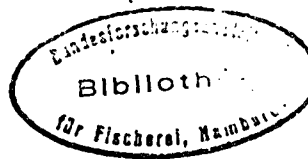


International Council for the
Exploration of the Sea



ICES C.M. 1993/D:52
ref. G+H+J

Using Prior and Current Information to Estimate Age Composition: a new kind of age-length key.

John M. Hoenig

Department of Fisheries and Oceans, P.O. Box 5667, St. John's, Nfld A1C 5X1, Canada

Dennis M. Heisey

Department of Surgery, G5316 Clinical Sciences Center, 600 Highland Avenue,
University of Wisconsin, Madison, WI 53706, USA

R.C. Hanumara

Department of Statistics, University of Rhode Island, Kingston, RI 02881, USA

Age composition is often estimated by measuring the lengths of a large number of fish and aging a small portion of the measured fish. The small sample cross-classified by length and age can be used to estimate classification rates and these, in turn, can be used in conjunction with the estimated length distribution to obtain an estimate of the population age composition. There are two approaches to this problem characterized by the way in which the classification rates are defined. The simplest approach uses estimates of the probability $P(i|j)$ that a fish is actually age i given that the length is j . The more complicated approach uses estimates of the probability $P(j|i)$ that the length is j given that a fish is actually age i . The latter approach involves estimating more parameters and is less precise than the former. However, it avoids the necessity for the cross-classified sample to be from the same population as the population from which the large sample was drawn for estimating the length composition. In this paper, we show that the two approaches can be combined when there are multiple samples. For example, one might have two samples obtained by random sampling of the population, and a third, cross-classified, sample from another population, such as the one in the previous year, with different composition but identical classification probabilities $P(j|i)$. We also show how to modify the method to allow for a fixed number to be aged from each length category.

1. INTRODUCTION

Each year, fishery scientists examine otoliths and other skeletal hard parts from thousands of fish in order to determine the age of individual fish; this information is then used to estimate the age composition in the catches. Typically, lengths are determined for a large sample of fish and the lengths and corresponding ages are determined for a much smaller sample. The length frequency information can be obtained at low cost; it provides information on the age composition in the population since age and length are correlated.

Information is often available from previous years but this prior information is not combined with the current data on ages and lengths in order to estimate the age composition in the current year. This seems a pity since the data from previous years is essentially free and can be used to help interpret the length frequency data from the current year.

In this paper, we review the two basic approaches to age-length keys - forward or classic keys and inverse keys. The forward key requires information from the current year whereas the inverse key can make use of information from previous years. When the forward key is appropriate, it is more efficient (i.e., has lower variance) than the inverse key. This assertion is consistent with experience and can also be justified by the fact that the inverse key involves estimating more parameters than the forward key. (We also have a proof, which is available upon request, that the variance of the forward key is smaller than that of the inverse key.) We show how to combine the two approaches in a single analysis to achieve even greater efficiency. We then generalize the inverse key approach and the combined approach to allow for length stratification. The goal is to use information from previous years to aid in the estimation of age composition in the current year. An added benefit is that information from the current year provides for revised estimates with higher precision of the age composition in previous years.

We illustrate the logic of the two approaches by considering a simple case in which the age and length are determined for each of n fish; each examination assigns an age of 1 or 2 to the fish and a length of 1 or 2 to the fish. The result is a 2×2 cross-classified table as in Figure 1. There is also a sample of size N on which observations are made on only the length.

Figure 1. Notation for the results of determining the length and age of n fish (left) and for determining the length of N additional fish (right).

		length j			length j		
		1	2		1	2	
age i	1	n_{11}	n_{12}	$n_{1.}$	Y_1	Y_2	N
	2	n_{21}	n_{22}	$n_{2.}$			
		$n_{.1}$	$n_{.2}$	n			

1.1 Approach 1 - the forward or classic key

Assume that the n cross-classified fish and the N fish examined for just the length are simple random samples from the same population. Then, the probability $P(i|j)$ that a fish is actually of age i ($i = 1$ or 2), given that it has length j ($j = 1$ or 2), is the same for both samples. One can estimate these conditional probabilities by

$$\hat{P}(i|j) = q_{ij} = n_{ij}/n_{.j}$$

where the $\hat{}$ symbol denotes an estimate and the rest of the notation is as in Figure 1. Denote the 2×2 matrix with elements q_{ij} by Q and the vector of length proportions by

$$E = \begin{bmatrix} \frac{Y_1 + n \cdot 1}{N + n} \\ \frac{Y_2 + n \cdot 2}{N + n} \end{bmatrix} = \begin{bmatrix} e_1 \\ e_2 \end{bmatrix}.$$

Then, intuitively, the age composition A might be estimated by

$$\hat{A} = Q E$$

where $\hat{A} = [\hat{a}_1, \hat{a}_2]^T$ is the vector of estimated proportions at age. Thus,

$$\hat{a}_i = \sum_{j=1}^2 \hat{P}(ij) e_j.$$

In this example, the conditional probabilities are estimated from the first sample and the marginal probabilities for length, $P(j)$, are estimated from both samples as e_j .

This estimator can be shown to be of maximum likelihood (see Tenenbein 1970; Hochberg 1977; Jolayema 1990). It is also an example of stratified random sampling where the units are post-stratified by length (Swensen 1988). Related estimators, which are not fully efficient, are discussed by White and Castleman (1981) and Hand (1986).

In fisheries research, this approach is usually modified slightly. The number of fish aged from each length category, n_j , is fixed by the investigator (see section 3.1). The only change to the estimation procedure is to estimate the length composition, E , by the proportion of fish observed in each class:

$$E = \begin{bmatrix} \frac{Y_1}{N} \\ \frac{Y_2}{N} \end{bmatrix}$$

The estimator is still of maximum likelihood. The variance changes, however.

1.2 Approach 2 - the inverse key

It may happen that the cross-classified sample and the length sample are obtained from different populations. For example, in the first year of a study both the lengths and ages might be recorded for a sample but in the second year only the lengths are recorded. The conditional probabilities $P(i|j)$ from the first year will not be applicable to the results in the second year if the population composition has changed (Kimura 1977; Westrheim and Ricker 1978). To see this, consider the probability that a fish is actually age 1 given that the length is 1. If, in the first year, all fish are age 1 then all of the fish of length 1 will in fact be age 1 ($P(i=1|j=1) = 1$). If, in the second year, none of the fish are age 1 then none of the fish of length 1 will in fact be age 1 ($P(i=1|j=1)$ is now 0).

There may be a way out of this dilemma. The probability that a fish is length j given that it is age i , $P(j|i)$, might not vary with the population composition. Thus, the age composition of the fish population will change each year as a variable number of young fish are recruited into the population and thus the probability that a fish is a certain age given its size, $P(\text{age} = i | \text{length} = j)$, will vary from year to year. On the other hand, the distribution of size about age, $P(\text{length} = j | \text{age} = i)$, should not change much as the population changes in composition except inasmuch as the growth may be somewhat dependent on environmental conditions. It thus may be entirely reasonable to suppose that $\text{Prob}(\text{length} = j | \text{age} = i)$ is constant from one year to the next or one area to the next.

Define the matrix P to have elements p_{ij} given by

$$p_{ij} = n_{ij}/n_i.$$

where the n_{ij} are the cell counts from a cross-classified sample from a *prior* time. Also, define the vector E to have elements

$$E = \begin{bmatrix} Y_1/N \\ Y_2/N \end{bmatrix},$$

that is, the vector E contains estimates of the marginal probabilities $P(j)$ obtained from just the length survey in the *current* year. Then, intuitively, the age composition should be related to the length composition by

$$E = P^T \hat{A}.$$

The age composition can be estimated by premultiplying each side by the generalized inverse of P^T (assuming the number of length classes is \geq the number of age classes). Thus,

$$\hat{A} = (PP^T)^{-1} PE \quad (1)$$

Equation (1) is a least squares estimator.

It can be seen that when the estimates from (1) are feasible, they are maximum likelihood estimates. Clark (1981) developed a fitting procedure which restricts the parameter estimates to the feasible region. Hoenig and Heisey (1986) developed a model with a more realistic error structure in which the uncertainty in both the classification rates and the length composition is accounted for explicitly as functions of the sample sizes.

This general approach has appeared in the applied literature a number of times, e.g., as a hypothetical example of correcting deer age composition (Searle 1966 p. 93-4); as a method of correcting stock composition estimates for mixed fisheries - see Worlund and Fredin (1962), Fukuhara et al. (1962), Berggren and Lieberman (1978), Pella and Robertson (1978), and van Winkle et al. (1988); as a means of estimating prevalence of diseases - see Rogan and Gladen (1978), Greenland and Kleinbaum (1983), Hand (1986); as a correction for misclassification in a fourfold table relating disease status to risk factors (Kleinbaum et al. 1982 and references therein); as a method of correcting estimates of deer harvest composition obtained from hunter reports (D. Ingebrigtsen, MN Department of Natural Resources, pers. comm.); and for converting length-frequency distributions to age-

frequency distributions (Clark 1981; Bartoo and Parker 1983; Kimura and Chikuni 1987; Hoenig and Heisey 1987).

Thus, there are two approaches to using estimates of classification probabilities to convert a vector of length frequencies to an estimate of age composition. Method 1 is straightforward, is well known, and requires that the classification probabilities be estimated from a random sample of the population to which they will be applied. Method 2 is more complicated and involves a backward or inverse type of reasoning. For method 2, the classification rates are conditional on the age rather than on the length. The method has been repeatedly derived in the applied literature but does not appear to be well established in the statistical literature. In the next section, we show how the two methods can be combined.

2. COMBINED METHOD

In the previous section we used matrix notation to make clear the difference between the two general approaches to age-length keys. In order to show how the two approaches can be combined, it is easier to work with the likelihood for the data. The likelihood for the full data set is the product of the likelihoods for each datum.

2.1 Three samples

Assume that we have three samples of fixed size. Sample 1 is a random sample of size n_1 collected during a prior time period or from a nearby location. All n_1 fish were classified according to age and length. Sample 2 is a random sample of size n_2 from the population of interest; all fish in this sample are classified by both variables. Sample 3 is a random sample of size N_2 from the population of interest and all fish are classified according to just the length. The subscript denotes the population (e.g., time period) from which the sample was drawn. We denote the count of fish with age classification i and length classification j in samples 1 and 2 by n_{ij1} and n_{ij2} , respectively. The count of fish in sample 3 with length classification j is denoted by Y_{j2} . We assume that $P(j|i)$ for sample 1 is the same as for samples 2 and 3, and we denote this by $P(j|i)_2$. In general, subscripts on probabilities are used to denote the population or populations to which the probabilities apply.

Likelihood for method 1.

Approach 1 utilizes the information in samples 2 and 3. The likelihood for samples 2 and 3 is the product of independent multinomials and can be written

$$\Lambda_1 \propto \prod_{i=1}^2 \prod_{j=1}^2 [P(i|j)_2 P(j)_2]^{n_{ij2}} \prod_{j=1}^2 P(j)_2^{Y_{j2}}.$$

There are six parameters in Λ_1 but only three parameters need to be estimated: two conditional probabilities and one marginal probability $P(j)_2$. This is because there are three constraints that must be satisfied: $P(i=1|j=1)_2 + P(i=2|j=1)_2 = 1$, $P(i=1|j=2)_2 + P(i=2|j=2)_2 = 1$, and $P(j=1)_2 + P(j=2)_2 = 1$.

The goal is to estimate the proportion $P(i)_2$ that is age i and, by the invariance principle of maximum likelihood estimation, this can be accomplished by

$$\hat{P}(i)_2 = \sum_{j=1}^2 \hat{P}(ij)_2 \hat{P}(j)_2 .$$

Likelihood for method 2.

Approach 2 utilizes the information in samples 1 and 3. The likelihood is again the product of two multinomials

$$\Lambda_2 \propto \prod_{i=1}^2 \prod_{j=1}^2 [P(j|i)_2 P(i)_1]^{n_{ij1}} \prod_{j=1}^2 \left[\sum_{i=1}^2 P(j|i)_2 P(i)_2 \right]^{Y_{j2}} .$$

There are eight parameters in the model but only four parameters need be estimated: one $P(i)_1$ and one $P(i)_2$, and two conditional probabilities. This is because of the constraints $P(i=1)_1 + P(i=2)_1 = 1$, $P(i=1)_2 + P(i=2)_2 = 1$, $P(1|1)_2 + P(2|1)_2 = 1$, and $P(1|2)_2 + P(2|2)_2 = 1$.

Combined likelihood.

The likelihood for all of the data can be written as

$$\Lambda_3 \propto \prod_{i=1}^2 \prod_{j=1}^2 [P(j|i)_2 P(i)_1]^{n_{ij1}} \prod_{i=1}^2 \prod_{j=1}^2 [P(j|i)_2 P(i)_2]^{n_{ij2}} \prod_{j=1}^2 \left[\sum_{i=1}^2 P(j|i)_2 P(i)_2 \right]^{Y_{j2}} .$$

Here, we have rewritten the likelihood for sample 2 (middle) and 3 (right) in terms of $P(j|i)_2$. However, it should be noted that this likelihood reduces to Λ_1 (the classic key) when $n_1 = 0$.

2.2 Combined method - generalizations

It may occur that four samples are available: the three samples discussed in section 2.1 plus a length-frequency sample from the same population as the population from which sample 1 was drawn. We can use this additional sample to obtain improved estimates of $P(i)_1$. This is of interest not only for its own sake but also because the $P(i)_1$ occur in the formulae for the variance-covariance matrix when this is calculated on the basis of expected information.

A general form of the likelihood is presented below which allows for I age classes, J length classes, and K surveys. Here, a survey refers to fish examined from the same time and place. In each survey, both variables are noted on a random sample and, optionally, just the length is noted on another random sample. Denote the number classified as length j in the kth survey for the sample in which just the length is noted by Y_{jk} (note that Y_{jk} can be 0). Then, the general form of the likelihood is proportional to

$$\Lambda_g \propto \prod_{i=1}^I \prod_{j=1}^J \prod_{k=1}^K [P(j|i) P(i)_k]^{n_{ijk}} \prod_{j=1}^J \prod_{k=1}^K \left[\sum_{i=1}^I P(j|i) P(i)_k \right]^{Y_{jk}} .$$

Here, $P(j|i)$ is assumed to pertain to all samples, $P(i)_k$ pertains to all samples within the k th survey, and n_{ijk} is the number of fish cross-classified as ij in the k th survey.

3. ALLOWING FOR STRATIFICATION BY LENGTH

Until now, we have assumed that the fish which are aged are a random sample of the fish which were measured for length. In practice, one is likely to consider the lengths when selecting the fish to age, e.g., one might choose equal numbers of fish from each length class. Indeed, there is good reason to consider the lengths when selecting the sample for aging: otherwise one might obtain by chance a sample in which none of the fish were from a particular length class and one would not be able to estimate some of the classification rates.

3.1 Classic key

Haitovsky and Rapp (1992) modified Approach 1 to allow for fixed numbers from each length category to be aged. As indicated earlier, the estimates of age composition are the same under length stratification as under simple random sampling but the estimates of variance differ. The likelihood is proportional to

$$\Lambda_{1s} \propto \prod_{i=1}^I \prod_{j=1}^J P(i|j)^{n_{ij}} \prod_{j=1}^J P(j)^{Y_j}.$$

3.2 Inverse key

Here, we show how the inverse key (Approach 2) can be modified to allow for stratification by length. For ease of exposition, we revert back to the case where there are 2 ages and 2 lengths; extension to the general case of I ages and J length classes is straightforward. Assume that at a previous time a sample of N_1 fish was randomly selected and measured resulting in Y_{11} fish being classified as length 1 and Y_{21} fish as length 2, $Y_{11} + Y_{21} = N_1$. Suppose further that ages are determined for n_{11} and n_{21} fish of length 1 and 2, respectively. This results in a cross-classified table with fixed column totals of n_{11} and n_{21} with table entries of n_{ij1} where the i indexes the age classification and the j indexes the length classification.

The likelihood for the N_1 fish classified by just the length is simply a binomial

$$\begin{aligned} \Lambda_{N_1} &= \binom{N_1}{Y_{11}} P(j=1)_1^{Y_{11}} P(j=2)_1^{Y_{21}} \\ &= \binom{N_1}{Y_{11}} \left[\sum_{i=1}^2 P(j=1|i) P(i)_1 \right]^{Y_{11}} \left[\sum_{i=1}^2 P(j=2|i) P(i)_1 \right]^{Y_{21}}. \end{aligned}$$

The likelihood for the cross-classified table is the product of two binomials, one for each column

$$\Lambda_{n_{11}, n_{21}} = \binom{n_{11}}{n_{111}} P(i=1|j=1)_1^{n_{111}} P(i=2|j=1)_1^{n_{211}} \binom{n_{21}}{n_{121}} P(i=1|j=2)_1^{n_{121}} P(i=2|j=2)_1^{n_{221}}.$$

Now, in the population at large,

$$P(i|j)_1 = \frac{P(j|i) P(i)_1}{\sum_i P(j|i) P(i)_1}$$

by Bayes rule. Substituting this into the product of the likelihoods Λ_{N_1} and $\Lambda_{n,11,n,21}$ yields the likelihood for the prior data; it is proportional to

$$\begin{aligned} \Lambda_{\text{prior}} \propto & \left[\sum_{i=1}^2 P(j=1|i=1) P(i)_1 \right]^{Y_{11}} \left[\sum_{i=1}^2 (P(j=2|i) P(i)_1) \right]^{Y_{21}} \\ & \left[\frac{P(j=1|i=1)P(i=1)_1}{\sum_{i=1}^2 P(1|i) P(i)_1} \right]^{n_{111}} \left[\frac{P(j=1|i=2)P(i=2)_1}{\sum_{i=1}^2 P(1|i) P(i)_1} \right]^{n_{211}} \left[\frac{P(j=2|i=1)P(i=1)_1}{\sum_{i=1}^2 P(2|i) P(i)_1} \right]^{n_{121}} \\ & \left[\frac{P(j=2|i=2)P(i=2)_1}{\sum_{i=1}^2 P(2|i) P(i)_1} \right]^{n_{221}} \end{aligned}$$

There are three unknowns for the prior data: two conditional probabilities (e.g., $P(j=1|i=1)$ and $P(j=1|i=2)$) and one marginal probability (e.g., $P(i=1)_1$). Thus, when the prior data have been stratified by length, the likelihood still contains information on the conditional probabilities that are used to model the current data. Note that it is necessary to know the results of the length survey, i.e., the Y_{j1} , unlike for the estimators in Section 1 for which knowledge of the Y_{j1} is optional.

The likelihood for the current length sample is a multinomial

$$\Lambda_{\text{current}} \propto \left[\sum_{i=1}^2 P(j=1|i) P(i)_2 \right]^{Y_{12}} \left[\sum_{i=1}^2 P(j=2|i) P(i)_2 \right]^{Y_{22}}$$

The full likelihood for the inverse key under length stratification is simply the product of Λ_{prior} and Λ_{current} .

3.3 Combined key

The data for the current survey are handled in the same way as those from prior surveys when the cross-classified table is generated by fixing the numbers in each length category. These results generalize easily to the case where there are I age categories and J length categories. The complete likelihood for all of the data (prior and current) is proportional to

$$\Lambda \propto \prod_{i=1}^I \prod_{j=1}^J \prod_{k=1}^K \left[\frac{P(j|i) P(i)_k}{\sum_{i=1}^I P(j|i) P(i)_k} \right]^{n_{ijk}} \prod_{j=1}^J \prod_{k=1}^K \left[\sum_{i=1}^I P(j|i) P(i)_k \right]^{Y_{jk}}$$

4. VARIANCE ESTIMATION

Variances and covariances of the estimates of age composition (and classification rates) can be estimated using standard methods for maximum likelihood estimation. For example, if Newton's method is used to find the values of the parameters which maximize the logarithm of the likelihood function (i.e., the maximum likelihood estimates), then the values of the mixed second partial derivatives of the log-likelihood will be obtained as a by-product. The matrix of second partial derivatives evaluated at the parameter estimates is called the information matrix; the negative of the inverse of this matrix is an estimate of the variance-covariance matrix.

The variance-covariance matrix can also be obtained when the parameters are estimated using iteratively reweighted least squares (see section 6).

5. TEST OF ASSUMPTIONS

The combined key is based on the assumption that the conditional probabilities of length given age do not vary from population to population. This assumption can be tested using a likelihood ratio test. For example, suppose that some fish are aged last year and this year. We construct two likelihoods for the data: the first is under the assumption that the $P(j|i)$ are the same in both years (this is the restricted model); the second is under the assumption that there are different $P(j|i)$ in the different years (this is the full model). Note that for the full model we estimate more parameters. The ratio of the maximum of the likelihood for the restricted model divided by the maximum of the likelihood for the full model is the likelihood ratio test statistic λ ; the asymptotic distribution of $-2 \ln \lambda$ is χ^2 with degrees of freedom equal to the difference in the degrees of freedom for the two models.

6. FITTING THE MODEL BY IRLS

Maximizing the likelihoods presented in this paper is not a technically difficult task. However, it is worth noting that the maximum likelihood estimates can be found by iteratively reweighted least squares. Thus, procedures such as PROC NLIN in SAS can be used to obtain the estimates, variance-covariance matrix, likelihood value, etc. The reader is referred to Green (1984) and Jennrich and Moore (1975) for details.

7. DISCUSSION

The methods considered here combine two previously unrelated approaches. The general approach allows one to use previous information whose cost is essentially free. Often, the previous information on age will have been collected according to a scheme in which the population is (post-)stratified by the length categories. This presents no problem if the estimate of the population length composition is known. Hoenig and Heisey (1986) did not stress this point; it appears that some people have used the inverse key approach

without accounting for the fact that the number of fish aged from each length category was fixed by the investigator and was not necessarily proportional to the number of fish in the length category.

The use of previous data is based on the assumption that the classification rates $P(\text{length}|\text{age})$ have not changed from sample to sample. This assumption can be tested using standard methods such as a likelihood ratio test.

We have not encountered any computational difficulties with the general approach. This may be due to the fact that good starting values can usually be obtained by using one of the traditional approaches in a preliminary analysis to obtain starting values.

REFERENCES

- Bartoo, N. and K. Parker. 1983. Stochastic age-frequency estimation using the von Bertalanffy growth equation. *Fish. Bull., U.S.* 81:91-96.
- Berggren, T.J., and J.T. Lieberman. 1978. Relative Contribution of Hudson, Chesapeake, and Roanoke Striped Bass, *Morone saxatilis*, Stocks to the Atlantic Coast Fishery. U.S. National Marine Fisheries Service Fisheries Bulletin, 76, 335-345.
- Clark, W. 1981. Restricted least-squares estimates of age composition from length composition. *Can. J. Fish. Aquat. Sci.* 38:297-307.
- Fukuhara, F.M., S. Murai, J.J. LaLanne, and A. Sribhibhadh. 1962. Continental Origin of Red Salmon as Determined from Morphological Characters. *International North Pacific Fisheries Commission Bulletin* 8, 400-408.
- Green, P.J. 1984. Iteratively Reweighted Least Squares for Maximum Likelihood Estimation, and some Robust and Resistant Alternatives. *J.R. Statist. Soc. B* 46(2):149-192.
- Greenland, S. and D.G. Kleinbaum. 1983. Correcting for Misclassification in Two-way and Matched-pair Studies. *International Journal of Epidemiology*, 12, 93-97.
- Haitovsky, Y. and J. Rapp. 1992. Conditional Resampling for Misclassified Multinomial Data with Applications to Sampling Inspection. *Technometrics* 34, 473-483.
- Hand, D.J. 1986. Estimating Class Sizes by Adjusting Fallible Classifier Results. *Computations & Mathematics with Applications*, 12A, 289-299.
- Hochberg, Y. 1977. On the Use of Double Sampling Schemes in Analyzing Categorical Data with Misclassification Errors. *Journal of the American Statistical Association*, 72, 914-921.
- Hoening, J.M. and D.M. Heisey. 1987. Use of a Loglinear Model with the EM Algorithm for Correcting Estimates of Stock Composition and Converting Length to Age. *Transactions of the American Fisheries Society*, 116, 232-243.
- Jennrich, R.I. and R.H. Moore. 1975. Maximum Likelihood Estimation by Means of Non-linear Least Squares. *Proc. Amer. Statist. Soc. Statistical Computing Section*, 57-65.

- Jolayema, E.T. 1990. Relative Frequency Estimation in Multiple Outcome Measurement with Misclassification. *Biometrical Journal*, 6, 707-711.
- Kimura, D. 1977. Statistical assessment of the age-length key. *J. Fish. Res. Bd. Can.* 34:317-324.
- Kimura, D. and S. Chikuni. 1987. Mixtures of empirical distributions: an iterative application of the age-length key. *Biometrics* 43:23-35.
- Kleinbaum, D.G., L.L. Kupper and H. Morgenstern 1982. *Epidemiologic Research: Principles and Quantitative Methods*, Belmont, CA: Wadsworth
- Pella, J.J. and T.L. Robertson 1978. Assessment of Composition of Stock Mixtures. *U.S. National Marine Fisheries Service Fisheries Bulletin*, 76, 415-423.
- Rogan, W.J. and B. Gladen 1978. Estimating Prevalence from the Results of a Screening Test. *American Journal of Epidemiology*, 107, 71-76.
- Searle, S. 1966. *Matrix Algebra for the Biological Sciences (Including Applications in Statistics)*, New York: Wiley.
- Swensen, A.R. 1988. Estimating Change in a Proportion by Combining Measurements on a True and a Fallible Classifier. *Scandinavian Journal of Statistics*, 15, 139-145.
- Tenenbein, A. 1970. A Double Sampling Scheme for Estimating from Binomial Data with Misclassification. *Journal of the American Statistical Association*, 65, 1350-1361.
- Van Winkle, W., K.D. Kumar and D.S. Vaughan (1988), "Relative Contributions of Hudson River and Chesapeake Bay Striped Bass Stocks to the Atlantic Coastal Population," *American Fisheries Society Monograph*, 4, 255-266.
- Westrheim, S.J. and W.E. Ricker. 1978. Bias in using an age-length key to estimate age-frequency distributions. *J. Fish. Res. Bd. Can.* 35:184-189.
- White, B.S. and K.R. Castleman (1981), "Estimating Cell Populations," *Pattern Recognition*, 13, 365-370.
- Worlund, D.D. and R.A. Fredin (1962), "Differentiation of Stocks," pp. 143-153 in *Symposium on Pink Salmon*, H.R. MacMillan Lectures in Fisheries, University of British Columbia, Vancouver.