A Simple Method for Estimating Survival Rate from Catch Rates from Multiple Years

JOHN M. HOENIG AND TODD GEDAMKE*[1]

*Virginia Institute of Marine Science, College of William and Mary,
Post Office Box 1346, Gloucester Point, Virginia 23062, USA*

Abstract.—Survival rates can be estimated from annual surveys by tracking the abundance of one or more cohorts, as measured by catch per unit of sampling effort, from one year to the next. However, it can be difficult to attain reasonable precision unless sampling effort is extensive. Indeed, estimates of survival exceeding 100% are not infrequently obtained. We show that data from several years can be analyzed simultaneously to obtain a single estimate of survival under the assumption that survival is constant over the period analyzed. The method requires that only a single cohort be identified and separated from the other age-groups. Thus there are minimal data requirements. Estimates of the survival of goosefish *Lophius americanus* obtained by this method compare favorably with estimates obtained by analyzing changes in mean length over time.

Annual survival rate, *S*, can be estimated from annual surveys in various ways. Most methods are based on the relationships

$$N_{a+1,\,t+1} = SN_{a,t}$$

and

$$I_{a,t} = qN_{a,t},$$

where $N_{a,t}$ is the number of animals of age *a* alive at the time of the survey in year *t*, $I_{a,t}$ is the expected value of an index of abundance of the animals of age *a* in year *t*, and *q* is the catchability coefficient. The survival rate for age *a* in year *t* can thus be estimated as

$$\hat{S}_t = \frac{\hat{N}_{a+1,\,t+1}}{\hat{N}_{a,t}} = \frac{\hat{I}_{a+1,\,t+1}}{\hat{I}_{a,t}},$$

where the "hat" symbol denotes an estimate. This method was used as far back as 1934 (Graham 1934, 1938). When several years of data are available, an index of abundance can be regressed on the previous year's value for the same cohort. The assumptions are that survival and catchability (more properly, availability to the survey gear) are constant over time and age and that the ages of the animals observed in the survey can be determined.

Sometimes it is difficult to determine ages for older animals. However, it may be possible to identify one age-group, say age 3, and divide the survey catch into animals of that age and animals above that age. If one age-group can be identified and separated from the others and two years of data are considered, then the survival rate can be estimated by

$$\hat{S}_t = \frac{\hat{N}_{\ge a+1,\,t+1}}{\hat{N}_{\ge a,t}} = \frac{\hat{I}_{\ge a+1,\,t+1}}{\hat{I}_{\ge a,t}}, \qquad (1)$$

where the notation $\ge a$ means all ages greater than or equal to *a*. That is, age-groups are pooled and the abundance of a group of cohorts is followed from one year (when they are age *a* and above) to the next (when they are age *a* + 1 and above). Heincke (1913, cited in Ricker 1975) was apparently the first to suggest pooling data over ages, although he formulated the estimation procedure in terms of a cross-sectional catch curve (age composition observed in one year) rather than as a longitudinal catch curve (changes in abundance of specified cohorts monitored over time). Gulland (1955) pooled ages and computed survival from the change in abundance of a group of cohorts from one year to the next. A generalization of this to include more than two years of data has not been given previously and is developed below.

Development of Multiyear Estimators

We consider just two age-groups in the population: recruits and previously recruited animals. Recruits are defined to be those animals that will join the previously recruited animals the next year (if they survive the year). The relationship between the number of previously recruited animals in year *t* + 1 (N_{t+1}) and the number of recruits (R_t) and previously recruited animals (N_t) in year *t* is

$$N_{t+1} = SN_t + \phi^* R_t, \qquad t = 1, \ldots, T-1,$$

TABLE 1.—Stratified mean number of goosefish caught per tow in surveys off the northeastern United States, by size. The surveys were conducted in the fall of each year.

Survey year	Northern region		Southern region	
	>40 cm	>30 cm	>44 cm	>33 cm
1963	1.035	1.218	0.266	0.330
1964	0.777	0.938	0.311	0.333
1965	1.137	1.384	0.489	0.492
1966	0.172	0.234	0.189	0.189
1967	0.225	0.272	0.259	0.259
1968	0.228	0.307	0.411	0.411
1969	0.266	0.282	0.375	0.389
1970	0.120	0.168	0.439	0.451
1971	0.764	1.081	0.246	0.282
1972	0.289	0.558	0.455	0.476
1973	0.128	0.167	0.250	0.270
1974	0.349	0.442	0.260	0.276
1975	0.196	0.263	0.370	0.409
1976	0.231	0.253	0.577	0.616
1977	0.188	0.192	0.495	0.539
1978	0.258	0.357	0.408	0.432
1979	0.234	0.451	0.440	0.481
1980	0.344	0.577	0.338	0.346
1981	0.115	0.254	0.130	0.130
1982	0.430	0.577	0.341	0.418
1983	0.201	0.222	0.379	0.445
1984	0.234	0.336	0.234	0.305
1985	0.107	0.168	0.411	0.475
1986	0.048	0.099	0.189	0.232
1987	0.086	0.120	0.245	0.274
1988	0.168	0.274	0.217	0.260
1989	0.084	0.120	0.180	0.232
1990	0.148	0.254	0.219	0.349
1991	0.077	0.147	0.294	0.406
1992	0.042	0.139	0.234	0.315
1993	0.109	0.168	0.219	0.356
1994	0.079	0.204	0.344	0.709
1995	0.115	0.168	0.308	0.502
1996	0.182	0.224	0.208	0.342
1997	0.116	0.172	0.217	0.323
1998	0.053	0.157	0.157	0.376
1999	0.105	0.227	0.528	0.954
2000	0.159	0.251	0.588	0.924
2001	0.377	0.523	0.501	0.776
2002	0.238	0.325	0.575	0.660

where ϕ^* and S are the survival rates of the recruits and previously recruited animals, respectively, and T is the number of years of survey data. In terms of the expected values of indices of abundance, we have

$$I_{t+1} = SI_t + \phi r_t, \quad t = 1, \ldots, T-1, \quad (2)$$

where r_t is the expected index of recruits in year t and ϕ subsumes the survival of recruits and the selectivity of the survey gear for recruits. If the indices I_t are independent, then equation (2) is in the form of a multiple linear regression with no intercept. Parameter estimates can be found easily by minimizing the sum of squared deviations between observed indices, $\hat{I}_{t+1}$, and the predictions obtained from the previous year's indices, $\hat{I}_t$ and $\hat{r}_t$, that is by minimizing

$$\sum_{t=1}^{T-1}(\hat{I}_{t+1} - \hat{S}\hat{I}_t - \hat{\phi}\hat{r}_t)^2. \quad (3)$$

The model in equation (2) is quite general because it allows recruits to have a different catchability or survival rate (or both) than the previously recruited animals. However, the estimates of S and ϕ may be highly negatively correlated and unstable unless appreciable contrast is observed in the recruitment over time. Therefore, it will often be necessary to adopt additional assumptions.

An alternative is based on the idea that the parameter ϕ is likely to be close to the value of S. Recruits may have a higher natural mortality than previously recruited animals but probably have a lower fishing mortality and a lower catchability so that, on balance, it may be reasonable to set ϕ equal to S to obtain an estimate of survival. Thus, equation (2) would be replaced by

$$I_{t+1} = S(I_t + r_t), \quad t = 1, \ldots, T-1. \quad (4)$$

Example: Goosefish

Goosefish *Lophius americanus* are captured in the annual groundfish trawl survey conducted by the National Marine Fisheries Service's Woods Hole Laboratory in the fall of each year. However, the survey was not designed to sample this species and the catches are always low, ranging from 14 to 196 animals per survey. Despite this, the mean length data from the survey have proved useful for estimating mortality rates even though the mean lengths vary greatly from year to year (see Gedamke and Hoenig 2006). Estimates were made for two regions: the northern management area, comprising the Gulf of Maine, southern New England, and Georges Bank, and the southern management area, comprising the waters from Rhode Island to North Carolina. It is of interest to see how well those values are reproduced when mortality is estimated from catch rates (Table 1) instead of mean lengths.

Goosefish are believed to be fully vulnerable to the survey trawl when they reach 30 cm in total length (NEFSC 2002). We determined the first fully vulnerable age-class using the von Bertalanffy growth models developed by the National Marine Fisheries Service. The growth parameters for the northern region are as follows: $L_\infty = 126.0$ cm and $K = 0.1080$/year; those for the southern region are 129.2 cm and 0.1198/year (NEFSC 2002). No values were given for the parameter t_0, so a value of 0.0 years was assumed for both regions. It is seen that fish from age 2.5 to age 3.5 have predicted lengths of 29.8–39.7 cm in the northern region. Consequently, fish in the size range 30–40 cm are taken to be the recruits and all fish above 40 cm are
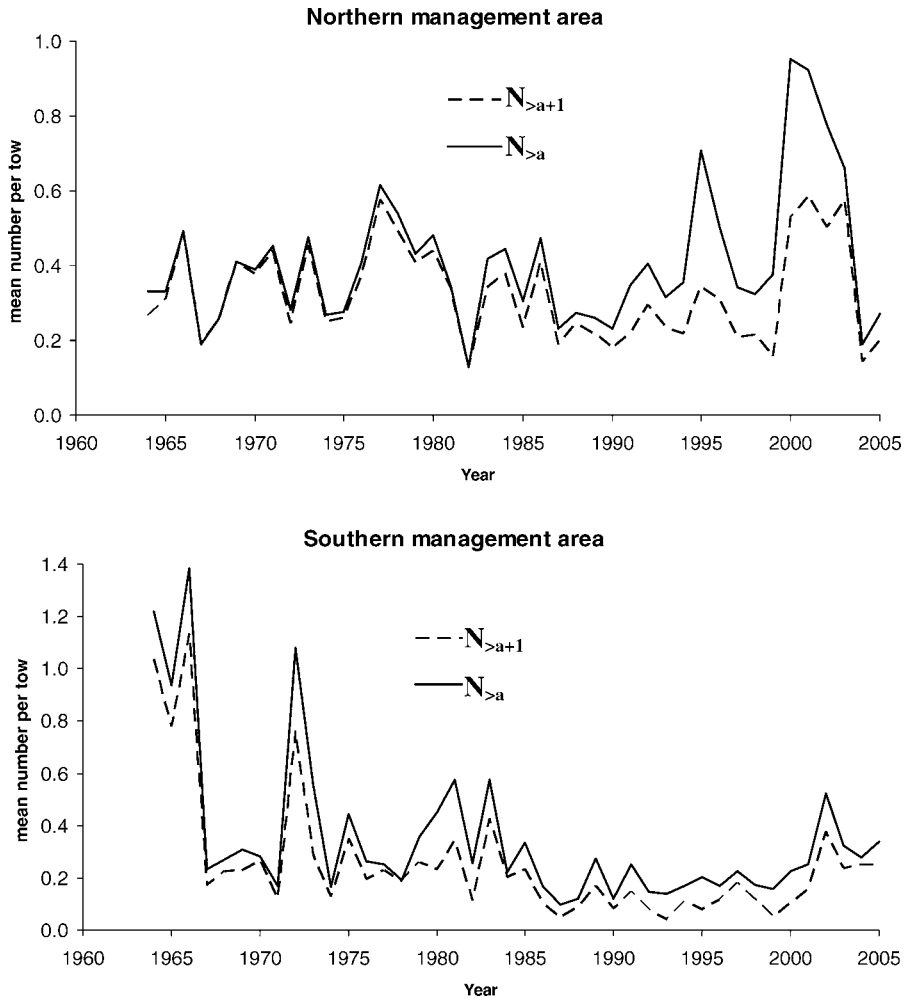
FIGURE 1.—Catch rates of goosefish in National Marine Fisheries Service groundfish surveys in two management regions off the northeastern coast of the United States. The dashed line denotes catch rates of previously recruited fish ($N_{>a+1}$); the solid line denotes catch rates of recruiting and previously recruited fish combined ($N_{>a}$).

considered previously recruited animals. For the southern region, the recruits are fish in the range 33–44 cm.

We apply equation (1) to the catch rate data in Table 1 and Figure 1 to obtain annual estimates of survival rate and then convert these to estimates of the instantaneous rates of total mortality, Z, according to the formula $Z = -\log_e(S)$ (Figures 2, 3). Data from the 1963 and 1964 surveys give rise to an estimate of survival between the surveys, that is, between fall 1963 and fall 1964. We refer to this as the survival in 1964 because most of the time interval is in 1964. Not unexpectedly, the results are highly variable and often

infeasible (i.e., estimates of the mortality rate are negative).

We computed the arithmetic mean of the results from equation (1) over the periods of stable mortality identified by Gedamke and Hoenig (2006) from their analysis of mean sizes (Table 2; Figures 2, 3). Gedamke and Hoenig found that the periods of stable mortality identified through their analysis of length data corresponded well with recorded changes in the fishery (i.e., price rises and concomitant increases in landings). Averaging the annual estimates gives results roughly similar to those obtained by Gedamke and Hoenig, but the agreement is strong only in one comparison.
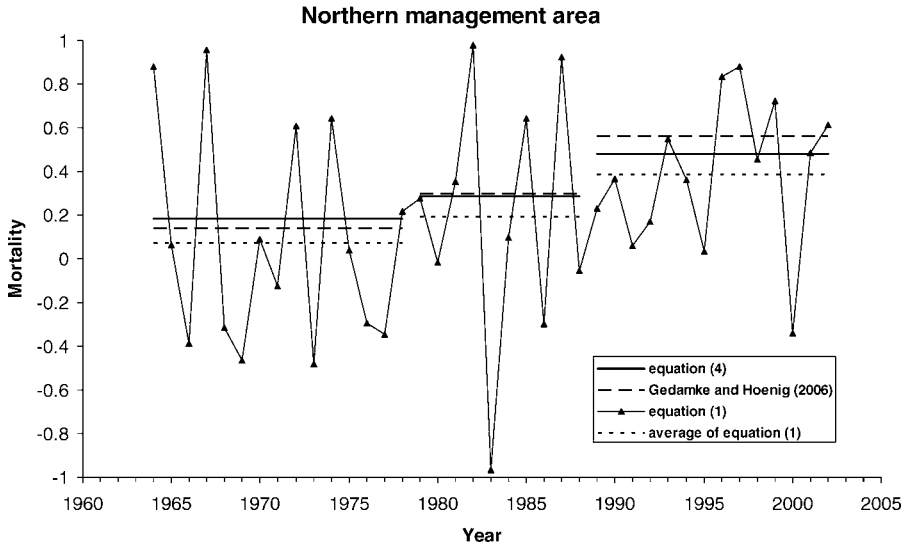
Northern management area

FIGURE 2.—Estimates of the total instantaneous mortality rate ($Z = -\log_e[S]$, where S is the survival rate) for the northern management area in three time periods obtained by Gedamke and Hoenig (2006) from the length frequencies of the survey catches. Also shown are the annual estimates of Z obtained from equation (1) along with their averages over the three time periods and the results of applying equation (4).

We applied the estimators based on equations (2) and (4) to those same ranges of years. The results based on equation (2) were poor and are not shown here.

The results of applying the estimator based on equation (4) are extremely close to the results of Gedamke and Hoenig (2006) for four of the five comparisons. Only for the period 1963–1976 for the southern region was there a large discrepancy ($Z = 0.33$/year based on mean lengths and 0.55/year based on equation 4).
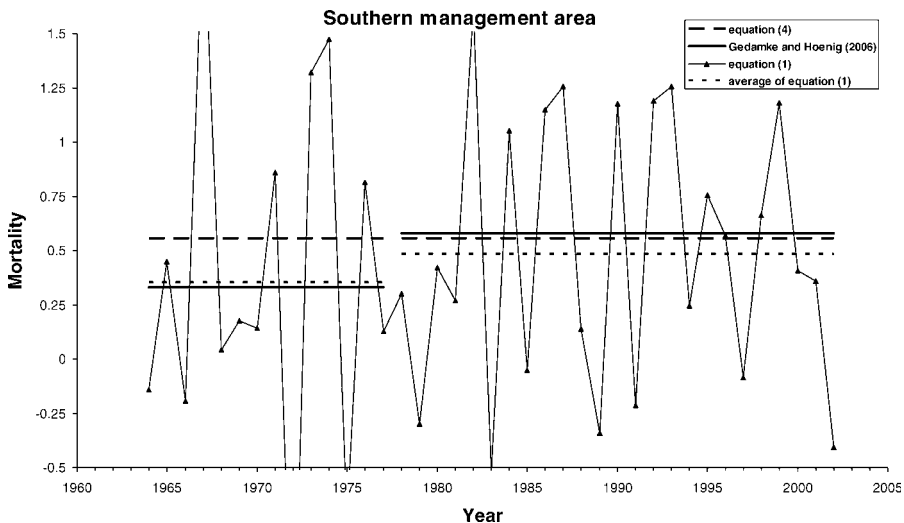
Southern management area

FIGURE 3.—Estimates of the total instantaneous mortality rate for the southern management area in two time periods obtained by Gedamke and Hoenig (2006) from the length frequencies of the survey catches. Also shown are the annual estimates obtained from equation (1) along with their averages over the two time periods and the results of applying equation (4). Note that two extreme values of the estimator based on equation (1) are not shown.

TABLE 2.—Estimates of the total mortality rate for goosefish in two management areas by three methods. The estimates from lengths are from Gedamke and Hoenig (2006); see text for equations.

Period	From lengths	From equation (4)	Average from equation (1)
Northern management area			
1963–1977	0.14	0.18	0.07
1978–1988	0.29	0.28	0.19
1989–2002	0.55	0.48	0.39
Southern management area			
1963–1976	0.33	0.55	0.37
1977–2002	0.58	0.56	0.47

Sensitivity Analysis

Assumption that $\phi = S$.—A key assumption in applying equation (4) is that the parameter ϕ is equal to S. We specify departures from this assumption by the equation $\gamma S = \phi$. Then equation (2) can be written

$$I_{t+1} = SI_t + \gamma Sr_t, \quad t = 1, \ldots, T - 1.$$

The effect of assuming that $\gamma = 1$ when it is really some other value can be determined by multiplying all the recruitment index values by the other value and then reestimating the survival rate. We did this for the northern management region data from 1963 to 1977 and converted the results to the instantaneous rates of total mortality, Z. The computed value of Z is an increasing function of the value of γ and is described by the equation computed $Z = 0.0554\gamma + 0.123$, such that when $\gamma = 1$ the computed Z is 0.1784, as reported in Table 2. If ϕ is really 90% of S (i.e., $\gamma = 0.9$), the estimate of Z would be 0.17286 and the percent difference in results would be

$$100(0.1784 - 0.17286)/0.17286 = 3\%.$$

Thus, a 10% error in the specification of γ results in a 3% error in the estimate of Z. The estimator based on equation (4) is insensitive to model misspecification, at least when applied to the goosefish data from 1963 to 1977.

Effects of systematic errors in age composition.—There are two effects of systematic errors in the specification of age composition. First, if only a portion of the animals in age-group 1 (the recruits) is included in the analysis, the result will be a positive bias in the survival estimator. This is easily seen from equation (1), where the denominator is made smaller by the exclusion of some recruits. Similarly, including some animals that will not reach the size of the second age-group in 1 year in the new-recruits category results in a

negative bias. Similar results hold for the other models (based on equations 2 and 4).

The second consequence of systematic misspecification of age composition is that the estimators are no longer unaffected by recruitment variability. We demonstrate this by considering the estimator in equation (1). Suppose the fraction α of age-group 1 (the recruits) included in the analysis is constant. Then, as recruitment approaches zero, the estimator in equation (1) approaches the survival rate, S. As recruitment increases without bound, the estimator tends to $S/(1 - \alpha)$. If recruitment is overestimated (some animals that will not grow into age-group 2 in 1 year are included in age-group 1) but it approaches zero, then the estimator tends to the true survival rate S. If recruitment is overestimated and it approaches infinity, the estimator is too low and tends to $S/(1 + \alpha)$. These conclusions are justified in the appendix.

Discussion

We have presented a new approach to estimating survival rate from multiple years of survey data. The big advantages of this approach are that one does not need to know the magnitude of the landings and does not have to be able to age the catch beyond identifying the recruits. We presented a general model (equation 2) that allows for recruits to have a different survival or catchability than previously recruited animals. However, in practice it is often necessary to adopt an additional assumption to obtain reasonably precise and stable estimates. The more general model should be less biased than the alternatives but at the expense of greater variance.

For the goosefish example, it appears that the survey catch rates are more variable than the mean lengths and consequently that the precision of the estimates from the new approach is less than those based on the Gedamke and Hoenig (2006) approach based on mean lengths. However, it should be noted that both methods can be applied to data from the trawl survey. For goosefish, it is reassuring that the two approaches gave very similar results for five of the six time–area comparisons. The estimates obtained by averaging the two-year estimates from equation (1) appeared reasonable in most cases. However, averaging a set of numbers that contains nonsensical values (survivals greater than 1.0) may be troubling and hard to justify.

We relied on the model described by equation (4) because the survey catches of goosefish were low and the catch rates highly variable. In cases where a species is better sampled, it may be worthwhile to use the model in equation (2).

Our example involved survey catch rate data. It is

also possible to apply the method to commercial catch rate data provided these represent relatively short periods of time. For example, the catch rate in the first two months of year $t + 1$ can be compared with the catch rate in the same period in year t. Restricting the analysis to short periods of time within a year is necessary for two reasons. First, catch rate in a time interval is proportional to average abundance in that interval. Thus, over an extended period the effort may be large and the commercial catch rate will reflect both the initial abundance and the depletion of the population. The second reason for using a restricted period of time is that catchability, recruitment, and other factors may change seasonally so that catch rate is harder to interpret.

In the appendix we investigate the impact of errors in determining the fraction of the survey catch that will recruit in the next year and place bounds on these errors. In cases where errors in determining the catch composition may be appreciable, the survival estimators are more properly thought of as providing an index of survival rather than absolute survival. That is, the estimators can be used to monitor trends in the survival rate over time rather than to quantify the level of survival.

Acknowledgments

References

Gedamke, T., and J. M. Hoenig. 2006. Estimating mortality from mean length data in nonequilibrium situations, with application to the assessment of goosefish. Transactions of the American Fisheries Society 135:476–487.

Graham, M. 1934. The North Sea cod. Journal du Conseil Permanent International pour l'Exploration de la Mer IX:159–171.

Graham, M. 1938. Growth of cod in the North Sea and use of the information. Rapports et Procès-Verbeaux des Réunions, Conseil Permanent International pour l'Exploration de la Mer 108:57–66.

Gulland, J. A. 1955. Estimation of growth and mortality in commercial fish populations. United Kingdom Ministry of Agriculture and Fisheries, Fisheries Investigations (Series 2) 18(9).

Heincke, F. 1913. Investigations on the plaice. General report 1: plaice fishery and protective regulations, part 1. Rapports et Procès-Verbeaux des Réunions, Conseil International pour L'Exploration de la Mer 17A:1–153.

NEFSC (Northeast Fisheries Science Center). 2002. Report of the 34th Northeast Regional Stock Assessment Workshop: Stock Assessment Review Committee consensus summary of assessments. National Marine Fisheries Service, Reference Document 02-06, Woods Hole, Massachusetts.

Ricker, W. E. 1975. Computation and interpretation of biological statistics of fish populations. Fisheries Research Board of Canada Bulletin 191.

Appendix: Effects of Systematic Errors in Age Composition on Survival Estimates

Referring to Figure A.1, we note that the survival estimator in equation (1) can be viewed geometrically as a ratio of areas

$$\hat{S} = \frac{b_2}{a_1 + b_1}, \tag{A.1}$$

where $\hat{S}$ is the estimated survival, a_1 is the abundance of recruits in year 1, and b_1 and b_2 are the abundances of previously recruited animals in years 1 and 2, respectively. If a portion of the animals that will recruit in year 2, say, αa_1, is excluded from the recruitment, then the estimate of survival will be biased upwards. Similarly, if some animals are included in the recruitment that will not, in fact, recruit in year 2 (a_1 is specified as too large), the estimate will be biased downwards.

We now consider how the magnitude of the recruitment affects this result. We note that
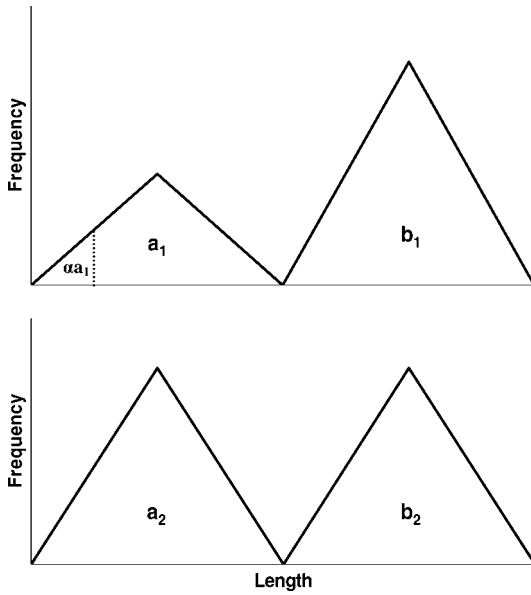
$$b_2 = S(a_1 + b_1) \tag{A.2}$$

when the survival of the recruits is the same as the survival of the previously recruited animals. If a portion of the recruitment, say, αa_1, is excluded from the calculation of survival, then equation (A.1) becomes an erroneous estimate of survival, $\hat{S}_{err}$, given by

$$\hat{S}_{err} = \frac{b_2}{a_1 + b_1 - \alpha a_1} = \frac{S(a_1 + b_1)}{a_1 + b_1 - \alpha a_1}.$$

Suppose a_1 approaches zero. This implies that the amount of recruitment not included in the calculation, αa_1, also approaches zero. Then, $\hat{S}_{err}$ approaches $S(b_1/b_1) = S$. On the other hand, suppose a_1 approaches ∞ while α remains constant. Then

$$\hat{S}_{err} \rightarrow S\left(\frac{a_1}{a_1 - a_1\alpha}\right) = \frac{S}{1 - \alpha}.$$

Similarly, it can be shown that if recruitment is

FIGURE A.1.—Hypothetical length frequency distributions showing two modes representing one identifiable age-class (a_i) and one "composite" or "plus" group consisting of multiple age-classes (b_i) for years $i = 1, 2$. The text discusses the effect on the estimated survival rate of not including the area marked αa_1 as part of recruitment in year 1. The length frequencies are shown as triangular distributions merely for convenience.

overstated by an amount α but recruitment approaches zero, then the erroneous estimator approaches the true survival rate, S. And, under the same circumstances, if recruitment approaches infinity, the survival estimate approaches $S/(1 + \alpha)$.