtained the number of candidate males which were genetically sampled from the database. Candidate males (i.e. males who are potential fathers to calves) are males who were alive and older than 5 years of age during the calf's year of conception. The age of 5 was used as the cut off for candidate males, as the actual age at which right whale males become sexually mature is unknown, but the average age at which females have their first calf is 8 (Frasier, Hamilton, et al., 2007). Therefore, this approach is conservative (i.e., it is considering more males as candidate fathers than likely actually exist). This reduced the probability that a true candidate male would be inappropriately excluded as a potential father (Frasier, Hamilton, et al., 2007). Pedigree analyses were conducted using a program called CERVUS which is a publically available computer program that is commonly used by molecular ecologists to assess paternity (Marshall, Slate, Kruuk, & Pemberton, 1997; Slate, Marshall, & Pemberton, 2000). CERVUS compares the genetic profiles of known mother-calf pairs, and from those, infers which alleles must have been inherited from the father. It then searches the list of candidates for males whose genetic profile is consistent with that inferred from each mother-calf pair. CERVUS allows paternity to be assigned using multiple techniques, and I used the exclusion method, which is the method used for this species, and is also the method used in human forensic paternity cases (Frasier, Hamilton, et al., 2007). An example of the exclusion method would be to compare the genetics of a mother-calf pair to the genetics of all potential candidate males. Each candidate male that has a genetic incompatibility with the mother-calf pair are excluded as the potential father. The elimination of candidate males continues until, in an ideal case, there is only one remaining male with no genotypic incompatibilities with the mother-calf pair. This male would then be assigned paternity of that calf.

For this study, pedigree analysis could result in one of three scenarios. In the scenarios where all but one male was excluded as a potential father, paternity was assigned. In scenarios where multiple males could not be excluded, paternity was not assigned (i.e. paternity was unresolved, but it was concluded that the true father was sampled). Finally, in cases where all candidate males were excluded (i.e. no candidate male had a genome which was compatible with the mother-calf dyad), paternity was not assigned and it was concluded that the true father has not been sampled. Thus, for each calf where all candidate males were excluded as the father, one male was inferred to exist (i.e. a male that has not been genetically sampled). These are referred to as “inferred individuals”.

Pedigree analysis was also conducted for calves who had been genetically sampled, but whose mother was not. In these cases pedigree analysis was used to assign both maternity and paternity. The number of inferred individuals resulting from pedigree analyses were totaled for females, males, and for the whole population.

Finally, the number of breeders for each year was calculated. This was divided into sampled breeders (B_s), which were sampled parents identified through photo-identification and/or pedigree analysis; and inferred breeders (N_{in}), which were the number of un-sampled breeders inferred through the pedigree analyses. These values were then organized for each year as the total number of breeders for females, males, and the whole population.

Ideally, analyses of these data would have been conducted for each year. However, there were too few calves born in each year on which to base the inferred individuals. Therefore, I had to develop a strategy for how to group years appropriately. I chose to analyse the data in rolling sets of five years to result in larger sample sizes per period, but having the period lengths short enough so that an individual is not both a calf and a parent during the same time period. Right whale females become sexually mature at 8 years of age on average, with the youngest recorded mom being 5 (Kraus et al., 2007). This is

why 5-year periods were chosen. In this way, my estimate for 1990 included data from 1988-1992, the estimate for 1991 included data from years 1989-1993, and so on.

Values obtained from queries from the MySQL Database were then organized to be used in the calculations of this study. The values needed for each 5 year period were the total number of individuals sampled (N_s), the total number of individuals inferred (N_{in}), and the total number of breeders sampled (B_s). The analyses were conducted three times, once for all individuals, a second time for just females, and a third time for just males. From these values, the probability of an individual being sampled (P_s) and the probability of an individual being a breeder (P_b) can be estimated as described in equations 2 and 3, respectively.

A Bayesian approach was used to estimate the most likely estimates of P_s and P_b , as well as the uncertainty associated with these estimates. The two estimated parameters (P_s and P_b) are ratios where one of two outcomes are possible; a “success”, or a “failure”. For example, a “success” in the case of P_s , is when an individual has been sampled, and a “failure” is when an individual has not been sampled. The data therefore represent a binomial distribution, where the goal is to estimate the probability of “success” (p). As a prior for p I used a uniform distribution between 0 and 1 (i.e. I considered all values between 0 and 1 to be equally probable). All calculations and statistical analyses were conducted in the program R (R Core Team, 2017) via the RStudio (RStudio Team, n.d.) interface. A Markov Chain Monte Carlo (MCMC) was used to estimate the posterior distribution. The MCMC samples randomly selected values from the priors. The likelihood of a selected value is calculated given the data. Then, the MCMC selects a new value. If this new value has a higher likelihood than the current value, the MCMC selects this new value. In this way, the MCMC can find which values are most likely by calculating how long it remains at a value. The MCMC process thus finds the most likely value for a parameter through many random samples of values. Using this method, the MCMC finds the most likely probability of being

sampled (P_s) as well as the uncertainty around that estimate. This process is conducted with JAGS (Plummer, 2003) (communicating with R via the package runjags (Denwood, 2016) through R-studio. This process is again repeated to estimate P_b and the corresponding uncertainty.

Finally, the values calculated above were used to estimate the number of invisible individuals (N_{iv}) using equations 4-8. With all of the required values obtained (N_s , N_{in} , and N_{iv}), the total abundance estimate for each period is then calculated (equation 1).

Results

The abundance estimates obtained for the total population using the pedigree-informed approach are similar to those based solely on the photo-identification data (Figure 1). These estimates show a slow increase from 1980-2010, followed by a decline in recent years (from 2011-2013). A similar pattern was also found by Pace et al. (2017). In addition to these total estimates, I also conducted the analyses independently for each sex because the patterns of inference may be different for each.

For males, the pedigree-informed abundance estimates were generally larger than those obtained based solely on the photo-identification data, with the one exception being for 2014 (Figure 2). Despite this, in many years the photo-identification estimates still fall within the 95% highest density interval (HDI) of the pedigree-informed estimates. However, there are three clusters of years in which the pedigree data suggest a substantially larger number of males than does the photo-identification data. These are the periods of 1990-1991, 2000-2003, and 2007-2011. These three clusters correspond to years when calf numbers were particularly high, and therefore so was the ability to infer additional males.

For females, the pedigree-inferred estimates were generally lower than those based on photo-identification (Figure 3). While both methods show a slow increase in female abundance, the widest range between these two estimates (i.e. where the photo-identification estimates fall outside of the 95% HDI of the pedigree-informed estimates) occur in the years 2004 – 2011. During these years there were not any inferred females, making the pedigree-informed method – as implemented – reduce to just the number of sampled individuals, which is known to be an under-estimate.

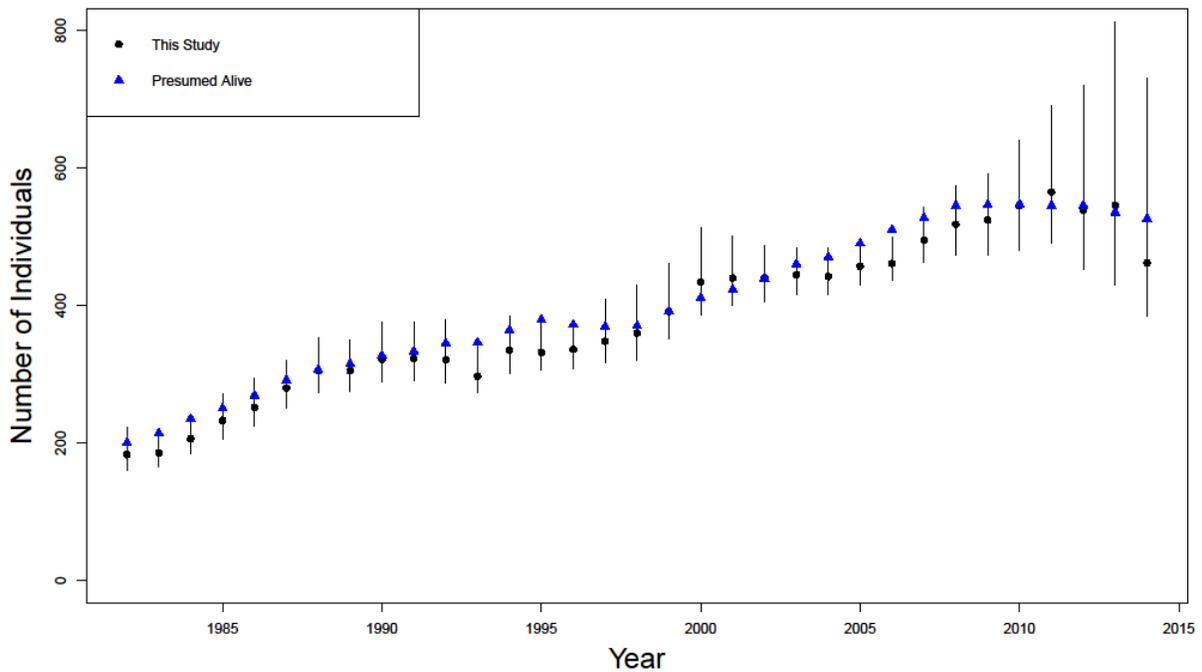


Figure 1: Abundance estimates for the entire population, showing results from the pedigree-informed estimate compared to the presumed alive estimate (based off of sightings data alone).

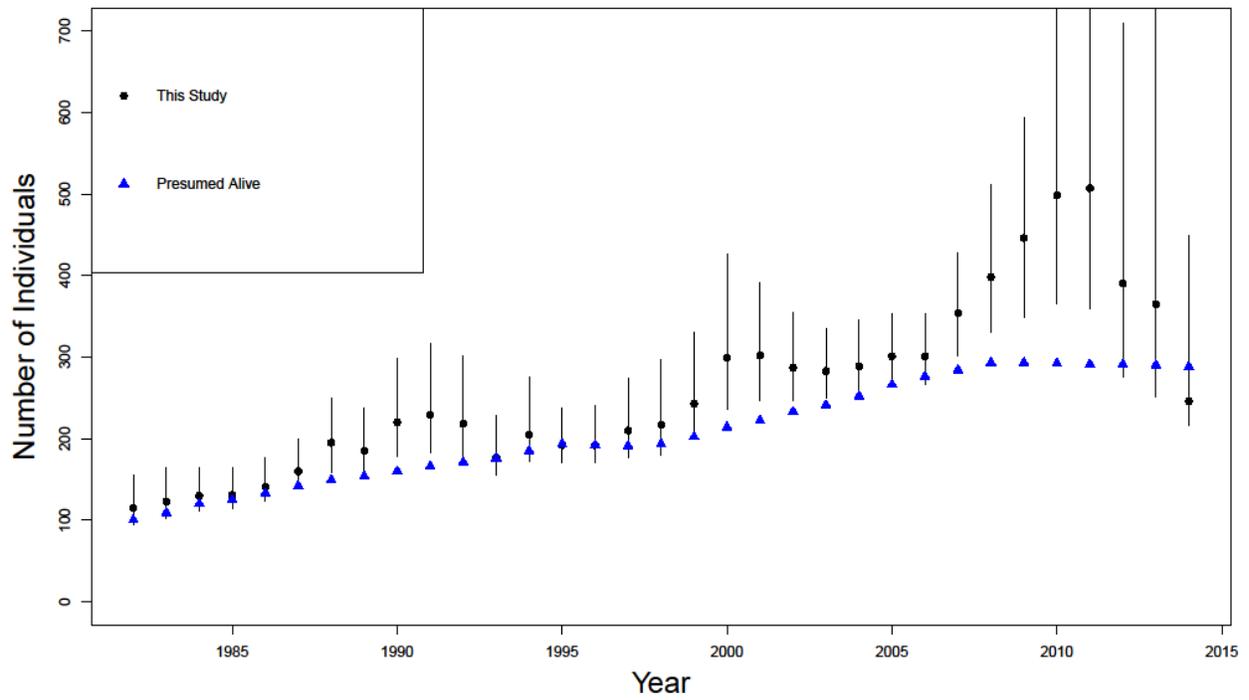


Figure 2: Abundance estimates for the male population, showing results from the pedigree-informed estimate compared to the presumed alive estimate (based off of sightings data alone).

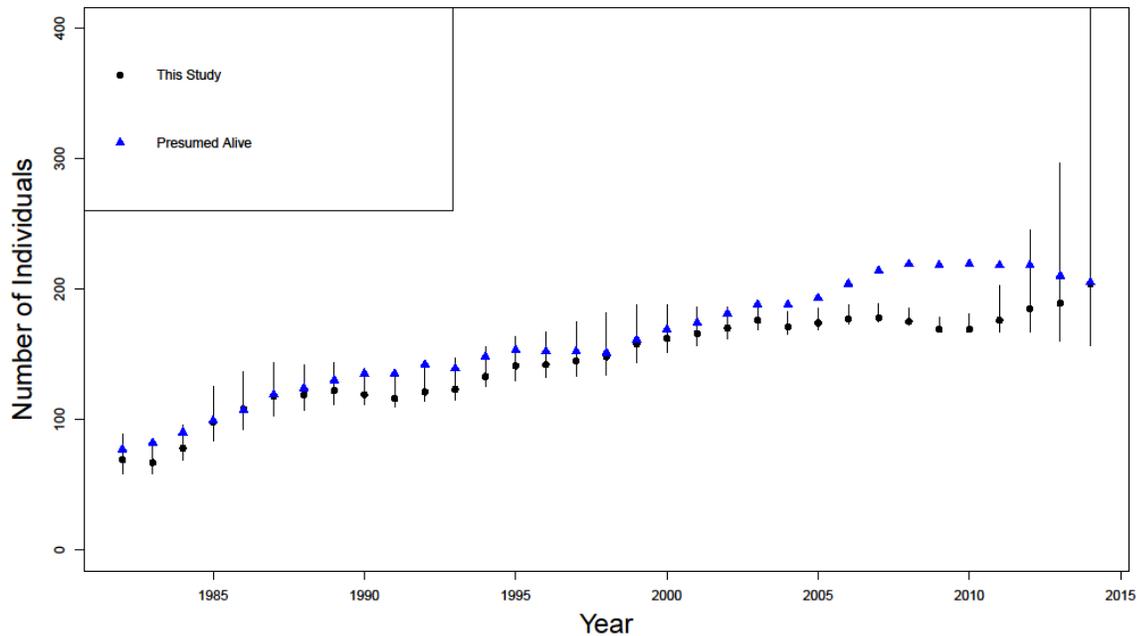


Figure 3: Abundance estimates for the female population, showing results from the pedigree-informed estimate compared to the presumed alive estimate (based off of sightings data alone).

Discussion

The abundance estimates found for the total population are similar to the presumed alive abundance estimates obtained from just the photo-identification data (Figure 1). Generally, the abundance estimate from the presumed alive method are within the range of the 95% highest density interval (HDI) of the pedigree-informed estimates. This result suggests that the presumed alive estimate is similar to the actual abundance of this population, and that, contrary to my expectations, there is not likely a large number whales who have not been photo-identified.

The pedigree-based abundance estimate for the female portion of the population is generally lower than that based just on the photo-identification data (Figure 2). This can be attributed to the data required for the pedigree-based method. As can be seen by equations 1, 2, and 3, the pedigree-based method is only appropriate when the number of inferred individuals is not zero. Otherwise, this estimate reduces to just N_s , the number of sampled individuals, which will almost always be smaller than the true abundance. In the majority of years, the number of females that could be inferred was very low, and in many cases was zero, and therefore resulted in underestimates for the number of females. This is likely due to the fact that there is just one calving ground for this species, which is intensively surveyed for management purposes. Therefore, this represents a critical habitat to which *all* mothers must go, resulting in most females being photo-identified at some point in their lives. There is no such “requirement” for males in this species, resulting in a higher probability of males not being photo-identified than females.

The pedigree-based abundance estimate for the male portion of the species is generally larger than the photo-identification estimate (Figure 3). This can be attributed to the inclusion of inferred individuals in this study. By inferring paternity of calves, males that have not been genetically sampled can be inferred to exist if all other genetically sampled males are excluded as a potential father. This allows the inclusion of males who have gone undetected via direct sampling methods to be included in the abundance estimate. The years where the largest difference is seen between the two methods (i.e. where the photo-identification estimate falls below the 95% HDI of the pedigree-based estimate: 1990-1991, 2000-2003, and 2007-2011), correlate with years with the largest sample size of calves. This could result for two different reasons. First, these years correlate with the highest numbers of sampled calves, resulting in more opportunities to infer males (i.e., more power to infer individuals). Under this interpretation, these years would be more representative of “the truth” than in years with smaller calf

numbers. This suggests that there are indeed a substantial number of “missing” males in this species. The second possibility is that in these years it is possible that the same male fathered more than one calf, and therefore, our assumption of each inferred father being unique may be incorrect, resulting in an over-estimate of inferred males. Frasier et al. (2007) conducted analyses of calves (using data from 1980-2006) and found that this assumption was valid. These authors concluded from analyses of calves for which all sampled males were excluded as fathers, that each of these calves was fathered by a different male. However, similar analyses should be conducted on this data set to test whether or not this assumption is still valid, and therefore to determine which of these two scenarios is the appropriate interpretation.

The abundance estimates for all three analyses (total, male, and female) indicate a decline in the population from 2012-2014 (Figure 1, Figure 2, Figure 3). However, my results cannot be interpreted as support for the proposed population decline in recent years (Pace et al., 2017). This analysis did not include the calves born in 2015 and 2016 as they had not been included in the data set at the time of analysis. Thus, as a result of the 5 year method, estimates from 2014 and 2013 may be considered less accurate. Additionally, the abundance estimates produced from the photo-identification data can also be considered less robust for the end of the time series (Pace et al., 2017). This is because individuals have a better chance of being added to abundance estimates as time continues. Therefore, for recent years, abundance estimates will be lower than in reality.

The exclusion method was used in this study to assign parentage. However, it should be noted that there are potential errors with using this method. For instance, a potential parent could be eliminated as the result of a genotyping error (Marshall et al., 1997). Since the method used to estimate abundance is highly dependent on the number of inferred individuals (i.e. paternity assignment), the

results of this study could potentially be influenced by this aspect. However, a study by Frasier et al. (2006) conducted an analysis of the genotyping errors within this database. This study concludes that the genotyping error of this database is <1%. Thus, if this aspect has influenced my results, it is unlikely to contribute significantly.

Overall, these analyses suggest that there is not a substantial number of “missing” individuals for this species, as was hypothesized by some researchers in the past (e.g. Frasier, McLeod, et al., 2007; Hamilton et al., 2007). The total abundance estimates produced from both methods are very similar (Figure 1) which suggests that there are few un-sampled individuals that are not also accounted for in the photo-identification data. However, my results indicate the potential for more males than are currently accounted for, but more analyses are needed before any such statements can be made.

Acknowledgements

I would like to thank the North Atlantic Right Whale Consortium for allowing me to use their data. I would also like to thank Tim Frasier for his contribution and guidance, as well as those who have supported me and provided me with feedback throughout my project.

References

- Amstrup, S. C., McDonald, T. L., & Manly, B. F. J. (2005). *Handbook of Capture-Recapture Analysis*. Princeton, NJ: Princeton University Press.
- Best, P., Reeb, D., Rew, M., Palsbøll, P., Schaeff, C., & Brandao, a. (2005). Biopsying Southern Right Whales: Their Reactions and Effects on Reproduction. *Journal of Wildlife Management*, 69(3), 1171–1180. doi:10.2193/0022-541X(2005)069[1171:BSRWTR]2.0.CO;2
- Brillant, S. W., Vanderlaan, A. S. M., Rangeley, R. W., & Taggart, C. T. (2015). Quantitative estimates of the movement and distribution of North Atlantic right whales along the northeast coast of North America. *Endangered Species Research*, 27(2), 141–154. doi:10.3354/esr00651
- Brown, M. W., Brault, S., Hamilton, P. K., Kenney, R. D., Knowlton, A. R., Marx, M. K., ... Kraus, S.-D. (2001). Sighting heterogeneity of right whales in the western North Atlantic: 1980 -1992. *Journal of Cetacean Research and Management Special Issue*, (Special Issue 2), 245–250.
- Brown, M. W., Kraus, S. D., & Gaskin, D. E. (1991). Reaction of North Atlantic right whales (*Eubalaena glacialis*) to skin biopsy sampling for genetic and pollutant analysis. *Report of the International Whaling Commission, Special Issue*, 13, 81–89.
- Brown, M. W., Kraus, S. D., Slay, C. K., & Garrison, L. P. (2007). Surveying for Discovery, Science, and Management. In S. D. Kraus & R. M. Rolland (Eds.), *The Urban Whale: North Atlantic Right Whales At The Crossroads* (pp. 105–137). Cambridge, MA: Harvard University Press.
- Buckland, S. T., Anderson, D. R., Burnham, K. P., Laake, J. L., Borchers, D. L., & Thomas, L. (2001). Introduction to distance sampling: estimating abundance of biological populations. *New York New York USA*, 10, 432. doi:10.1007/s00531-004-0408-5

- Caughley, G., & Gunn, A. (1996). *Conservation biology in theory and practice*. Massachusetts: Blackwell Science.
- Cole, T. V. N., Hamilton, P., Henry, A. G., Duley, P., Pace III, R. M., White, B. N., & Frasier, T. (2013). Evidence of a North Atlantic right whale *Eubalaena glacialis* mating ground. *Endangered Species Research*, 21(1), 55–64. doi:10.3354/esr00507
- Denwood, M. J. (2016). runjags: An R package providing interface utilities, model templates, parallel computing methods and additional distributions for MCMC models in JAGS. *Journal of Statistical Software*, 71(9), 1–25. doi:10.18637/jss.v071.i09
- Frasier, T. R., Hamilton, P. K., Brown, M. W., Conger, L. A., Knowlton, A. R., Marx, M. K., ... White, B. N. (2007). Patterns of male reproductive success in a highly promiscuous whale species: The endangered North Atlantic right whale. *Molecular Ecology*, 16(24), 5277–5293. doi:10.1111/j.1365-294X.2007.03570.x
- Frasier, T. R., McLeod, B. A., Gillett, R., Brown, M. W., & White, B. N. (2007). Right whales past and present as revealed by their genes. In S. D. Kraus & R. M. Rolland (Eds.), *The Urban Whale: North Atlantic Right Whales At The Crossroads* (pp. 200–231). Cambridge, MA: Harvard University Press.
- Gillett, R. M., Frasier, T. R., Rolland, R. M., & White, B. N. (2010). Molecular identification of individual North Atlantic right whales (*Eubalaena glacialis*) using free-floating feces. *Marine Mammal Science*, 26(4), 917–936. doi:10.1111/j.1748-7692.2010.00380.x
- González-Vicente, L., Díaz, D., Mallol, S., & Goñi, R. (2012). Tag loss in the lobster *Palinurus elephas* (Fabricius, 1787) and implications for population assessment with capture-mark-recapture methods. *Fisheries Research*, 129–130, 1–7. doi:10.1016/j.fishres.2012.05.014

- Gore, M. A., Frey, P. H., Ormond, R. F., Allan, H., & Gilkes, G. (2016). Use of photo-identification and mark-recapture methodology to assess basking shark (*Cetorhinus maximus*) populations. *PLoS ONE*, *11*(3), 1–22. doi:10.1371/journal.pone.0150160
- Hamilton, P. K., Knowlton, A. R., & Mar, M. K. (2007). Right Whales Tell Their Own Stories: The Photo-Identification Catalog. In S. D. Kraus & R. M. Rolland (Eds.), *The Urban Whale: North Atlantic Right Whales At The Crossroads* (pp. 75–104). Cambridge, MA: Harvard University Press.
- Hestbeck, J. B., Nichols, J. D., & Malecki, R. A. (1991). Estimates of movement and site fidelity using mark-resight data of wintering Canada geese. *Ecology*, *72*(2), 523–533. doi:10.2307/2937193
- Israel, J. A., & May, B. (2010). Indirect genetic estimates of breeding population size in the polyploid green sturgeon (*Acipenser medirostris*). *Molecular Ecology*, *19*(5), 1058–1070. doi:10.1111/j.1365-294X.2010.04533.x
- Kraus, S. D., Pace III, R. M., & Frasier, T. R. (2007). High Investment, Low Return: The Strange Case of Reproduction in *Eubalaena glacialis*. In S. D. Kraus & R. M. Rolland (Eds.), *The Urban Whale: North Atlantic Right Whales At The Crossroads* (pp. 172–199). Cambridge, MA: Harvard University Press.
- Kraus, S. D., & Rolland, R. (2007a). *The Urban Whale*. (S. D. Kraus & R. M. Rolland, Eds.), *North Atlantic Right Whales at the Crossroads*. Cambridge, MA: Harvard University Press. Retrieved from http://books.google.com/books?id=-m6r-SfaSy4C&printsec=frontcover&dq=intitle:The+Urban+Whale&hl=&cd=1&source=gbs_api%5Cnpapers3://publication/uuid/1B1ACAF2-B2BC-47B0-B4DD-DE27AC75DD01
- Kraus, S. D., & Rolland, R. M. (2007b). Right Whales in the Urban Ocean. In S. D. Kraus & R. M. Rolland (Eds.), *The Urban Whale: North Atlantic Right Whales At The Crossroads* (pp. 1–38). Cambridge,

MA: Harvard University Press.

Krebs, C. J. (1989). *Ecological Methodology*. New York, NY: Harper Collins.

Langtimm, C. A., Beck, C. A., Edwards, H. H., Fick-Child, K. J., Ackerman, B. B., Barton, S. L., & Hartley, W.

C. (2004). Survival Estimates For Florida Manatees From The Photo-Identification Of Individuals.

Marine Mammal Science, 20(3), 438–463. doi:10.1111/j.1748-7692.2004.tb01171.x

Marshall, T. C., Slate, J., Kruuk, L. E. B., & Pemberton, J. M. (1997). Statistical confidence for likelihood-based paternity. *Mol. Ecol*, (1998), 639–655.

Mowat, G., & Strobeck, C. (2000). Estimating Population Size of Grizzly Bears Using Hair Capture, DNA Profiling, and Mark-Recapture Analysis. *The Journal of Wildlife Management*, 64(1), 183.

doi:10.2307/3802989

Noren, D. P., & Mocklin, J. A. (2012). Review of cetacean biopsy techniques: Factors contributing to successful sample collection and physiological and behavioral impacts. *Marine Mammal Science*, 28(1), 154–199. doi:10.1111/j.1748-7692.2011.00469.x

Pace, R. M. I., Corkeron, P. J., & Kraus, S. D. (2017). State space mark-recapture estimates reveal a recent decline in abundance of North Atlantic right whales. *Ecology and Evolution*, (May), 1–12.

doi:10.1002/ece3.3406

Palsboll, P. J., Allen, J., Bérubé, M., Clapham, P. J., Feddersen, T. P., Hammond, P. S., ... Øien, N. (1997).

Genetic tagging of humpback whales. *Nature*, 388(6644), 767–769. doi:10.1038/42005

Pearse, D. E., Eckerman, C. M., Janzen, F. J., & Avise, J. C. (2001). A genetic analogue of “mark-recapture” methods for estimating population size: An approach based on molecular parentage assessments.

- Molecular Ecology*, 10(11), 2711–2718. doi:10.1046/j.0962-1083.2001.01391.x
- Petit, E., & Valiere, N. (2006). Estimating population size with noninvasive capture-mark-recapture data. *Conservation Biology*, 20(4), 1062–1073. doi:10.1111/j.1523-1739.2006.00417.x
- Pettis, H. M., Pace, R. M. I., Schick, R. S., & Hamilton, P. K. (2017). North Atlantic Right Whale Consortium annual report card.
- Plummer, M. (2003). JAGS: A program for analysis of Bayesian graphical models using Gibbs sampling. *Proceedings of the 3rd International Workshop on Distributed Statistical Computing*, 20–22.
- R Core Team. (2017). R: A language and environment for statistical computing. Vienna, Austria: R Foundation for Statistical Computing. Retrieved from <https://www.r-project.org/>
- Reeves, R. R., Read, A. J., Lowry, L., Katona, S. K., & Boness, D. J. (2007). Report of the North Atlantic Right Whale Program Review. In *Report prepared for the Marine Mammal Commission* (pp. 1–69).
- Rice, W. R., & Harder, J. D. (1977). Application of multiple aerial sampling to a mark-recapture census of white-tailed deer. *The Journal of Wildlife Management*, 41, 197–206.
- RStudio Team. (n.d.). RSTUDIO: Integrated development environment for R. Boston, MA: RStudio Inc. Retrieved from <https://www.rstudio.com/>
- Schaeff, C. M., Kraus, S. D., Brown, M. W., & White, B. N. (1993). Assessment of the population structure of western North Atlantic right whales (*Eubalaena glacialis*) based on sighting and mtDNA data. *Canadian Journal of Zoology*, 71(2), 339–345. doi:10.1139/z93-047
- Slate, J., Marshall, T. C., & Pemberton, J. M. (2000). A retrospective assessment of the accuracy of the paternity inference program CERVUS. *Molecular Ecology*, 9, 801–808. doi:10.1046/j.1365-

294X.2000.00930.x

Weinrich, M. T., Kenney, R. D., & Hamilton, P. K. (2000). Right whales (*Eubalaena glacialis*) on Jeffrey's Ledge: a habitat of unrecognised importance? *Marine Mammal Science*, *16*(2), 326–337.
doi:10.1111/j.1748-7692.2000.tb00928.x

Wilson, B., Hammond, P. S., & Thompson, P. M. (1999). Estimating size and assessing trends in a coastal bottlenose dolphin population. *Ecological Applications*, *9*(1), 288–300. doi:10.1890/1051-0761(1999)009[0288:ESAATI]2.0.CO;2

Wilson, B., Hammond, P. S., & Thompson, P. M. (1999). Estimating size and assessing trends in a coastal bottlenose dolphin population. *Ecological Applications*, *9*(1), 288–300. doi:10.1890/1051-0761(1999)009[0288:ESAATI]2.0.CO;2