

Improving Size Estimates of Open Animal Populations by Incorporating Information on Age

BRYAN F. J. MANLY, TRENT L. McDONALD, STEVEN C. AMSTRUP, AND ERIC V. REGEHR

Around the world, a great deal of effort is expended each year to estimate the sizes of wild animal populations. Unfortunately, population size has proven to be one of the most intractable parameters to estimate. The capture–recapture estimation models most commonly used (of the Jolly–Seber type) are complicated and require numerous, sometimes questionable, assumptions. The derived estimates usually have large variances and lack consistency over time. In capture–recapture studies of long-lived animals, the ages of captured animals can often be determined with great accuracy and relative ease. We show how to incorporate age information into size estimates for open populations, where the size changes through births, deaths, immigration, and emigration. The proposed method allows more precise estimates of population size than the usual models, and it can provide these estimates from two sample occasions rather than the three usually required. Moreover, this method does not require specialized programs for capture–recapture data; researchers can derive their estimates using the logistic regression module in any standard statistical package.

Keywords: logistic regression, capture–recapture, population size estimation

The ages of many long-lived animals can be determined with accuracy and relative ease (e.g., see Scheffer 1950, Stoneberg and Jonkel 1966, Turner 1977, Burbidge 1981, Fay 1982, Harrison 1983, Moen et al. 1991, Stewart et al. 1996, Pyle 1997, Thomas et al. 1997, Calvert and Ramsay 1998, Germano 1998, Sibley 2000). However, this information is not used directly in the standard Jolly–Seber (Jolly 1965, Seber 1965) capture–recapture method of estimating population size, although it is sometimes used indirectly to allow estimates of survival rates, for example, to vary with age. When we investigated whether age information might be better used in estimating the size of animal populations, we found that this can be done quite simply. It results in a method of estimation that (a) provides estimates from just two sample occasions, rather than the three usually required; (b) may considerably improve the precision of estimates; and (c) involves calculations that can be carried out using the logistic regression option available in many standard statistical packages. This means that specialized computer programs such as MARK (White and Burnham 1999) are not required.

In historic terms, our proposed new method is related to the early death-only model for capture–recapture data (Darroch 1959, Jolly 1965), reverse-time models (Nichols et al. 1986, 1998, Pradel 1996, Pradel et al. 1997, Pradel and

Lebreton 1999), and the original method of Manly and Parr (1968). Here, we start by comparing it with the Manly and Parr (M&P) method, because the relationship is straightforward to explain.

Computation of M&P population size estimates is simple. Suppose that a group of animals are known to be alive when a random sample of the population is taken and that a fraction \hat{p} of the animals in this group are captured. This fraction is an estimate of p , the average probability of capture at the time. Hence the population size (N) at the sample time is estimated by

$$(1) \quad \hat{N} = n / \hat{p},$$

where n is the total number of animals captured in the sample.

For large samples, the M&P estimator has little bias, and an estimator of the variance can be calculated with a simple

Bryan F. J. Manly (e-mail: bmanly@west-inc.com) and Trent L. McDonald are statisticians at Western EcoSystems Technology Inc., 217 South First Street, Laramie, WY 82070. Steven C. Amstrup is a research wildlife biologist for the US Geological Survey, Biological Resources Division, Alaska Science Center, 1011 East Tudor Road, Anchorage, AK 99503. Eric V. Regehr is a graduate student in the Department of Zoology, University of Wyoming, Laramie, WY 82070. © 2003 American Institute of Biological Sciences.

equation (Manly 1969). The accurate estimation of N depends on the accurate estimation of p . This, in turn, depends largely on the number of animals known to be alive at the sampling time. In the traditional M&P method, the set of animals known to be alive when a sample is taken is composed of individuals that were captured in an earlier sample and seen or recaptured during a later sampling. Because this determination is based on observed capture histories, the sets of animals known to be alive on the sample occasions are usually small when capture probabilities are low.

The critical point we raise here is that the group of animals known to be alive on a particular occasion can be expanded when age information is available and animals are known to have been in the sampled population since their birth. For example, Pacific walrus (*Odobenus rosmarus divergens*) can be accurately aged on first capture as either 0 (the first year of life), 1 (the second year of life), 2 (the third year of life), or 3+ (three years old or more) on the basis of the length of the tusks in relation to the breadth of the snout (Fay 1982). Similarly, northern spotted owls (*Strix occidentalis caurina*) can be placed into four similar age groups on the basis of plumage (Moen et al. 1991). The age of polar bears (*Ursus maritimus*) can be determined accurately, regardless of when they are first captured, by removing and sectioning a vestigial premolar tooth (Calvert and Ramsay 1998). If a sample of these animals is taken in a given year and their age determined, then we know that the animals aged 1 or more were alive in the previous year, whether they were captured in that year or not. Likewise, the animals aged 2 or more were alive two years before, the animals aged 3+ were alive three years before, and so on.

Thus, in the analysis of capture–recapture data from animals that can be aged to 3 years, like walrus and spotted owl, those known to be alive in sample year X will consist of (a) those sampled both before and after year X , (b) those seen in year $X + 1$ that are aged 1 or more, (c) those seen in year $X + 2$ that are aged 2 or more, and (d) those first seen in year $X + 3$ that are aged 3 or more (counting individual animals only once). Similarly, when we analyze capture–recapture data from animals like the polar bear, which can be aged accurately when they are first captured, the animals known to be alive in sample year X consist of those sampled both before and after X , plus any other animals that are known to have been in the population because of their age in a subsequent capture. Burbidge (1981) used age information in this

way when estimating the size of a population of the endangered western swamp tortoise (*Pseudemys umbrina*) in Australia, but this innovation has apparently not been recognized by other biologists.

The use of age information to estimate population size relies on an implicit assumption that animals enter the population being sampled at age 0. Thus, this method would not be valid for a population that animals entered through immigration at a later age. Applied to such a population, this method would tend to underestimate probabilities of capture and overestimate population sizes. One reviewer suggested that blood parasites, isotopic markers, or something similar also may be available in some cases to provide “natural marks” that confirm that an animal was available for sampling at a particular occasion, though we have not explored those ideas in this article.

The first consequence of incorporating age data into the M&P method is that the size of an open population can be estimated with data from just two years. This is because captured animals aged 2 or more in the second year are known to have been in the population in the first sampling year. The proportion of these animals captured in the first year then provides the estimate \hat{p} needed for equation 1. Considering the needs of most resource managers, being able to derive an estimate in two years rather than the usual three is in itself a considerable benefit.

Including age data is also likely to reduce the standard errors of population size estimates. We have demonstrated this by programming a computer spreadsheet to simulate six years of capture–recapture data. In year 1 of this simulation, there were exactly 1000 animals in the population, all aged 1. For each of the following five years, 100 additional animals were added, each aged 1 when it entered. We assumed that the exact age of each animal could be accurately determined when first captured. Each animal had a survival probability of 0.9 per year, and consequently the population size remained at about 1000 animals. After simulated sampling, we estimated population size with the standard Jolly–Seber method, the standard M&P method, and the M&P method incorporating the known ages of animals. Standard errors were estimated using the usual equations in each case. Table 1 shows that the estimated standard errors were highest for the standard M&P method and lowest for the M&P method using age information. Standard errors were 10 to 51 percent

Table 1. Estimates of the population size N by Jolly–Seber and Manly and Parr methods, with estimated standard errors, for a simulated set of capture–recapture data.

Day	True N	Jolly–Seber estimates		Manly and Parr usual estimates		Manly and Parr, using age	
		Estimated N	Standard error	Estimated N	Standard error	Estimated N	Standard error
1	1000	Not available		Not available		999	47
2	1006	802	92	780	100	937	45
3	1014	1023	105	1102	145	1004	59
4	1014	972	95	1040	128	954	63
5	1024	1225	167	1209	195	1275	150
6	1019	Not available		Not available		Not available	

lower for the M&P method using age information than they were for the Jolly–Seber method. Not surprisingly, the greatest improvements in precision afforded by the age-modified M&P method were realized during the early sample periods, when the largest number of animals was known to be alive.

For long-lived animals, the greatest reduction in standard errors can be expected to occur when there are numerous (e.g., more than six) sampling occasions. The need for large numbers of animals in the “known to be alive” category also suggests that reliable estimates from two sample studies will depend on the samples being large on both sampling occasions.

Current state-of-the-art analyses of open population capture–recapture data build on the conditional likelihood approach formulated by Lebreton and colleagues (1992), with the estimated probabilities of survival and capture being those that maximize the likelihood of the observed data, conditional on the first capture of each animal. Typically, an attempt is made to define these parameters in terms of covariates, as described by Lebreton and colleagues. For example, a logistic function might be used to relate the capture probabilities in different years to a measure of the weather conditions and a measure of the capture effort. In this way the number of parameters to be estimated is reduced, which should lead to improved estimation of the survival and capture probabilities.

Although not previously noted, the traditional M&P method is also a conditional likelihood–based approach, and covariates can be built into the M&P method in a manner similar to that proposed by Lebreton and colleagues (1992). A likelihood function for the traditional M&P method can be constructed by conditioning on both the first and last capture times of each animal (i.e., by regarding the first and last capture times as fixed nonrandom events, which allows us to calculate the probabilities of the captures and noncaptures between these two times, but not at these two times). If age information is available, then a likelihood function can be constructed conditional on the last capture time of an animal. For example, consider an experiment with samples taken every

year for 10 years. Suppose that in year 5 an animal is captured at age 3 and that the animal has the capture pattern shown in table 2. The table indicates not only the age of the animal but also the period in which it was known to be alive, assuming that no capture is possible at age 0. The probability (P) of the observed capture pattern in this case, conditional on the last capture, is then

$$(2) \quad P = (1 - p_3)(1 - p_4)p_5(1 - p_6)(1 - p_7)p_8,$$

where p_i is the probability of capture in year i . This would be the contribution to the overall likelihood function. Incorporating age into M&P analysis and including covariates simplifies modeling because the likelihood function no longer involves survival parameters. Therefore, the problem of deciding on a model for these survival parameters disappears.

Also, incorporating age and conditioning on the last capture simplifies the mechanics of estimation. Equation 2 shows that when the probabilities of capture patterns are combined, the likelihood is the same as would be obtained by simply listing all of the potential captures of each animal when it was known to be alive (excluding the last capture) and determining whether the animal was captured. This means that if the probabilities of capture are related to covariates by means of a logistic function (Pollock et al. 1984, Alho 1990), then any software that performs logistic regression can be used for the calculations. Reducing the estimation to a logistic regression is important, because the interpretation of regression-type relationships is familiar to most biologists. It is also important because researchers no longer need specialized capture–recapture software, such as MARK (White and Burnham 1999), to carry out the calculations.

Suppose, for example, that a 10-year study is carried out, as described for equation 2. Suppose also that the probabilities of capture in year i for animal j are assumed to be given by

$$(3) \quad p_{ij} = \frac{\exp(\beta_0 + \beta_1 X_{ij1} + \beta_2 X_{ij2})}{1 + \exp(\beta_0 + \beta_1 X_{ij1} + \beta_2 X_{ij2})},$$

where X_{ij1} is the average temperature in the i th year and X_{ij2} is the age of the j th animal. This is a logistic function, and the coefficients β_0 , β_1 , and β_2 can be estimated using the logistic regression option in standard statistical packages. Observations simply come from a list of the potential captures (excluding last captures), whether or not these captures occurred, and from the corresponding values of the two covariates. If necessary, the model can also be extended to allow the probability of capture to be a function of the age of an animal.

In general, once a logistic function has been fitted, the probability of capture can be estimated for each of the n_i animals captured on the i th sampling occasion. If these estimated probabilities are denoted by \hat{p}_{ij} for $j = 1, 2, \dots, n_i$, then a Horvitz–Thompson estimator of the population size (N) at the sample time is

$$\hat{N}_i = \sum_{j=1}^{n_i} 1 / \hat{p}_{ij}.$$

Table 2. A possible capture–recapture pattern for an animal captured for the first time at age 3 in year 5.

Year	Capture ^a	Age
1	0	—
2	0	—
3	0	1
4	0	2
5	1	3
6	0	4
7	0	5
8	1	6
9	1	7
10	0	—

a. 1, capture; 0, no capture.

A population size estimate is therefore available for all sampling occasions for which capture probabilities can be estimated. Variances for the estimators can be calculated using the equation of McDonald and Amstrup (2001).

A paper describing the use of the proposed method with polar bear and other data sets is under preparation. We are also preparing papers that describe simulations comparing our proposed method with others currently in use, examine the effects of errors in aging and other failures of assumptions, and consider the extension of the method to multistage models that allow for migration (Schwarz and Seber 1999).

Acknowledgments

We thank Dr. Andrew A. Burbidge of the Department of Conservation and Land Management, Western Australia, for mentioning to B. F. J. M. that he had used information on the ages on capture of western swamp tortoises to improve the estimates of population size with the Manly and Parr method. This led us to examine the feasibility of this idea and subsequently to develop it and produce the general logistic regression approach to population size estimation that is proposed here. We would also like to thank three reviewers whose comments enabled us to improve the article through the incorporation of some extra ideas and the correction of some minor errors in the text.

References cited

- Alho JM. 1990. Logistic regression in capture–recapture models. *Biometrics* 46: 623–635.
- Burbidge AA. 1981. The ecology of the western swamp tortoise *Pseudemys dura umbrina* (Testudines: Chelidae). *Australian Wildlife Research* 8: 203–223.
- Calvert W, Ramsay MA. 1998. Evaluation of age determination of polar bears by counts of cementum growth layer groups. *Ursus* 10: 449–453.
- Darroch JN. 1959. The multiple-recapture census, II: Estimation when there is immigration or death. *Biometrika* 46: 336–351.
- Fay FH. 1982. Ecology and biology of the pacific walrus, *Odobenus rosmarus divergens* Illiger. *North American Fauna* 74: 1–279.
- Germano DJ. 1998. Scutes and age determination of desert tortoises revisited. *Copeia* 1998: 482–484.
- Harrison P. 1983. *Seabirds: An Identification Guide*. Boston: Houghton Mifflin.
- Jolly GM. 1965. Explicit estimates from capture–recapture data with both death and immigration—stochastic model. *Biometrika* 52: 225–247.
- Lebreton J-D, Burnham KP, Clobert J, Anderson DR. 1992. Modeling survival and testing biological hypotheses using marked animals: A unified approach with case studies. *Ecological Monographs* 62: 67–118.
- Manly BFJ. 1969. Some properties of a method for estimating the size of mobile animal populations. *Biometrika* 56: 407–410.
- Manly BFJ, Parr MJ. 1968. A new method of estimating population size, survivorship, and birth rate from capture–recapture data. *Transactions of the Society for British Entomology* 18: 81–89.
- McDonald TL, Amstrup SC. 2001. Estimation of population size using open capture–recapture models. *Journal of Agricultural, Biological, and Environmental Statistics* 6: 206–220.
- Moen CA, Franklin AB, Gutierrez RJ. 1991. Age determination of subadult northern spotted owls in northwest California. *Wildlife Society Bulletin* 19: 489–493.
- Nichols JD, Morris RW, Brownie C, Pollock KH. 1986. Sources of variation in extinction rates, turnover and diversity of marine invertebrate families during the Paleozoic. *Paleobiology* 12: 421–432.
- Nichols JD, Boulinier T, Hines JE, Pollock KH, Sauer JR. 1998. Estimating rates of local extinction, colonization, and turnover in animal communities. *Ecological Applications* 8: 1213–1225.
- Pollock KH, Hines JE, Nichols JD. 1984. The use of auxiliary variables in capture–recapture and removal experiments. *Biometrics* 40: 329–340.
- Pradel R. 1996. Utilization of capture–mark–recapture for the study of recruitment and population growth rate. *Biometrics* 52: 703–709.
- Pradel R, Lebreton J-D. 1999. Comparison of different approaches to the study of local recruitment. *Bird Study* 46 (suppl.): S74–S81.
- Pradel R, Johnson AR, Viallefant A, Nager RG, Cezilly F. 1997. Local recruitment in the greater flamingo: A new approach using capture–mark–recapture data. *Ecology* 78: 1431–1445.
- Pyle P. 1997. *Identification Guide to North American Birds*. Bolinas (CA): Slate Creek Press.
- Scheffer VB. 1950. Growth layers on the teeth of pinnipedia as an indication of age. *Science* 112: 309–311.
- Schwarz CJ, Seber GAF. 1999. A review of estimating animal abundance, III: *Statistical Science* 14: 427–456.
- Seber GAF. 1965. A note on the multiple-recapture census. *Biometrika* 52: 249–259.
- Sibley DA. 2000. *The Sibley Guide to Birds*. New York: Alfred A. Knopf.
- Stewart REA, Stewart BE, Stirling I, Street E. 1996. Counts of growth layer groups in cementum and dentine in ringed seals (*Phoca hispida*). *Marine Mammal Science* 12: 383–401.
- Stoneberg RP, Jonkel CJ. 1966. Age determination of black bears by cementum layers. *Journal of Wildlife Management* 30: 411–414.
- Thomas RB, Beckman DW, Thompson K, Buhlmann KA, Gibbons JW, Moll DL. 1997. Estimation of age for *Trachemys scripta* and *Deirochelys reticularia* by counting annual growth layers in claws. *Copeia* 1997: 842–845.
- Turner JC. 1977. Cemental annulations as an age criterion in North American sheep. *Journal of Wildlife Management* 41: 211–217.
- White GC, Burnham KP. 1999. Program MARK: Survival estimation from populations of marked animals. *Bird Study* 46 (suppl.): S120–S138.