Mortality Estimation: Biased Results from Unbiased Ages

Jay Barlow

U.S. National Marine Fisheries Service, Southwest Fisheries Center, P.O. Box 271, La Jolla, CA 92038, USA

Barlow, J. 1984. Mortality estimation: biased results from unbiased ages. Can. J. Fish. Aquat. Sci. 41: 1843–1847.

Estimates of mortality rates from age distributions are biased by imprecision in age estimation, even if age estimates are unbiased. I have derived a method for predicting the magnitude of this bias from information on the precision of age determination. Monte Carlo simulations show that bias can be accurately predicted. The commonly used Chapman–Robson mortality estimator is shown to be robust to imprecision in age determination if all age-classes are included. Errors are likely, however, if one or more age-classes are excluded or if other mortality estimators are used. Biases can be corrected if the distribution of age-estimation errors is known.

L'estimation des taux de mortalité d'après les distributions des âges est biaisée en raison de l'imprécision de la détermination de l'âge, même si les estimations des âges ne sont pas biaisées. Nous avons mis au point une méthode par laquelle on peut connaître l'importance de ce biais d'après des renseignements sur la précision de la détermination de l'âge. Les simulations de Monte Carlo révèlent qu'il est possible de déterminer le biais avec exactitude. Nous avons constaté que l'estimation de Chapman–Robson, qui est couramment employée pour déterminer la mortalité, peut absorber l'imprécision de la détermination de l'âge lorsqu'on emploie toutes les classes d'âge. Toutefois, il est probable qu'il y ait des erreurs si l'on exclut une ou plusieurs classes d'âge ou si l'on emploie d'autres estimateurs pour déterminer la mortalité. Il est possible de corriger le biais si l'on connaît la distribution des erreurs des estimations de l'âge.

Received March 2, 1984
Accepted August 22, 1984
(J7709)

Reçu le 2 mars 1984
Accepté le 22 août 1984

Age–frequency distributions have frequently been used to estimate mortality rates via an application of stable population theory. Many methods have been proposed for extracting this mortality information from age distributions (Deevey 1947; Chapman and Robson 1960; Caughley 1966; Barlow 1982). All of these methods have assumed that ages could be determined without error. In fact, error in age estimation is a very real problem (Beamish and Fournier 1981; Chang 1982). Clearly, mortality rates could be biased if age estimation is biased. Here I address a slightly different question: can biased mortality estimates result from imprecise, but unbiased ages?

The assumptions required for estimating mortality rate (or its complement, survival rate) from age frequencies are stringent. Stated briefly, a population must be in stable age distribution and its growth rate must be known. These assumptions are treated in detail elsewhere (Caughley 1966; Barlow 1982). Despite stringent assumptions, use of these methods is common in fishery, marine mammal, and other wildlife research. The assumption of accurate age determinations has always been implicit, though usually unspoken.

Given a population with a zero growth rate, a stable age distribution, and a constant mortality rate with age, Chapman and Robson (1960) have derived a minimum variance, unbiased estimator of that population's survival rate based solely on sampled age frequencies. Because the Chapman–Robson (C–R) method is unbiased in cases where aging is deterministic, this method was chosen to examine potential biases when errors occur in age determination. Standard errors of the predicted survival rates can also be estimated from the C–R method.

The C–R survival rate estimator can be expressed in vector notation as

$$(1)\quad \hat{s} = \boldsymbol{n}\cdot\boldsymbol{v}' \Big/ \left(\sum_{i=1}^{\infty} n_i + \boldsymbol{n}\cdot\boldsymbol{v}' - 1\right)$$

where $\boldsymbol{n}$ is the vector of observed frequencies by age-class and $\boldsymbol{v}$ is a vector defined such that $v_i = i - 1$. This survival rate estimator can be reformulated in terms of two parameters: sample size and a scaled mean age (Chapman and Robson 1960, Eq. 4). It is trivial to show that an unbiased error in estimating ages will not affect the expected value of the scaled mean age. Imprecision in estimating ages will, therefore, not affect the expected value of C–R survival rate if all age-classes are included. Frequently the first several age-classes are excluded because their abundance cannot be estimated without bias (e.g. Barlow 1982, Chap. 1). The exclusion of the first age-classes can introduce a bias in the estimation of survival rates when age estimation is subject to errors. In this paper, I derive an expression that predicts the magnitude of this bias and use Monte Carlo simulations to verify the ability of this expression to predict bias accurately.

I only consider the special case for which age estimation is unbiased. In fact, there are few situations where the probability of counting too many annual layers would equal the probability of counting too few. It will be assumed that these types of errors can be corrected with calibration studies.

Methods

To remain consistent with the terminology of Chapman and Robson (1960) and common usage, ages and errors in ages are represented as discrete integers (days, years, etc.). The distribution of aging errors is seldom uniform over all ages. Most commonly, young animals can be aged with greater precision than older animals. The probability distribution of age estimates for a given age-class can be given as columns of a transition probability matrix (e.g. Table 1). The (i, j)th element of such a matrix, $\boldsymbol{P}$, would represent the probability that an individual in age-class j would be interpreted as belonging to age-class i. The product of the transition matrix times the true age distribution of a sample, $\boldsymbol{n}$, would thus give the age distribution that might be expected to be observed given $\boldsymbol{n}$, $\boldsymbol{m}$:

$$(2)\quad \boldsymbol{P}\cdot\boldsymbol{n} = \boldsymbol{m}.$$

The observed age distribution would be expected to differ from the true age distribution unless (trivially) $\boldsymbol{P}$ is the identity matrix, or (miraculously) $\boldsymbol{n}$ is an eigenvector of $\boldsymbol{P}$ with a unit eigenvalue.

When errors in aging are considered, the expected C–R survival rate could be estimated by substituting $\boldsymbol{m}$ for $\boldsymbol{n}$ in Equation 1:

$$(3)\quad \hat{s} = \frac{\boldsymbol{m}\cdot\boldsymbol{v}'}{\sum_{i=k}^{\infty} m_i + \boldsymbol{m}\cdot\boldsymbol{v}' - 1}$$

where k is the first age-class included in the estimation and $\boldsymbol{v}$ is defined such that $v_i = 0$ if $i \leqslant k$, and $v_i = i - k$ if $i > k$.

In this study, the expected value of $\hat{s}$ was calculated for a range of values of the actual survivorship ($s = 0.25$–0.75) and the first age-class to be included ($k = 1$–10). The expected age distribution of a sample was estimated as the geometric series corresponding to the true survivorship. The distribution of aging errors was determined by the probability transition matrix (Table 1). For all ages, the aging errors were symmetric about the true age; hence, the expected age of any individual would equal its true age. Biases due to aging errors were estimated as the difference between the survival rate calculated using Equations 2 and 3 and the true survival rate.

TABLE 1. Probability distributions of aging errors by age-class. Elements indicate the probability that an individual with a given true age will be assigned to the indicated categories of estimated age. Note that errors are symmetrically distributed about the principle diagonal within columns. Note also that aging precision is assumed constant for individuals older than 6.

Estimated age	True age																
	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	·	· ·
1	1.0	0.1															
2		0.8	0.2	0.1	0.1	0.1											
3		0.1	0.6	0.2	0.1	0.1	0.1										
4			0.2	0.4	0.2	0.1	0.1	0.1									
5				0.2	0.2	0.1	0.1	0.1	0.1								
6				0.1	0.2	0.2	0.1	0.1	0.1	0.1							
7					0.1	0.1	0.2	0.1	0.1	0.1	0.1						
8					0.1	0.1	0.1	0.2	0.1	0.1	0.1	0.1					
9						0.1	0.1	0.1	0.2	0.1	0.1	0.1	0.1				
10						0.1	0.1	0.1	0.1	0.2	0.1	0.1	0.1	0.1			
11							0.1	0.1	0.1	0.1	0.2	0.1	0.1	0.1	0.1		
12								0.1	0.1	0.1	0.1	0.2	0.1	0.1	0.1	·	· ·
13									0.1	0.1	0.1	0.1	0.2	0.1	0.1		
14										0.1	0.1	0.1	0.1	0.2	0.1		
15											0.1	0.1	0.1	0.1	0.2		
·												·			·		
·												·				·	
·												·					·

Monte Carlo simulations were used to test the ability of the described procedures to estimate bias resulting from errors in age determinations. Age distributions were generated by randomly drawing discrete samples (with replacement) from a population with a geometric series distribution. This is analogous to a population with constant recruitment and with a survival rate that does not vary with age. In order to mimic error in age determination, a random integer was added to or subtracted from each age selected from the population. Again the distribution of errors was taken from Table 1.

For purposes of illustration, mean values of the C–R estimate were calculated from simulations using three different survival rates (0.3, 0.5, and 0.7) and three different starting ages (1, 2, and 3). These fall within the range of values used in the deterministic approach. The simulation was run with 1000 age distributions (with errors) for each of these nine permutations. Sample sizes for "aged" individuals of all age-classes were set so that the nominal sample for survival rate estimation was 200. Survival rates were only estimated for those individuals whose "apparent" age was greater than or equal to a given starting age (1, 2, or 3); hence, sample size varied slightly about this nominal value due to stochastic effects. Means and standard deviations of survival rates were calculated for each set of 1000 age distributions.

Results

The expected survival rate estimates from the C–R method are shown in Fig. 1 for a range of true survival rates and number of excluded age-classes. As predicted, the method is unbiased if all age-classes are included. If the initial age-class(es) are excluded, an appreciable bias can result. For the given error matrix, the magnitude of this bias is greater when survival rates are low and smaller when survival rates are high. When the first included age-class is 10 or greater, the bias disappears completely.

The results of the Monte Carlo simulation confirm these results. The average survival rates estimated using the C–R method are given in Table 2 for each of the nine combinations of true survival rates and starting ages. Each represents the average estimate from 1000 randomly selected age distributions with simulated error in age determination. Also tabulated are the survival rates predicted using Equations 2 and 3 and the expected (geometric series) age distributions. The magnitude of the bias observed in the simulations was in good agreement with the bias predicted from Equation 3. Observed standard errors in the survival rate estimates were only trivially different from those predicted by the C–R formulations (J. Barlow, unpubl. data).

Discussion

Age distributions typically resemble a declining exponential function. Errors in aging will usually change the shape of such a

TABLE 2. Results of Monte Carlo simulations of survival rate estimation with errors in age determinations. Mean survival rate estimates are averages of 1000 C–R estimates from randomly selected and transformed age distributions. Predicted survival rate estimates are from Equations 2 and 3, using a geometric series age distribution (corresponding to the actual survival rate) for n.

	First age-class included		
	1	2	3
s	0.3	0.3	0.3
Mean $\hat{s}$	0.2993	0.3492	0.3589
Predicted $\hat{s}$	0.3000	0.3490	0.3597
s	0.5	0.5	0.5
Mean $\hat{s}$	0.4995	0.5254	0.5452
Predicted $\hat{s}$	0.5000	0.5250	0.5446
s	0.7	0.7	0.7
Mean $\hat{s}$	0.6996	0.7079	0.7221
Predicted $\hat{s}$	0.7000	0.7090	0.7225

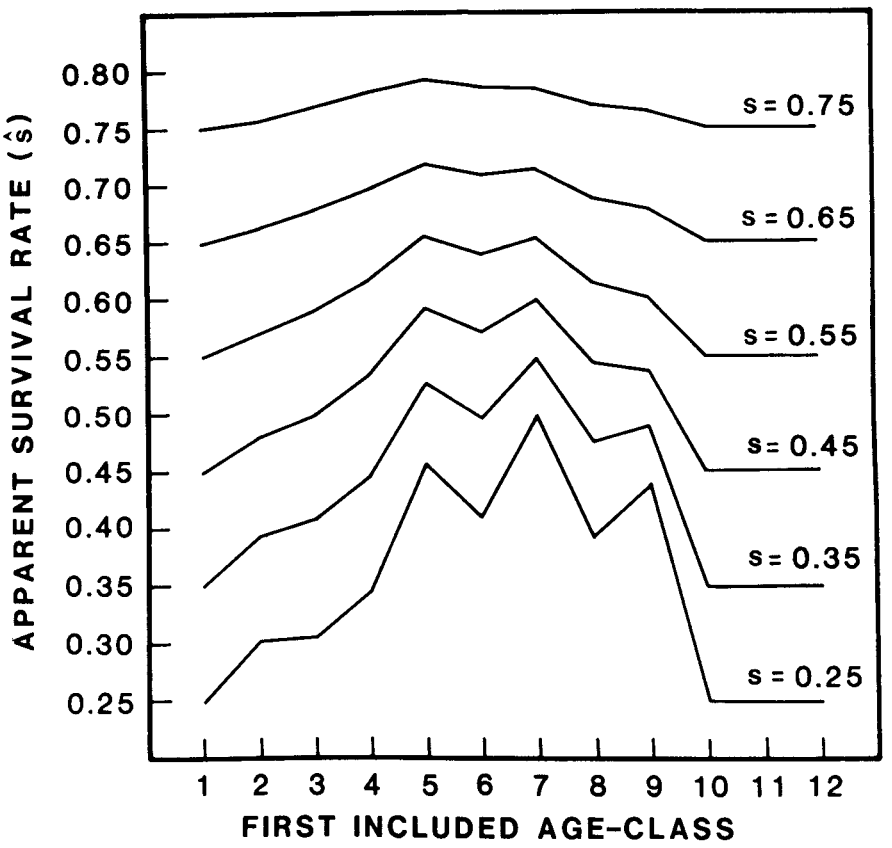

FIG. 1. Apparent survival rates from the C–R method as a function of the first age-class to be included in the estimation. Each line represents a different value of the true survival rate. Errors in age determination are as given in Table 1.

function. It is somewhat fortuitous that the C–R estimator is robust to such changes (if errors are unbiased and all age-classes are included). In the example used here, the method was also unbiased for age-classes 10 and greater. This age is the greatest possible "apparent" age corresponding to the true age at which aging precision becomes constant (Table 1). The C–R method is thus also unbiased in cases where aging precision does not vary with age.

The implications of these results go far beyond the simple examples given here. From information presented thus far, it would seem obvious that bias can be avoided by using the C–R estimator and including all age-classes for which age determination is imprecise. Unfortunately, this solution has a greater conceptual than practical value.

The C–R survival estimator is valid only in the special case where survival rates are constant with age. The decision to use this method is typically based on whether the observed age frequencies fit an expected geometric series age distribution. Robson and Chapman (1961) have formulated a stepwise test that allows the elimination of age-classes that do not meet this expectation. One problem is that errors in age estimation change the shape of the observed age–frequency distribution. Although the true age distribution of a sample may follow a geometric series, the observed distribution may be significantly different, and thus might result in a decision not to use the C–R method or to erroneously eliminate the first age-class(es).

To demonstrate this, Monte Carlo simulations were again used, with the first age-class being chosen by a stepwise chi-square test (Robson and Chapman 1961). Survival rates of 0.3, 0.5, and 0.7 were used, a simulated sample of 1000 aged individuals was drawn, and random errors were introduced in age estimates as before. Using the method of Robson and Chapman to determine the first age-class to be included, mean estimates of survival rates were 0.353, 0.532, and 0.715, respectively, for s = 0.3, 0.5, and 0.7, indicating an appreciable positive bias. Ironically, tests that are intended to avoid biases may introduce new biases.

A different problem exists if survival rates change with age: the C–R method is not valid. Most techniques for estimating age-varying survival rates depend on the slope of an age–frequency curve (Caughley 1966; Siler 1979; Barlow 1982, Chap. 2). Because errors in aging typically affect the shape of the observed age–frequency distributions (hence, their slopes), these methods will be biased even if all age-classes are included.

Bias in survival rate estimation resulting from aging errors is thus a complex problem. A general solution for all methods of survival rate estimation does not exist. For the C–R method, a simple approach can be used to estimate expected bias. This

approach can be visualized in Fig. 1. For a given error matrix, $\boldsymbol{P}$, the true survival rate can be interpolated from this figure if the apparent survival rate and initial age-class are specified.

A similar, but more rigorous approach would be to calculate expected age–frequency distributions for a range of true survival rate values. These true-age distributions are then transformed by $\boldsymbol{P}$ to give estimated-age distributions. The probabilities that the observed age distribution could have been drawn from each of these estimated-age distributions are calculated, and $\hat{s}$ is estimated as the value corresponding to the transformed age distribution that maximizes this probability.

Acknowledgments

I thank D. Chapman, J. Cooke, D. DeMaster, T. Gerrodette, D. Goodman, A. MacCall, R. Methot, J. Pella, and S. Reilly for their helpful comments and suggestions.

References

BARLOW, J. 1982. Methods and applications in estimating mortality and other vital rates. Ph.D. dissertation, University of California, San Diego, CA. 177 p.

BEAMISH, R. J., AND D. A. FOURNIER. 1981. A method of comparing the precision of a set of age determinations. Can. J. Fish. Aquat. Sci. 38: 982–983.

CAUGHLEY, G. 1966. Mortality patterns in mammals. Ecology 47: 906–918.

CHANG, W. Y. B. 1982. A statistical method for evaluating the reproducibility of age determination. Can. J. Fish. Aquat. Sci. 39: 1208–1210.

CHAPMAN, D. G., AND D. S. ROBSON. 1960. The analysis of a catch curve. Biometrics 16: 354–368.

DEEVEY, E. S. JR. 1947. Life tables for natural populations of animals. Q. Rev. Biol. 22: 283–314.

ROBSON, D. S., AND D. G. CHAPMAN. 1961. Catch curves and mortality rates. Trans. Am. Fish. Soc. 90: 181–189.

SILER, W. 1979. A competing-risk model for animal mortality. Ecology 60: 750–757.